



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE GEOLOGÍA

**MODELAMIENTO ESPECTRAL PREDICTIVO DEL CONTACTO ENTRE LA ZONA
PRIMARIA Y SECUNDARIA DEL YACIMIENTO LOS BRONCES, CHILE**

MEMORIA PARA OPTAR AL TÍTULO DE GEÓLOGA

THIARE PALOMA SALAZAR JERIA

PROFESOR GUÍA:
MARTIN REICH MORALES

MIEMBROS DE LA COMISIÓN:
FERNANDO BARRA PANTOJA
DIEGO SANHUEZA MARAMBIO

SANTIAGO DE CHILE

2024

MODELAMIENTO ESPECTRAL PREDICTIVO DEL CONTACTO ENTRE LA ZONA PRIMARIA Y SECUNDARIA DEL YACIMIENTO LOS BRONCES, CHILE

Actualmente, la incorporación de técnicas de *Machine Learning* se ha posicionado como una estrategia clave para transitar hacia una minería más verde y tecnológica. Si bien existen variadas técnicas, el éxito en la implementación de estas depende de las características de los datos, los objetivos y necesidades del estudio. Por otro lado, en el yacimiento Los Bronces, la identificación del contacto entre la zona primaria y secundaria es fundamental para optimizar el procesamiento y la extracción del mineral. Este contacto marca la transición entre roca primaria más competente y con mayor mineralización hipógena y una roca secundaria más fracturada. Así, este trabajo tiene como objetivo evaluar y comparar modelos de *machine learning* para predecir el contacto entre la zona primaria y secundaria del yacimiento Los Bronces a partir de espectros de reflectancia.

Se comparan cuatro modelos supervisados (*Random Forest* (RF), *Support Vector Machine* (SVM), *Logistic Regression* (LR), *Neural Network* (NN)) y un modelo no supervisado (*K-means*). La metodología de preprocesamiento incluye tres escenarios y busca minimizar las distorsiones en los espectros de reflectancia hacia longitudes de onda extremas. Los modelos se construyen con datos de los laboratorios de Perth y Santiago. También, se propone un método de búsqueda doble de hiperparámetros para los modelos supervisados, utilizando primero *RandomizedSearchCV* y luego *GridSearchCV*. Posteriormente, se evalúan métricas de rendimiento como exactitud, precisión, *recall*, *f1-score*, matriz de confusión, homogeneidad, completitud y *v-measure*.

Los modelos LR, SVM y NN mostraron mejor rendimiento, alcanzando métricas de exactitud, precisión, *recall* y *f1-score* sobre el 90% en la mayoría de los casos. Por otro lado, *K-means* no fue capaz de capturar la complejidad de los datos, con métricas cercanas a 0 y llegando a clasificar más de la mitad de los datos erróneamente. En cuanto a RF, no presentó mejores métricas que los otros modelos supervisados y no se recomienda su uso dado su alto costo computacional. Por otro lado, el escenario 1 presentó las métricas más bajas, mientras que el escenario 2 presentó los mejores resultados, seguido del 3, por lo que los modelos son robustos incluso con espectros que presentan valores negativos y reflectancias superiores a 1. Además, la primera búsqueda con *RandomizedSearchCV* demostró ser suficiente en la mayoría de los casos.

Finalmente, se recomienda para investigaciones futuras añadir una componente litológica a la caracterización de los dominios, ya que se identificó alta presencia de BXA, BXBM y BXPR en el dominio primario.

*Dedico mi memoria de título,
A mis sobrinos, para inspirarles.
A Scarlet, por acompañarme y motivarme.
A mi hermana y mi hermano, por inspirarme.
A mi madre y mi padre, por apoyarme y comprenderme.
A Nicky, Menta, Minina y Harry, por darme amor incondicional.*

AGRADECIMIENTOS

Este trabajo no habría sido posible sin el apoyo y colaboración de muchas personas, a quienes me gustaría expresar mi más profundo y sincero agradecimiento.

En primer lugar, agradezco a mi comisión de la memoria por su orientación, paciencia y valiosos consejos que dieron forma a este documento final. Al profesor Martin Reich, a quien admiro enormemente. Sinceramente agradecida por su constante disposición y apoyo en el desarrollo de este trabajo. Al profesor Fernando Barra, mi profesor co-guía, por sus consejos, recomendaciones y opiniones de la investigación que me han guiado para llevar este barco a buen puerto. Finalmente, a Diego Sanhueza, a quién conocí en mi segunda práctica profesional, su confianza y guía me ayudaron enormemente a visualizar la real importancia de este trabajo y desarrollar una investigación científica-tecnológica sin perder u olvidar su aplicación industrial, que es como realmente podemos hacer cambios, vinculando la academia con la industria.

Mi agradecimiento especial a la empresa Anglo American, por proporcionar los datos necesarios que, sin su colaboración, este trabajo no habría sido posible. Sobre todo, agradecida por la oportunidad de realizar mis prácticas profesionales en la división Los Bronces, donde conocí profesionales que hoy admiro y estimo, y de quienes no he recibido más que ayuda desinteresada. Agradecimiento especial al equipo de la Testigoteca Las Tórtolas (Alejandra Rioseco, Diego Ortiz, César Montenegro, Carla Barra, Álvaro Contreras), con quienes compartí comidas y arduas horas de trabajo en terreno. Sin olvidar al equipo de Geomecánica, particularmente a Carlos Barros, quién despertó en mí el bichito de la innovación para y por una minería más tecnológica.

A mis colegas de Graiph, por su apoyo técnico y moral. Gracias por las discusiones constructivas y por compartir su experiencia y conocimientos. Particularmente a Marcelo Surjan, a quién admiro, por su ayuda sincera y buen liderazgo, dándome la oportunidad de trabajar y adquirir experiencia en el mundo real del innovador. Además de MJ y Dani, a quienes admiro y agradezco el aprender de ellas. A Diego, Seba y Manu, por su buena onda y sincera disposición de enseñar. Agradecida de compartir con grandes personas, trabajar en un equipo enérgico y lleno de águilas.

A Kimie Suzuki, mi profesora guía de la tesis, a quien admiro y estimo, gracias por acompañarme y apoyarme durante todo este proceso de trabajo. A Alejandra, Kristian y Raymi, mi team favorito de postgrado, gracias por su buena onda y apoyo. A Eduardo, a quien considero mi mejor amigo de la universidad, muchas gracias por apoyarme, moti-

varme y ayudarme durante todo plan común, por esos cafecitos camino a los auxiliares de la tarde y por todos esos almuerzos que compartimos juntos. A Javier, a quien conocí por los ramos de geología, gracias por tu amabilidad y sinceridad, fue bastante agradable compartir terrenos contigo. También, a Matias, mi partner geoestadístico, sé que llegará bien lejos, fue bastante enriquecedor aprender de ti, gracias por todas esas tardes y noches conversando por Discord de la vida, de la u y de cosas randoms.

A mi papá y mi abuelo, quiénes despertaron mi interés por la minería. Mi padre, Jaime, a quien lleno de preguntas sobre sus trabajos, sobre las minas en las que trabajó y quien se ríe cuando le digo que escriba un libro. Muchas gracias por pasar horas contándome tus historias y gracias por tu trabajo, ahora puedes descansar de la industria y escuchar las historias que te contaremos Bryan y yo.

A mi mamá, Rosa, quien estuvo para todos mis estados de ánimos durante la carrera, agradezco diariamente tener la suerte de que seas mi madre, eres una hermosa persona que brilla, gracias por quedarte viendo dramas coreanos conmigo, sé que eso significaba que tuvieses que pasar de largo al trabajo o dormir 2 horas, pero lo hacías aunque te mandara a dormir.

A mi hermana, Mayra, a quien admiro enormemente, eres una persona de la cual aprendo día a día. Agradezco tu confianza plena en mí, incluso cuando yo no lo veía. A mi hermano, Bryan, a quien también admiro y sé que llegará lejos. Gracias a ustedes y mis padres, por darme todo lo necesario para estudiar y seguir creciendo.

A Scarlet, mi prima y mejor amiga, por estar día y noche durante toda mi vida, gracias por acompañarme en todo este viaje y por motivarme incluso en los momentos más duros.

A mis sobrinos, Maximo y Thomas, quienes llegaron hace un año y meses a mi vida. Espero inspirarles y verlos convertirse en grandes personas, gracias por llenar nuestras vidas de felicidad y amor puro.

A mi madrina, Juanita, gracias por amarme como una segunda hija y por prestarme su computador para el desarrollo de esta memoria. A mi padrino, Rachit, por entenderme y ayudarme. A mi tía Leyla, por sus palabras de ánimo y apoyo. A Paula, a quien quiero como una segunda hermana, gracias por estar con nosotros. A mi prima Anyelina, quien me acompaña a ver dramas, espero que te conviertas en una gran persona.

A Nicky, Harry, Menta y Minina, quienes han tenido que aguantarme, pero que siempre me devuelven amor sincero.

A Selena Gomez, Jonas Brothers, One Direction, Little Mix, The Weeknd, GOT7, EXO, ASTRO, SHINee, quienes han sido parte de mis playlists diarias. Gracias por tales obras de arte, sobretodo a Shinee por Replay/Stand by me y Cha Eun-woo por su cover de 10 minutes que fueron mi cable a tierra esos días sin dormir y que buscaba motivación. También agradecer a Eiichiro Oda, Hajime Isayama, Kohei Horikoshi, Jyoji Morikawa, Makoto Yukimura, Gege Akutami y Koyoharu Gotouge por crear mis animes favoritos.

A todas y todos ustedes, muchas gracias.

TABLA DE CONTENIDO

1. INTRODUCCIÓN	1
1.1. Motivación	3
1.2. Objetivos	6
1.2.1. Objetivo General	6
1.2.2. Objetivos Específicos	6
1.3. Hipótesis	6
1.4. Alcances	7
1.5. Estructura de la memoria	7
2. ANTECEDENTES GENERALES	9
2.1. Ubicación y accesos	9
2.2. Trabajos anteriores	13
3. ESPECTROSCOPIA	18
3.1. Espectroscopía Infrarroja por Transformada de Fourier (FTIR)	21
3.2. ASD Terraspec	22
4. MACHINE LEARNING	23
4.1. Aprendizaje supervisado	26
4.1.1. Clasificación	26
4.1.2. Regresión	26
4.2. Aprendizaje no supervisado	27
4.2.1. <i>Clustering</i>	27
4.3. Aprendizaje por refuerzo	29
4.4. Bosque aleatorio	29
4.5. Regresión Logística	31
4.6. Máquina de vectores de soporte	32
4.7. Redes neuronales	33
5. GEOLOGÍA REGIONAL	35
5.1. Generalidades	35
5.2. Estratigrafía	35
5.2.1. Rocas estratificadas	36

5.2.1.1.	<i>Formación Abanico (Eoceno - Mioceno)</i>	36
5.2.1.2.	<i>Formación Farellones (Mioceno inferior - medio)</i>	37
5.2.2.	Rocas intrusivas	37
5.2.2.1.	<i>Intrusivos Plutónicos (Mioceno)</i>	38
5.2.2.2.	<i>Intrusivos hipabisales a subvolcánicos (Mioceno Superior - Plioceno Inferior)</i>	38
5.2.2.3.	<i>Brechas hidrotermales (Mioceno Superior - Plioceno Inferior)</i>	38
5.2.3.	Depósitos no consolidados	39
5.2.3.1.	<i>Depósitos glaciares (Pleistoceno - Holoceno)</i>	39
5.2.3.2.	<i>Depósitos aluviales y aluvio-coluviales (Holoceno)</i>	39
5.2.3.3.	<i>Depósitos coluviales (Pleistoceno - Holoceno)</i>	39
5.2.3.4.	<i>Depósitos de flujo de detritos (Holoceno)</i>	39
5.2.3.5.	<i>Depósitos antrópicos (Holoceno - actual)</i>	40
5.3.	Geología Estructural	40
5.4.	Geotectónica	40
5.5.	Marco Metalogénico	41
5.6.	Geomorfología	42
6.	GEOLOGÍA DEL YACIMIENTO	45
6.1.	Generalidades	45
6.2.	Litologías	46
6.2.1.	Rocas ígneas intrusivas	46
6.2.1.1.	Batolito San Francisco	46
6.2.1.2.	Pórfidos intra-mineral a tardi-mineral	49
6.2.2.	Rocas ígneas extrusivas	50
6.2.2.1.	Andesita (AND)	50
6.2.2.2.	Riolita (RIO)	50
6.2.3.	Brechas	51
6.2.3.1.	Brechas magmático-hidrotermales de emplazamiento profundo	51
6.2.3.2.	Brechas hidrotermales de emplazamiento medio	52
6.2.3.3.	Brechas freático-hidrotermales de emplazamiento somero	53
6.3.	Alteraciones Hidrotermales	54
6.3.1.	Alteración Potásica (Grupo Potásico)	56
6.3.2.	Alteración Sericita-Gris Verde (Grupo Potásico)	56
6.3.3.	Alteración Albita - Turmalina - Epidota (Grupo Propilítico)	57
6.3.4.	Alteración Clorita - Epidota (Grupo Propilítico)	57
6.3.5.	Alteración Cuarzo - Sericita (Grupo Fílico)	57
6.3.6.	Alteración Clorita - Sericita - Arcillas (Grupo Fílico)	57
6.4.	Vetillas	58
6.4.1.	Vetillas tempranas	58

6.4.1.1.	Vetillas Clase M	58
6.4.1.2.	Vetillas Clase EB	58
6.4.1.3.	Vetillas Clase A	58
6.4.1.4.	Vetillas Clase B	59
6.4.2.	Vetillas transicionales	59
6.4.2.1.	Vetillas Clase C	59
6.4.2.2.	Vetillas Clase C1	59
6.4.2.3.	Vetillas Clase C2	60
6.4.2.4.	Vetillas Clase C3	60
6.4.3.	Vetillas tardías	61
6.4.3.1.	Vetillas Clase D	61
6.4.3.2.	Vetillas Clase DT	61
7.	CONTACTO PRIMARIO-SECUNDARIO	63
8.	METODOLOGÍA	65
8.1.	Base de datos	65
8.2.	Preprocesamiento de los datos	67
8.2.1.	Sondajes	68
8.2.2.	Espectros ASD-FTIR	69
8.3.	Modelo supervisado	71
8.3.1.	Bosque aleatorio	73
8.3.2.	Maquina de vectores de soporte	74
8.3.3.	Regresión Logística	75
8.3.4.	Redes neuronales	76
8.4.	Modelo no supervisado	77
8.5.	Comparación de modelos	77
8.6.	Implementación	78
9.	RESULTADOS	79
9.1.	Preprocesamiento de los datos	79
9.1.1.	Sondajes	79
9.1.1.1.	Filtro 1: Mantener litologías con 100%	79
9.1.1.2.	Compositación	79
9.1.1.3.	Filtro 2: Eliminar filas sin SampleID	81
9.1.1.4.	Filtro 3: Separar en muestras de Perth y Santiago	81
9.1.1.5.	Filtro 4: Ajustar nombre de las muestras a la nomenclatura solicitada	82
9.1.1.6.	Filtro 5: Separar muestras sin espectro	82
9.1.1.7.	Filtro 6: Separar muestras sin contacto Primario/Secundario	83
9.1.2.	Espectros ASD-FTIR	83
9.1.2.1.	Estadísticas	83

9.1.2.2.	Tratamiento de negativos	85
9.1.2.3.	Tratamiento de reflectancia sobre 1	88
9.1.2.4.	Filtro 6: Separar muestras sin espectros	89
9.1.2.5.	Integrar columna del contacto	91
9.2.	Modelo supervisado	92
9.2.1.	Preparación de los datos	92
9.2.1.1.	Normalización de los datos	92
9.2.1.2.	Codificar la variable categórica	93
9.2.1.3.	Dividir en conjunto de entrenamiento, prueba y validación	93
9.2.2.	Bosque aleatorio	94
9.2.2.1.	Búsqueda mejores hiperparámetros	94
9.2.3.	Máquina de vectores de soporte	97
9.2.3.1.	Búsqueda mejores hiperparámetros	97
9.2.4.	Regresión Logística	99
9.2.4.1.	Búsqueda mejores hiperparámetros	99
9.2.5.	Redes neuronales	101
9.2.5.1.	Búsqueda mejores hiperparámetros y selección	101
9.3.	Modelo no supervisado	102
9.3.1.	K-means	102
9.4.	Comparación de modelos	104
9.5.	Implementación	106
9.5.1.	Empaquetamiento de la solución	107
10.	DISCUSIÓN	112
11.	CONCLUSIÓN	115
12.	BIBLIOGRAFÍA	116
ANEXOS		124
ANEXO A.	INTRODUCCIÓN	124
ANEXO B.	MARCO TEÓRICO	126
ANEXO C.	RESULTADOS	131
C.1.	Bosque aleatorio	132
C.2.	Máquina de vectores de soporte	136
C.3.	Regresión Logística	144
C.4.	Redes neuronales	152
C.5.	K-means	154
C.6.	Comparación de modelos	156

ÍNDICE DE TABLAS

3.1.	Principales rangos de absorción en longitudes de onda con grupos moleculares y minerales asociados a estas. Modificado de Ni et al., 2020.	20
4.1.	Pros y contras de las técnicas <i>K-means</i> , Agrupamiento jerárquico y DBSCAN. Extraído de Mery y Morales (2023).	29
9.1.	Estadísticas básicas de Espectros ASD-FTIR Perth, donde en cada fila se tienen los valores del promedio (<i>mean</i>), la desviación estándar (<i>std</i>), mínimo (<i>min</i>), máximo (<i>max</i>), cuantiles 25 %, 50 % y 75 % para cada longitud de onda de la columna.	84
9.2.	Estadísticas básicas de Espectros ASD-FTIR Santiago, donde en cada fila se tienen los valores del promedio (<i>mean</i>), la desviación estándar (<i>std</i>), mínimo (<i>min</i>), máximo (<i>max</i>), percentil 25 %, 50 % y 75 % para cada longitud de onda de la columna.	85
9.3.	Columnas eliminadas de Espectros ASD-FTIR Perth.	85
9.4.	Conjuntos de datos derivados del preprocesamiento de los datos de Espectros ASD-FTIR Perth y Santiago, corresponden a los <i>inputs</i> de los modelos a crear.	91
9.5.	Cantidad de datos de cada set para cada variable.	93
9.6.	Mejores hiperparámetros encontrados en la primera (<i>RandomizedSearchCV</i>) y segunda búsqueda (<i>GridSearchCV</i>) para el clasificador <i>Random Forest</i> , con los datos de <i>Perth_1_Spectrum_Con_Contacto</i> y <i>Perth_2_Spectrum_Con_Contacto</i> . En color amarillo se destacan los valores distintos entre una búsqueda y otra.	94
9.7.	Métricas de evaluación del modelo <i>Random Forest</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para <i>Perth_1</i>	96
9.8.	Hiperparámetros seleccionados del modelo <i>Random Forest</i> para clasificar el contacto primario-secundario, con los datos de <i>Perth_1</i> , <i>Perth_2</i> , <i>Santiago_1</i> y <i>Santiago_2</i> , según corresponda.	97
9.9.	Hiperparámetros seleccionados del modelo <i>Support Vector Machine</i> para clasificar el contacto primario-secundario, con los datos de <i>Perth_1</i> , <i>Perth_2</i> , <i>Santiago_1</i> y <i>Santiago_2</i> , según corresponda.	99

9.10.	Hiperparámetros seleccionados del modelo <i>Logistic Regression</i> para clasificar el contacto primario-secundario, con los datos de Perth_1, Perth_2, Santiago_1 y Santiago_2, según corresponda.	101
9.11.	Modelos seleccionados para cada conjunto de datos, luego de analizar sus métricas de rendimiento.	106
C.1.1.	Mejores hiperparámetros encontrados en la primera (<i>RandomizedSearchCV</i>) y segunda búsqueda (<i>GridSearchCV</i>) para el clasificador <i>Random Forest</i> , con los datos de escenario 1 (Santiago_1_Spectrum_Con_Contacto) y escenario 2 (Santiago_2_Spectrum_Con_Contacto). En color amarillo se destacan los valores distintos entre una búsqueda y otra.	132
C.1.2.	Métricas de evaluación del modelo <i>Random Forest</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para el escenario 2 de Perth (Perth_2).	133
C.1.3.	Métricas de evaluación del modelo <i>Random Forest</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Santiago_1.	134
C.1.4.	Métricas de evaluación del modelo <i>Random Forest</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Santiago_2.	135
C.2.1.	Mejores hiperparámetros encontrados en la primera (<i>RandomizedSearchCV</i>) y segunda búsqueda (<i>GridSearchCV</i>) para el clasificador <i>Support Vector Machine</i> , con los datos de Perth en escenario 3 (Perth_Spectrum_Con_Contacto), escenario 1 (Perth_1_Spectrum_Con_Contacto) y escenario 2 (Perth_2_Spectrum_Con_Contacto). En color amarillo se destacan los valores distintos entre una búsqueda y otra.	136
C.2.2.	Mejores hiperparámetros encontrados en la primera (<i>RandomizedSearchCV</i>) y segunda búsqueda (<i>GridSearchCV</i>) para el clasificador <i>Support Vector Machine</i> , con los datos de Santiago en escenario 3 (Santiago_Spectrum_Con_Contacto), escenario 1 (Santiago_1_Spectrum_Con_Contacto) y escenario 2 (Santiago_2_Spectrum_Con_Contacto). En color amarillo se destacan los valores distintos entre una búsqueda y otra.	137
C.2.3.	Métricas de evaluación del modelo <i>Support Vector Machine</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Perth.	138
C.2.4.	Métricas de evaluación del modelo <i>Support Vector Machine</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Perth_1.	139
C.2.5.	Métricas de evaluación del modelo <i>Support Vector Machine</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Perth_2.	140

C.2.6. Métricas de evaluación del modelo <i>Support Vector Machine</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Santiago.	141
C.2.7. Métricas de evaluación del modelo <i>Support Vector Machine</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Santiago_1.	142
C.2.8. Métricas de evaluación del modelo <i>Support Vector Machine</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearch</i> para Santiago_2.	143
C.3.1. Mejores hiperparámetros encontrados en la primera (<i>RandomizedSearchCV</i>) y segunda búsqueda (<i>GridSearchCV</i>) para el clasificador <i>Logistic Regression</i> , con los datos de Perth en escenario 3 (Perth_Spectrum_Con_Contacto), escenario 1 (Perth_1_Spectrum_Con_Contacto) y escenario 2 (Perth_2_Spectrum_Con_Contacto). En color amarillo se destacan los valores distintos entre una búsqueda y otra.	144
C.3.2. Mejores hiperparámetros encontrados en la primera (<i>RandomizedSearchCV</i>) y segunda búsqueda (<i>GridSearchCV</i>) para el clasificador <i>Logistic Regression</i> , con los datos de Santiago en escenario 3 (Santiago_Spectrum_Con_Contacto), escenario 1 (Santiago_1_Spectrum_Con_Contacto) y escenario 2 (Santiago_2_Spectrum_Con_Contacto). En color amarillo se destacan los valores distintos entre una búsqueda y otra.	145
C.3.3. Métricas de evaluación del modelo <i>Logistic Regression</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Perth.	146
C.3.4. Métricas de evaluación del modelo <i>Logistic Regression</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Perth_1.	147
C.3.5. Métricas de evaluación del modelo <i>Logistic Regression</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Perth_2.	148
C.3.6. Métricas de evaluación del modelo <i>Logistic Regression</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Santiago.	149
C.3.7. Métricas de evaluación del modelo <i>Logistic Regression</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Santiago_1.	150
C.3.8. Métricas de evaluación del modelo <i>Logistic Regression</i> , derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con <i>RandomizedSearchCV</i> y <i>GridSearchCV</i> para Santiago_2.	151
C.4.1. Métricas de evaluación del modelo <i>Neural Network</i> al evaluar en el conjunto de prueba los mejores hiperparámetros para Perth, Perth_1 y Perth_2.	152

C.4.2. Métricas de evaluación del modelo <i>Neural Network</i> al evaluar en el conjunto de prueba los mejores hiperparámetros para Santiago, Santiago_1 y Santiago_2.	153
C.5.1. Métricas de evaluación del modelo <i>K-means</i> para Perth, Perth_1 y Perth_2.	154
C.5.2. Métricas de evaluación de los modelos <i>K-means</i> para Santiago, Santiago_1 y Santiago_2.	155
C.6.1. Métricas de evaluación en los modelos de Perth_1 al evaluar en el conjunto de validación.	156
C.6.2. Comparación final de métricas de evaluación en los modelos <i>Logistic Regression</i> y <i>Neural Network</i> de Perth_1 al evaluar en el conjunto de prueba.	157
C.6.3. Métricas de evaluación en los modelos de Perth_2 al evaluar en el conjunto de validación.	158
C.6.4. Comparación final de métricas de evaluación en los modelos <i>Support Vector Machine</i> y <i>Neural Network</i> de Perth_2 al evaluar en el conjunto de prueba.	159
C.6.5. Métricas de evaluación en los modelos de Santiago_1 al evaluar en el conjunto de validación.	160
C.6.6. Comparación final de métricas de evaluación en los modelos <i>Logistic Regression</i> y <i>Neural Network</i> de Santiago_1 al evaluar en el conjunto de prueba.	161
C.6.7. Métricas de evaluación en los modelos de Santiago_2 al evaluar en el conjunto de validación.	162
C.6.8. Comparación final de métricas de evaluación en los modelos <i>Logistic Regression</i> y <i>Neural Network</i> de Santiago_2 al evaluar en el conjunto de prueba.	163
C.6.9. Métricas de evaluación en los modelos de Perth al evaluar en el conjunto de validación.	164
C.6.10. Comparación final de métricas de evaluación en los modelos <i>Support Vector Machine</i> y <i>Neural Network</i> de Perth al evaluar en el conjunto de prueba.	165
C.6.11. Métricas de evaluación en los modelos de Santiago al evaluar en el conjunto de validación.	166
C.6.12. Comparación final de métricas de evaluación en los modelos <i>Support Vector Machine</i> y <i>Neural Network</i> de Santiago al evaluar en el conjunto de prueba.	167

ÍNDICE DE ILUSTRACIONES

1.1.	Esquema generalizado del patrón de alteración-mineralización en depósitos tipo pórfidos cupríferos, incluye el límite de sulfatos o primario-secundario, asociado en la mina Spence al límite de presencia de sulfatos. Extraído de Rimmelin (2019), modificado de Sillitoe (2010).	2
1.2.	Esquema generalizado del perfil de enriquecimiento supergeno típico de pórfidos cupríferos. Modificado de Makshev (2016) citado en Ávila (2019).	3
2.1.	Mapa de la zona de estudio. Elaboración propia.	11
2.2.	Diagrama de proceso de la Operación Los Bronces. Extraído de Anglo American (2012).	12
2.3.	Principales instalaciones de la Operación Los Bronces. Extraído de Anglo American (2021).	12
3.1.	Comportamiento de un rayo de luz al incidir sobre un material. Ilustración propia.	18
3.2.	Espectro electromagnético con sus respectivos rangos espectrales en nanómetros, se destacan las zonas de interés para la espectroscopía (VIS, NIR, SWIR, MWIR y LWIR). Además, se incluyen los principales procesos de absorción en los respectivos rangos. Modificado de Kerr et al., 2011.	19
3.3.	Funcionamiento de un interferómetro tipo Michelson. 1) Durante la medición, el haz de luz entra en el interferómetro y se dirige hacia el divisor de haz.2) El haz se divide y se proyecta respectivamente hacia un espejo fijo y un espejo móvil. Luego, 3) los haces de luz son reconducidos al divisor de haz donde se recombinan causando interferencia y 4) finalmente son dirigidos hacia la muestra. De esta manera, se adquiere simultáneamente la información espectral de todas las longitudes de onda. Modificado de Bruker, 2019.	21
4.1.	Desglose del campo de Inteligencia Artificial. Modificado de Jhean (2024).	25
4.2.	Tipos de <i>Machine Learning</i> . Modificado de Rajbanshi (2021).	25
4.3.	Visualización de un bosque aleatorio con tres árboles de decisión, donde los círculos azules corresponden a los nodos de decisión, mientras que los círculos verdes y rojos corresponden a los nodos hoja. Modificado de Mokari et al. (2023).	30
4.4.	Representación de una red neuronal artificial <i>multilayer feedforward</i> con tres capas: entrada, ocultas y salida. Modificado de Mokari et al. (2023).	34

5.1.	Mapa geológico regional de Chile central. En un recuadro color rojo se visualiza la ubicación del distrito Río Blanco - Los Bronces, junto con las formaciones Abanico y Farellones, unidades intrusivas y principales fallas activas durante el Neógeno. Modificado de Rojas (2022), basado en Farías et al., 2010, donde AFTB y MFTB corresponden a la faja plegada y corrida de Aconcagua y Malargüe, respectivamente.	36
5.2.	Principales Estructuras Regionales en el Distrito Río Blanco-Los Bronces. En el mapa se observan las tendencias de sistemas de fallas principales y pliegues, donde RB-SF y RC corresponden a abreviaciones de Batolito San Francisco-Río Blanco y Complejo plutónico Río Colorado, respectivamente. Modificado de Piquer, 2015.	43
5.3.	Mapa de los Andes de Chile central, La franja violeta indica la sub-franja de pórfidos cupríferos de Farellones, donde se distribuyen las minas y prospectos mineros de edades Mioceno Medio-Plioceno Inferior. En líneas grises segmentadas se muestra la posición de la Dorsal de Juan Fernández y los contornos de profundidad de la placa oceánica. Extraído de Rojas (2022), modificado de Cortés (2011) y basado en Gígola (2004) y Yañez et al. (2002).	44
6.1.	Mapa litológico de la mina Los Bronces. Extraído de Ortiz (2020), modificado por Vargas (2015) y basado en Abarzúa (2014).	47
6.2.	Modelo 3D de sólidos de alteración hidrotermal en la mina Los Bronces. En a) se observa el halo de alteración propilítica (verde), la cual es cortada y superpuesta por la alteración fílica (amarillo), mientras que en b) se visualiza el núcleo potásico profundo (pardo), de orientación NNW-SSE, cubierto y sobre impuesto por las alteraciones del tipo fílicas. Extraído de Abarzúa (2014).	55
6.3.	Vetilla Clase EB	62
6.4.	Vetilla Clase C	62
6.5.	Vetilla Clase C2	62
6.6.	Vetilla Clase A	62
6.7.	Vetillas Clase D y DT	62
6.8.	Vetillas Clase C3A y C3	62
6.9.	Fotografías de diferentes tipos de vetillas en la mina Los Bronces. Extraído de Rojas (2022).	62
7.1.	Modelo de bloques actual del contacto primario-secundario en la Mina Los Bronces, donde en color rojo se muestra el dominio primario, en verde el secundario y en naranja una transición entre estos.	64
8.1.	Curvas espectrales de muestras analizadas en los laboratorios de Anglo American ubicados en Perth, Australia. Datos sin ningún tipo de tratamiento.	66
8.2.	Curvas espectrales de muestras analizadas en los laboratorios de Anglo American ubicados en Santiago, Chile. Datos sin ningún tipo de tratamiento.	66

8.3.	Visualización 3D de sondajes en <i>Leapfrog Geo</i> con información del contacto primario-secundario, en color rojo tramos mapeados como roca primaria y en color verde tramos mapeados como roca secundaria.	67
8.4.	Visualización 3D de sondajes en <i>Leapfrog Geo</i> coloreados según litología reportada, donde los códigos corresponden tanto a litologías individuales como agrupadas. En color gris las trazas de los sondajes.	67
8.5.	Matriz de confusión para clasificación binaria. Elaboración propia.	72
9.1.	Distribución en la cantidad de tramos para cada litología presente, antes y después de eliminar las filas con menos del 100 % de la litología asociada. .	80
9.2.	Distribución de las litologías presentes en el dominio primario, basado en los datos de mapeo de la mina Los Bronces.	80
9.3.	Distribución de las litologías presentes en el dominio secundario, basado en los datos de mapeo de la mina Los Bronces.	81
9.4.	Distribución de los valores negativos según longitud de onda en Espectros ASD-FTIR Perth (muestras analizadas en laboratorios de Perth).	86
9.5.	Distribución de los valores negativos según longitud de onda en Espectros ASD-FTIR Santiago (muestras analizadas en laboratorios de Santiago). . . .	86
9.6.	Curvas espectrales de las muestras de Perth_1, resultante de eliminar filas con valores negativos de Espectros ASD-FTIR Perth.	87
9.7.	Curvas espectrales de las muestras de Santiago_1, resultante de eliminar filas con valores negativos de Espectros ASD-FTIR Santiago.	88
9.8.	Curvas espectrales de las muestras de Perth_2, resultante de eliminar columnas con valores negativos de Espectros ASD-FTIR Perth.	88
9.9.	Curvas espectrales de las muestras de Santiago_2, resultante de eliminar columnas con valores negativos de Espectros ASD-FTIR Santiago.	89
9.10.	Curvas espectrales de las muestras de Perth_1, resultante de eliminar columnas con valores de reflectancia sobre 1.	90
9.11.	Curvas espectrales de las muestras de Santiago_1, resultante de eliminar columnas con valores de reflectancia sobre 1.	90
9.12.	Distribución de los 50 primeros valores de a) X_Perth y b) X_scaled_Perth, donde X_Perth corresponde a los valores antes de ser normalizados, mientras que X_scaled_Perth presenta los datos transformados.	92
9.13.	Visualización 3D de sondajes en <i>Leapfrog Geo</i> para el escenario 2, coloreados según contacto estimado por los modelos SVM para datos analizados en laboratorios de Perth y LR para datos analizados en laboratorios de Santiago.	107
9.14.	Visualización 3D de sondajes en <i>Leapfrog Geo</i> para el escenario 3, coloreados según contacto estimado por los modelos SVM para datos analizados en Santiago y Perth.	107
9.15.	Interfaz para filtrar espectros según <i>SampleID</i>	108
9.16.	Interfaz para filtrar contacto primario-secundario.	109
9.17.	Entrenamiento de modelos SVM y LR.	110

9.18. Clasificación del contacto a partir de espectros.	111
A.0.1. Plan de trabajo de la memoria de título.	125
B.0.1. Modos vibracionales comúnmente observados en rocas y minerales de alteración, incluyendo sus posiciones en longitud de onda. ν - <i>stretching</i> /estiramiento, δ - <i>bending</i> /flexión, E - emisividad, R - reflectancia, * - Transmisión. Extraído de Laukamp et al., 2021.	127
B.0.2. Cont.	128
B.0.3. Cont.	129
B.0.4. Cont.	130
D.0.1. Distribución de las litologías presentes en el dominio primario, basado en la clasificación del dominio a partir de los modelos del escenario 2.	168
D.0.2. Distribución de las litologías presentes en el dominio secundario, basado en la clasificación del dominio a partir de los modelos del escenario 2.	169
D.0.3. Distribución de las litologías presentes en el dominio primario, basado en la clasificación del dominio a partir de los modelos del escenario 3.	169
D.0.4. Distribución de las litologías presentes en el dominio secundario, basado en la clasificación del dominio a partir de los modelos del escenario 3.	170

CAPÍTULO 1

INTRODUCCIÓN

Los depósitos tipo pórfido cuprífero se definen como grandes volúmenes de rocas hidrotermalmente alteradas, con una zonificación de alteración-mineralización que varía tanto vertical (profundidad) como lateralmente (figura 1.1) . Esta zonificación depende de la proximidad a la fuente termal, la penetración de fluidos meteóricos y los procesos de deformación posteriores a la formación del depósito. De acuerdo al modelo de Sillitoe (2010), estas zonas típicamente se organizan alrededor de un núcleo de alteración potásica, caracterizado por la presencia de biotita, feldespato potásico y minerales de cuarzo, que suele estar directamente asociado con la mineralización principal de cobre y, en ocasiones, oro y molibdeno. Este núcleo estaría rodeado por una extensa zona de alteración fílica (o sericítica), donde puede extenderse la mineralización de cobre y oro, que a su vez esta rodeada por zonas de alteración argílica, caracterizada por la presencia de arcillas, y propilítica, esta última hacia la periferia del sistema. Por lo tanto, esta zonificación refleja un gradiente de temperaturas decrecientes y química de fluidos en evolución desde el centro hacia los márgenes del sistema, delimitando zonas con condiciones características (pH, temperatura) para la formación de minerales específicos.

Por otra parte, dentro de los depósitos tipo pórfido cuprífero se pueden identificar dos macro zonas: zona hipógena y supérgena. La zona hipógena se encuentra en las profundidades del sistema, donde la mineralización original y la alteración están directamente relacionadas con los fluidos hidrotermales magmáticos y las altas temperaturas. Esta zona es rica en minerales sulfurados primarios como la calcopirita, pirita y ocasionalmente molibdenita, cuya estabilidad se ve comprometida en presencia de condiciones superficiales oxidantes sobre el nivel de aguas subterráneas, a las que el sistema es expuesto por procesos de exhumación y erosión. Como resultado, la alteración supergena transforma esta mineralogía hipógena, a través de procesos de oxidación, lixiviación y posterior precipitación, a una nueva mineralogía estable (supérgena).

Lo anterior, se puede ilustrar en un perfil de enriquecimiento supergeno (figura 1.2), donde por los procesos mencionados se forma una zona de gossan, lixiviada estéril, de óxidos y enriquecida (zona supergena) sobre la zona primaria o hipógena. En la parte más superficial, el gossan o sombrero de hierro señala la acción de la meteorización y la

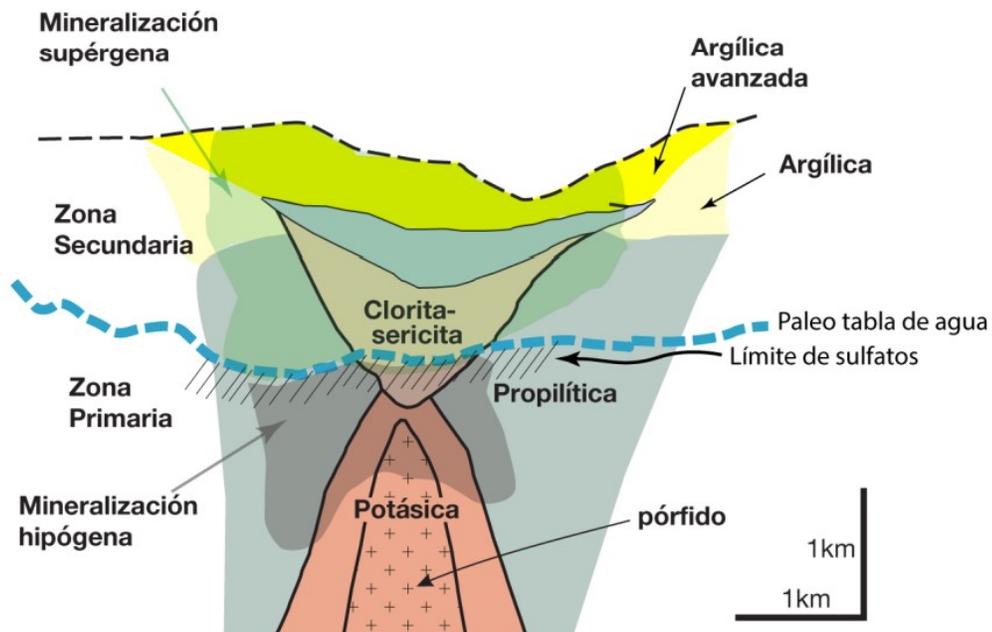


Figura 1.1: Esquema generalizado del patrón de alteración-mineralización en depósitos tipo pórfidos cupríferos, incluye el límite de sulfatos o primario-secundario, asociado en la mina Spence al límite de presencia de sulfatos. Extraído de Rimmelin (2019), modificado de Sillitoe (2010).

oxidación sobre los sulfuros de hierro, dejando tras de sí una capa rica en óxidos e hidróxidos de hierro. Bajo el gossan se encuentra la zona lixiviada estéril, donde la acción del ácido sulfúrico generado por la oxidación de sulfuros ha disuelto y arrastrado la mayoría de los metales, dejando un lecho de roca pobre en mineral. Más abajo, la zona de óxidos marca el comienzo de la concentración secundaria de minerales de cobre, donde los procesos de lixiviación y transporte de iones metálicos han depositado minerales de cobre oxidados como la cuprita, atacamita, malaquita y azurita. Finalmente, se tiene la zona de enriquecimiento, donde los minerales de cobre se encuentran con condiciones reductoras bajo el nivel freático, reaccionando con sulfuros hipógenos para precipitar sulfuros secundarios de cobre de alta ley, como la calcosina y covelina. Por lo tanto, reconocer esta zona de enriquecimiento puede transformar un depósito en una operación minera económicamente viable.

Consecuentemente, estos procesos degradan el estado original de la roca, afectando no sólo la composición mineral sino que también las propiedades físicas, tales como resistencia (competencia), porosidad y permeabilidad. De modo que, en la zona secundaria, cerca de la superficie, se encuentran rocas de mayor meteorización física, con ocurrencia de procesos como lixiviación y oxidación de sulfuros e hidratación de la anhidrita con la consecuente transformación a yeso secundario. En profundidad, en cambio, se presenta una roca geotécnicamente competente, sin fracturas o cuyas fracturas se encuentran selladas con yeso o anhidrita (roca primaria). La zona de transición entre estas unidades

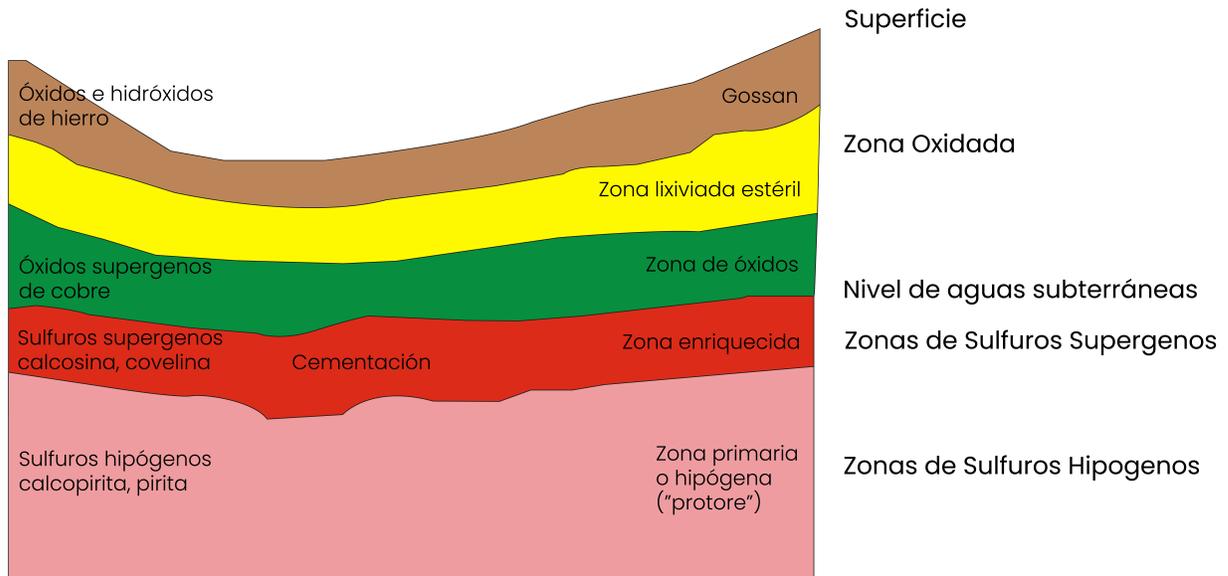


Figura 1.2: Esquema generalizado del perfil de enriquecimiento supergeno típico de pórfidos cupríferos. Modificado de Makshev (2016) citado en Ávila (2019).

se conoce como "contacto primario-secundario", "techo de sulfatos" o "límite de sulfatos", nombres que en este texto se utilizarán como sinónimos.

1.1. Motivación

En la operación de una mina intervienen varios equipos de trabajo, cuyas tareas deben ser coordinadas con especial cuidado para lograr una alta eficiencia y seguridad en la faena. Estos equipos son Geología, Planificación, Operaciones, Mantenimiento y Administración. En cuanto a Geología, son los encargados de reportar las características físicas, químicas y mineralógicas del material a extraer. Sumando la variable anterior a operación, mantenimiento, costos y plazos, el equipo de Planificación elabora el plan minero. Luego, Operaciones se encarga de movilizar el material en la mina a través de las operaciones unitarias de perforación, tronadura, carguío y transporte. En todo momento, el equipo de Mantenimiento debe velar por mantener, valga la redundancia, la disponibilidad mecánica de todos los equipos que participan en la faena. Finalmente, el equipo de Administración, debe brindar el apoyo necesario en manejo de recursos humanos, adquisiciones, contratos, entre otros, que mantengan la operación andando (Ramírez, 2018). Por consiguiente, la información que entrega el equipo de Geología impacta directamente a la planificación minera de la faena.

Con respecto a la planificación mina, en función del nivel de precisión de los datos y de la escala espacial de los períodos de duración, se tienen tres diferentes horizontes de planificación minera: Largo Plazo (LP), Mediano Plazo (MP) y Corto Plazo (CP). La planificación de LP se define en función de las reservas mineras disponibles, entrega

la planificación en un horizonte anual, estableciendo el tamaño de la mina, método y secuencia de explotación, capacidad de producción y el perfil de leyes de corte (Ramírez, 2018). Por otro lado, la planificación de MP, abarca un horizonte de tiempo trianual y anual, produce planes de producción orientados a metas productivas en el corto plazo definidas en el largo plazo, permitiendo retroalimentar la planificación LP. Finalmente, la planificación CP puede ser diaria, semanal, mensual y trimestral, debe recopilar la información operacional, analizando los recursos utilizados, de modo de retroalimentar la planificación de mediano plazo. En particular, la Superintendencia de Planificación de Operaciones de la División Los Bronces debe reportar un Plan Mensual (*Short Term Plan*) con un horizonte de dos meses, basado en los datos proporcionados por las diversas superintendencias de la mina. Este plan se complementa con un plan detallado que se entrega semanalmente, considerando la programación de las actividades presentes en el *Short Term Plan* e integrando detalles operacionales diarios no contemplados en los horizontes más amplios de la planificación. En ambos planes, el equipo de Geología es el encargado de entregar el Modelo de Bloques de la faena. Por lo tanto, reducir los períodos de estimación de las características geológicas del yacimiento, es fundamental para bajar las tasas de incertidumbre en el proceso minero y optimizar la planificación mina.

Por otro lado, la mina Los Bronces, a medida que profundiza sus frentes de producción, se acerca al contacto primario-secundario, donde la roca es más competente y presenta mayor mineralización hipógena. Esto trae consigo modificaciones en el procesamiento del material, ya que las características químicas, mineralógicas y físicas del material primario varían a la mena secundaria, lo que exige ajustar los procesos de conminución y flotación para mantener la eficiencia en la recuperación del mineral. La conminución es la etapa utilizada para reducir el tamaño de los materiales desde el chancado hasta la molienda y su eficiencia depende directamente de la dureza del material. Con respecto a la flotación, depende de la polaridad de los minerales, donde entre más polar la superficie generada en los minerales, más fácil es que se adhieran a moléculas como el agua. Más aún, los efectos de la mineralogía en el procesamiento de minerales se pueden acotar a tres categorías: disolución y precipitación de especies, efectos fisicoquímicos y mineralogía, y oxidación de sulfuros en molienda (Lois, 2015). Primero, la disolución de ciertos minerales puede liberar iones que compiten con los sulfuros o provocar la precipitación de óxidos que alteran las superficies de los minerales, afectando negativamente la recuperación. Por ejemplo, los cationes metálicos liberados pueden formar hidróxidos que reducen la flotabilidad de los sulfuros. En cuanto a efectos fisicoquímicos y mineralogía, el pH y el potencial de oxidación-reducción (Eh) son críticos para mantener condiciones óptimas que favorezcan la actuación de los colectores y eviten la formación de capas superficiales que depriman la flotabilidad. Finalmente, la oxidación de sulfuros durante la molienda, favorecida por la presencia de oxígeno disuelto, genera hidróxidos metálicos en las superficies de los minerales, reduciendo su flotabilidad natural.

En consecuencia, la ubicación del contacto primario-secundario, también conocido como techo de sulfatos, tiene un rol fundamental en el diseño y planificación minera, in-

fluenciando directamente las áreas de geometalurgia y geomecánica. Este límite no solo determina la distribución de la mineralización dentro del yacimiento, sino que también afecta la selección de las estrategias de extracción y procesamiento mineral.

En la división Andina de Codelco, este límite representa la superficie bajo la cual se produce la circulación de aguas subterráneas en el yacimiento y es utilizado para distinguir dos ambientes de calidad de roca contrastantes, roca secundaria de mayor fracturamiento versus roca primaria de escaso fracturamiento, esta última considerada como una unidad impermeable (Ferraz, 2017 en Contreras, 2018). Más aún, en la misma operación minera, se definen seis unidades geometalúrgicas de recuperación (UGM-1 a UGM-6), que se diferencian por su posición relativa respecto al techo de sulfatos, grado de alteración supérgena, composición mineral y mineralización. Estas unidades son esenciales para optimizar la recuperación de cobre, permitiendo a la operación adaptar sus estrategias de extracción y procesamiento a las características específicas de cada zona. Por ejemplo, la UGM-1, que comprende roca primaria bajo el techo de sulfatos, muestra una alta recuperación promedio de 95 % debido a la ausencia de alteración supérgena. En contraste, la UGM-4, roca secundaria con mayor efecto de alteración supérgena y contenido de limonitas, presenta porcentajes de recuperación entre 91 y 92 % (Luengo, 2018).

En el caso de la división El Teniente, la migración de los avances en la explotación de la roca secundaria a primaria, requirió un cambio en el método de explotación, pasando del *Block Caving*, utilizado en la extracción de mena secundaria, al *Panel Caving* (Varela, 2019). Este último es un método subterráneo masivo que socava la base de un panel de producción mediante perforación y tronadura para inducir el quiebre natural del macizo rocoso, aprovechando la gravedad. Se adapta especialmente a la roca primaria, caracterizada por ser menos fracturada y más competente que la roca secundaria, resultando en una fragmentación más gruesa del mineral. Por otra parte, las rocas primarias y secundarias pueden presentar diferencias en su reactividad ante explosivos. Contreras (2018) estudió la reactividad en muestras de roca primaria y secundaria de la división Andina, reconociendo sus implicancias en las operaciones de tronadura y seguridad en la mina. En su estudio, las rocas primarias presentaron un tiempo de inducción más largo cuando se exponen a nitrato de amonio, un componente común en los explosivos utilizados en minería (como el tipo ANFO utilizado en la mina Los Bronces), lo que sugiere que las rocas primarias son menos propensas a reacciones exotérmicas rápidas y potencialmente peligrosas con este tipo de explosivos. Las rocas secundarias, en cambio, mostraron una tendencia a reaccionar más rápido con el nitrato de amonio (tiempos de inducción más cortos), lo que se asocia a su mayor porosidad y la presencia de minerales alterados que pueden actuar como catalizadores en la reacción. Con todo lo anterior, las rocas secundarias implican un riesgo mayor de detonaciones no deseadas o prematuras durante las operaciones de tronadura, lo que recalca la importancia de delimitar el límite entre roca primaria y secundaria.

Por lo tanto, reconocer el contacto primario-secundario trasciende la geología y requiere una mirada multidisciplinaria para optimizar la precisión del mismo, ya que no sólo

influye en los modelos geológicos sino que impacta procesamiento mineral, tronadura, diseño y planificación minera. Actualmente, el contacto primario-secundario es actualizado de manera anual, por lo tanto es inminente optimizar la caracterización de este, bajo esta necesidad se desarrolla esta memoria, en la que se propone que mediante el desarrollo de modelos de *machine learning*, que permiten encontrar patrones en los datos y aprender de los mismos, es posible predecir y caracterizar el contacto primario-secundario, en este caso, a partir del análisis de espectros de reflectancia de muestras de la mina Los Bronces.

1.2. Objetivos

1.2.1. Objetivo General

Evaluar modelos de *machine learning* para predecir el contacto entre el dominio primario y secundario, a partir de espectros de reflectancia de muestras de la mina Los Bronces.

1.2.2. Objetivos Específicos

1. Establecer una metodología de preprocesamiento para los datos geológicos de mapeo y los espectros de reflectancia para asegurar su compatibilidad en la construcción de los modelos, considerando las limitaciones de los datos disponibles.
2. Desarrollar modelos supervisados para predecir el contacto primario-secundario en la Mina Los Bronces a partir de los espectros de reflectancia, obtenidos con el equipo combinado de ASD Terraspec y FTIR, y la información geológica de mapeo.
3. Desarrollar modelos no supervisados para caracterizar la zona primaria y secundaria, basados en espectros de reflectancia obtenidos con el equipo combinado de ASD Terraspec y FTIR.
4. Recomendar el modelo más adecuado para la predicción del contacto primario-secundario basado en las métricas de rendimiento obtenidas y la aplicabilidad a los datos de la mina Los Bronces.

1.3. Hipótesis

La caracterización del contacto primario-secundario es un proceso esencial en la operación, pero actualmente es actualizado de manera anual, lo que genera incertidumbre en la planificación mina al no contar con la información en períodos de tiempo más acotados. Por otro lado, constantemente se envían muestras a análisis por ASD Terraspec y/o espectrómetros FTIR, muestras que muchas veces cuentan con una clasificación del

dominio derivada del mapeo en terreno. Así, viendo la gran cantidad de datos que se generan diariamente en la faena, se abre una oportunidad de añadir valor a estos mediante la implementación de técnicas de *machine learning*. Se propone que la incertidumbre del contacto se puede minimizar al incorporar la predicción del mismo a partir de espectros electromagnéticos, obtenidos con equipos ASD Terraspec y/o espectrométros FTIR, y modelos de *machine learning*.

1.4. Alcances

El objeto de estudio corresponderá a espectros de reflectancia y cartillas de mapeo. Con respecto a los espectros de reflectancia, estos se obtienen de muestras de pulpa analizadas con espectrómetros FTIR y ASD Terraspec en laboratorios de la compañía ubicados en Perth y Santiago. En cuanto a las cartillas de mapeo, se tiene una con la descripción por tramos del contacto primario - secundario, mientras que la otra corresponde a la descripción litológica por tramos. Se analizan datos de entre los años 2001 y 2023. La zona de estudio corresponde a la mina Los Bronces, explotada por Anglo American Sur S.A., ubicada en la Cordillera de Los Andes Central a una altura aproximada de 3500 m.s.n.m en la comuna de Lo Barnechea, Región Metropolitana de Santiago.

1.5. Estructura de la memoria

La memoria se encuentra estructurada de la siguiente manera:

En el Capítulo 2 se presenta la ubicación y accesos de la zona de estudio, así también como un revisión bibliográfica de trabajo anteriores realizados en la mina Los Bronces.

En el Capítulo 3 se presenta una revisión bibliográfica del estado del arte en espectroscopía, enfocado en la espectroscopía de reflectancia, considerando la espectroscopía infrarroja por transformada de Fourier (FTIR) y ASD Terraspec.

En el Capítulo 4 se presenta una descripción de los métodos de *Machine Learning* existentes, dando mayor detalle de los utilizados para el desarrollo de esta memoria.

En el Capítulo 5 se presenta la descripción de la geología regional en la zona de estudio, a nivel distrital, considerando estratigrafía, geología estructural, geotectónica, marco metalogénico y geomorfología.

En el Capítulo 6 se detalla la geología del yacimiento Los Bronces, considerando definiciones locales de litologías, alteraciones hidrotermales y vetillas.

En el Capítulo 7 se presenta un descripción del contacto primario - secundario, según literatura.

En el Capítulo 8 se detalla la metodología de la memoria, considerando la descripción de la base de datos, preprocesamiento de los datos, descripción de los modelos supervisados a utilizar, así también como los modelos no supervisados. Finalmente, se detalla el paso de comparación de modelos y su implementación en faena.

En el Capítulo 9 se presenta un análisis de los resultados de la memoria, considerando las etapas anteriores de metodología.

En el Capítulo 10 se exponen los puntos de discusión de la memoria, posibles mejoras y recomendaciones para futuras investigaciones de este tipo.

Finalmente, en el Capítulo 11 se escriben las conclusiones del proyecto y los principales hallazgos.

En Anexos A.0.1 se incluye el plan de trabajo inicial de la memoria.

CAPÍTULO 2

ANTECEDENTES GENERALES

2.1. Ubicación y accesos

La zona de estudio corresponde a la mina Los Bronces, explotada por Anglo American Sur S.A.. Se ubica en el distrito geológico Los Bronces - Río Blanco - Los Sulfatos, que constituye depósitos de clase mundial de pórfidos de Cu y Mo de Chile Central. Esta operación minera se localiza en la Cordillera de Los Andes Central a una altura aproximada de 3500 metros sobre el nivel del mar (m.s.n.m) en la comuna de Lo Barnechea, Región Metropolitana de Santiago, aproximadamente a 65 km al noreste de Santiago (véase Figura 2.1).

El flujo de operación en la mina Los Bronces (figura 2.2) consiste en la extracción del material a través de explotación a cielo abierto, en bancos de 15 metros de altura. El mineral es tronado utilizando explosivos tipo ANFO y otros típicos en minería, para luego ser cargado por palas (eléctricas o hidráulicas) y/o cargadores frontales en camiones de alto tonelaje para su posterior transporte. El mineral extraído es llevado a un chancador primario, el que luego es transportado por medio de correas transportadoras al acopio de mineral grueso. Seguido de esto, se ingresa el mineral proveniente del acopio a una etapa de molienda para reducir el tamaño del material a partículas pequeñas y formar una pulpa de mineral, ya que en esta etapa se agrega agua de proceso. Así, la pulpa de mineral es enviada a estanques espesadores donde se extrae el agua hasta alcanzar la densidad requerida que permita su transporte hidráulico a la planta de flotación Las Tortólas. En cuanto al área de Las Tortólas, la pulpa ingresa a una planta de flotación colectiva compuesta por etapas de flotación primaria, remolienda de concentrados, flotación de limpieza y flotación de barrido. Como resultado, se obtiene un concentrado mixto de cobre y molibdeno, el cual es enviado a una planta de flotación selectiva, donde se separan los concentrados de cobre y molibdeno. En cuanto al mineral de baja ley, se envía a las pilas de lixiviación, mientras que el estéril es enviado a botadores. Finalmente, el material proveniente de las pilas de lixiviación es procesado en la Planta San Francisco, planta de extracción por solventes y electroobtención (SX-EW), para obtener cátodos de cobre. Con respecto a las instalaciones de la operación, estas abarcan el área de Los

Bronces, Mineroducto y Las Tortólas (Anglo American, 2012), indicadas en la Figura 2.3 y descritas a continuación.

- **Los Bronces**

El área de Los Bronces comprende el rajo Los Bronces, botaderos, plantas de molienda Los Bronces y Confluencia, planta de producción de cátodos San Francisco, campamento y talleres de mantención. Por lo que en el área no sólo se lleva a cabo el proceso extractivo, si no que también los procesos de chancado, molienda, riego, extracción por solvente (EX) y electro-obtención (EW) de cátodos desde el mineral oxidado.

En cuanto al acceso, la principal ruta de acceso hacia la mina Los Bronces es el camino Santiago-Farellones (G-21) hasta Corral Quemado aproximadamente, donde comienza el camino hacia Paso Marchant y Los Bronces (G-245).

- **Mineroducto**

El mineroducto es un Sistema de Transporte de Pulpa (STP) de 56 km de longitud, transporta el mineral sulfurado en pulpa entre Los Bronces y Las Tortólas. Este sistema cruza los valles de los esteros San Francisco, Arrayán y Colina, entre las comunas de Lo Barnechea y Colina.

- **Las Tórtolas**

El área comprende la planta de flotación y el tranque de relaves, por lo que es la operación encargada de la producción de concentrados de cobre y molibdeno. Esta localizada en el valle central de la Región Metropolitana, en la provincia de Chacabuco, comuna de Colina, aproximadamente a 30 kms. al norte de Santiago.

En cuanto al acceso, la ruta principal es por la Autopista Los Libertadores que une Santiago con Los Andes, aproximadamente en el kilómetro 30. Más aún, existe otro acceso desde el camino Quilapilún - Polpaico desde la Ruta 5 Norte.

En relación a la producción de cobre en el año 2023, se alcanzaron 215 500 toneladas de cobre fino y 1 937 toneladas de molibdeno. Según el informe del cuarto trimestre 2023 (Anglo American Chile, 2024), la división Los Bronces registró una baja de 32 % a 57 200 toneladas, debido a menores leyes y cantidad tratada por causa de la continua mayor dureza del mineral. Sumado a esto, informes anteriores señalan la severa sequía de la zona central como uno más de los grandes desafíos para la operación (Anglo American Chile, 2023a).

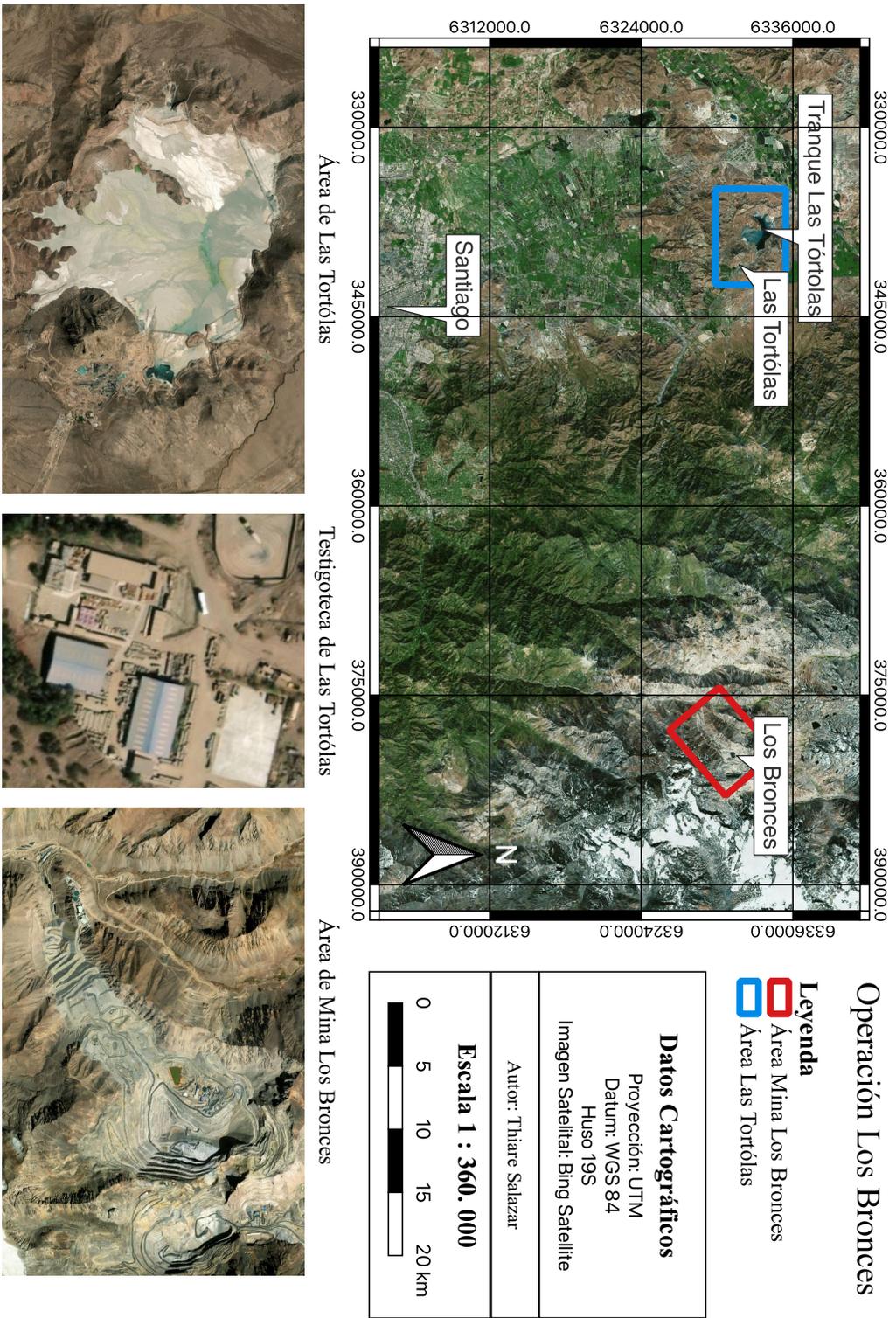


Figura 2.1: Mapa de la zona de estudio. Elaboración propia.

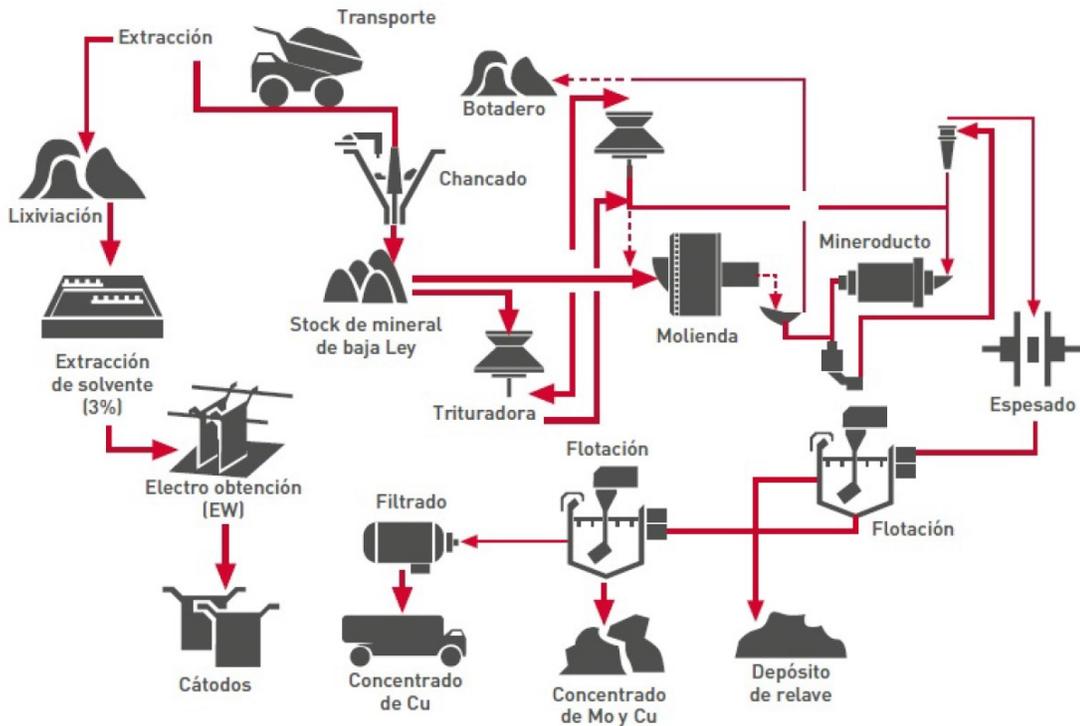


Figura 2.2: Diagrama de proceso de la Operación Los Bronces. Extraído de Anglo American (2012).

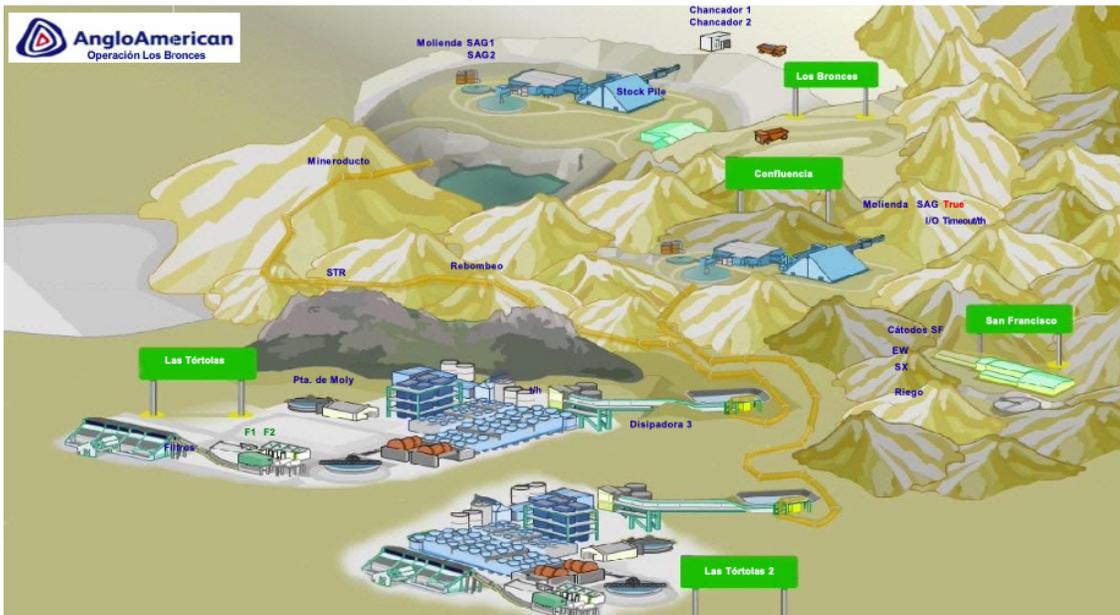


Figura 2.3: Principales instalaciones de la Operación Los Bronces. Extraído de Anglo American (2021).

2.2. Trabajos anteriores

En la zona se han realizado distintos estudios por parte de la compañía y otras/os investigadoras/es, por mencionar algunos tipos se tienen estudios de caracterización de la geología regional, geología económica, geología estructural, origen y factores de control de emplazamiento del yacimiento, así también como estudios innovadores, por ejemplo, mapeo de zonas de alteración mineral hidrotermal y supérgena mediante imágenes hiperespectrales satelitales (Sorrentino et al., 2024) y el diseño optimizado de plantas de procesamiento (Velásquez et al., 2020). A continuación, se detalla parte de la bibliografía revisada.

Warnaars et al. (1985) describen Los Bronces como un complejo de brechas superpuesto a un sistema temprano de pórfido cuprífero, formado al este del batolito San Francisco (intrusión peraluminosa y de carácter calco-alcalina). Los autores indican que el depósito está compuesto de siete brechas de turmalina, categorizadas por su ubicación, matriz, clastos, formas, tipos y grados de alteración y mineralización. Las cuales son, de más vieja a más joven, Brecha Fantasma, Central, Occidente, Infiernillo, Anhidrita, Gris Fina y Donoso. Finalmente, los autores interpretan que las brechas en Los Bronces se emplazaron explosivamente, seguido de un colapso producto de la despresurización de los fluidos hidrotermales. Por lo que, internamente, están localmente bien definidas, pero en otra parte, coalescen, se entrelazan o muestran contactos graduales.

Vargas et al. (1999) mediante petrografía, geoquímica y estudio de inclusiones fluidas en brechas Sur-Sur y La Americana del depósito Río Blanco - Los Bronces, señalan una relación temporal y espacial de la mineralogía de mena de las brechas.

Contreras (2003) modela la geología de la Mina Los Bronces mediante el estudio de sondajes profundos y considera que esta es el resultado de la evolución de un sistema de pórfido cuprífero que presenta todas las etapas de alteración. Más aún, describe dos tipos de pórfidos: Pórfido Cuarzomonzodiorítico (PQMD) y Pórfido Cuarzomonzonítico (PQD). Y, finalmente, define tres brechas más a las ya conocidas, estas son: Brecha Ígnea, Brecha de Magnetita y Brecha de Biotita.

Greß & Spröhnle (2003) redefinen las unidades litológicas separando la roca huésped Cuarzomonzonita (QM), en Cuarzomonzonita de grano grueso (QMG) y Cuarzomonzonita de grano fino (QMF). Además, destacan dos zonas de exploración: Extensión suroeste del Rajo Los Bronces y al lado este del complejo de brechas Los Bronces.

Skewes et al. (2003) centran su estudio en la Brecha Donoso, una de las 15 brechas del depósito Río Blanco - Los Bronces. Le asignan una edad entre 5.2 y 4.9 Ma, y describen la mineralogía de la matriz, los tipos de clastos, alteración y mineralización asociada a la brecha. Además, mediante estudios de petrología, inclusiones fluidas e isótopos estables, señalan que el origen de la brechización y mineralización metálica de la Brecha Donoso fue causado por la exsolución de fluidos magmáticos producto del enfriamiento del plutón parental

Davidson et al. (2005) estudian los precursores magmáticos de los fluidos hidrotermales que dieron origen al depósito Río Blanco-Los Bronces, ligando esto a los magmas silicatados y al transporte de metales.

Deckart et al. (2005) estudian la cronología de los eventos magmáticos e hidrotermales mediante los métodos U-Pb y $^{40}\text{Ar}-^{39}\text{Ar}$. Indican que la mineralización de Cu-Mo está temporalmente vinculada al emplazamiento de al menos tres tipos de pórfidos: cuarzo-monzonítico, feldespático y Don Luis, pertenecientes al conjunto de los pórfidos tardíos. Y, por lo tanto, fue contemporánea al ascenso de fundidos dacíticos a niveles subvolcánicos.

Jara (2005) realiza un estudio de bornitas en el sector Los Bronces y, mediante estudios de microscopía, define dos etapas de formación de la bornita. La primera etapa asociada a una alteración potásica con temperaturas mayores a 460°C, mientras que la segunda etapa estaría asociada a vetas tardías ricas en minerales de arsénicos, donde la bornita se presenta reemplazando a calcopirita en un ambiente a temperaturas menores a 228°C.

Muñoz et al. (2006) estudian la estratigrafía, geoquímica y edades de las rocas volcánicas pertenecientes a la Formación Abanico Este.

Barros (2007) realiza el estudio de distribución de As, Bi, Pb, Sb, Ag y Au en el sector Los Bronces mediante petrografía, calcografía, técnicas de absorción atómica y microsonda electrónica. Concluye que estos elementos mineralizan en etapa tardía del yacimiento, hospedándose en vetas tardías con orientación NE en el sector norte de la mina y con orientación NW en el sector Sur.

Irarrázaval et al. (2010) realizan un resumen de la historia del descubrimiento del yacimiento de clase mundial Los Sulfatos, indicando que la mineralización de cobre-molibdeno estaría relacionada a lo menos a dos centros porfídicos y a un complejo de brechas magmático-hidrotermal.

Toro et al. (2012) definen el distrito a través de tres dominios geológicos: el bloque Los Piches - Ortiga, el bloque San Manuel - El Plomo y el bloque Río Blanco - Los Bronces - Los Sulfatos. En cada uno de estos dominios, describen las litologías y las alteraciones hidrotermales presentes. Más aún, presentan un modelo evolutivo del sistema.

Agus (2011) investiga el uso de imágenes hiperespectrales de alta resolución, tomadas con el sistema SisuCHEMA, para evaluar los tipos de mica blanca y los estilos de alteración presentes en guijarros molidos de la mina Los Bronces. Estos guijarros corresponden a fragmentos de roca normalmente descartados durante la molienda que aún pueden contener concentraciones significativas de cobre, mayores a 0.4%. También, utiliza el algoritmo *Spectral Angle Mapper* (SAM) para la clasificación mineral de los datos y generación de mapas minerales. Más aún, estudia la relación de cristalinidad de las micas blancas, estimada comúnmente por la relación entre la profundidad de absorción característica de AlOH en 2200 nm y la absorción de H₂O en 1900 nm. Los resultados revelan un patrón en el que la mineralización de cobre se corresponde con muestras que exhiben longitudes de onda más cortas, una alta cristalinidad y absorciones intensas en

el grupo de la mica blanca, lo que sugiere que la cristalinidad de la mica blanca podría ser un indicador confiable de zonas con mayores leyes de cobre.

Abarzúa (2014) realiza una descripción de los principales tipos de alteración hidrotermal presentes en el yacimiento Los Bronces. Explora su distribución, temporalidad y criterios de identificación, dividiendo las alteraciones en los grupos potásico, propílitico y fílico. Adicionalmente, lleva a cabo un modelamiento espacial de las alteraciones en la mina

Dalm et al. (2014) emplean un espectrómetro portátil ASD Fieldspec3 para medir la reflectancia en el rango VNIR-SWIR de 150 muestras de la mina Los Bronces. Estas muestras, derivadas del material de gran tamaño producido por el molino semi-autógeno (SAG) de la planta, consisten en trozos de roca de aproximadamente 5-7 cm de diámetro. Los autores determinan que existe una relación indirecta entre la ley de cobre y la mineralogía activa en el rango VNIR-SWIR. Además, concluyen que la mica blanca de baja cristalinidad y la clorita podrían utilizarse como indicadores de desecho (*waste*), mientras que la presencia de mica blanca de alta cristalinidad y turmalina se asocian a muestras de mineral (*ore*). También, señalan que la mica blanca de baja cristalinidad está asociada con una alteración argílica, en contraste, la mica de alta cristalinidad estaría vinculada a una alteración fílica. De manera similar, la ocurrencia de clorita se asociaría con la alteración argílica y la turmalina con la alteración fílica. Cabe mencionar que, empleando regresión logística multivariable, postulan que es posible estimar de manera confiable la probabilidad de que la ley de Cu de una muestra esté por debajo de la ley de corte para el procesamiento mineral, utilizando su curva de reflectancia en el rango VNIR-SWIR.

Vargas (2015) realiza un estudio acerca de la distribución de plata en la mina Los Bronces, describiendo la existencia de cuatro especies de plata asociadas a eventos tardíos del sistema y en asociación con piritita, calcopiritita, tennantita, tetraedrita, esfalerita y galena. Describiendo este como el último evento mineralizador del sistema, en un ambiente de sulfidación intermedia con temperaturas menores a 250°C.

Quigley & Yildirim (2015) presentaron un estudio sobre la utilización de sistemas de registro de sondajes hiperespectrales automatizados, utilizando HCI-3 de Corescan y HyLogger-3 desarrollado por CSIRO, para caracterizar minerales y definir dominios geometalúrgicos en la mina Los Bronces. Su investigación confirmó la presencia de varios minerales en el rango SWIR, validados por XRD, como la moscovita (sericita), turmalina, yeso, clorita, biotita, y trazas de caolinita y montmorillonita. Así también como cuarzo y feldespatos identificados por HyLogger-3 en el rango TIR. Este estudio resalta el aporte de la tecnología hiperespectral en la mejora del diseño de flujos geometalúrgicos y la predicción de comportamientos de procesamiento mineral.

Sánchez (2017) hace uso del sistema HyLogger3TM para obtener espectros de reflectancia infrarroja de tramos de sondajes representativos de las alteraciones potásica y cuarzo sericita en el yacimiento Los Sulfatos. Se identificaron mineralogías predominantes como cuarzo, feldespatos potásicos, albita, turmalina, muscovita, yeso, flogopita y

clorita. Además, los datos se contrastaron con otras técnicas como difracción de rayos X, caracterización macroscópica, petrografía y geoquímica multielemento. Así, utilizando la información anterior, identificaron asociaciones mineralógicas características para los tipos de alteraciones en estudio. Finalmente, concluye que HyLogger junto con el software TSG es eficaz para la caracterización mineral y modelación.

Dalm et al. (2017) utilizan la cámara hiperespectral SisuCHEMA para analizar 43 muestras del estudio de Dalm et al. (2014) y producir mapas con la distribución de minerales activos en el rango SWIR, así como obtener la cristalinidad y composición de la mica blanca, y la composición de la clorita. Este estudio resalta las ventajas de las imágenes hiperespectrales sobre espectrómetros, como el utilizado en Dalm et al. (2014), incluyendo 1) la capacidad de proporcionar datos más completos de la composición mineral de una muestra, incluidos los minerales presentes en cantidades relativamente bajas; 2) permite la identificación y extracción de espectros minerales puros que se pueden utilizar para caracterizar la composición mineral y/o cristalinidad; 3) habilita la cuantificación de la composición mineralógica activa SWIR en la superficie de una muestra contando el número de píxeles que representan ciertos minerales y/o características minerales; 4) permite la caracterización y evaluación de la textura de las muestras a partir de mapas minerales.

Ortiz (2020) introduce nuevas técnicas para mejorar el modelo mineralógico de la mina Los Bronces. Se utiliza un nuevo enfoque con muestras de pozos de tronadura, análisis automatizados TIMA-X y análisis elemental FRX. Los resultados revelan una representación más precisa de minerales y detectan nuevas anomalías mineralizadas en comparación con el modelo anterior. También se identifican distribuciones de óxidos de cobre no estudiados previamente. Estos resultados tienen un alto potencial para futuras investigaciones en la mina Los Bronces.

Rojas (2022) emplea datos de televiewer óptico y acústico de la mina Los Bronces para identificar y caracterizar vetillas, fracturas y otras estructuras geológicas en profundidad. Lo anterior, le permite establecer tendencias en la orientación de las vetillas y asociarlas con zonas de alteración, del mismo modo, clasifica las vetillas según su tipo y correlaciona estas tendencias con estructuras geológicas mayores, teniendo potencial implicancias para comprender la dirección del emplazamiento de estructuras mineralizadas.

Sorrentino et al. (2024) evalúan la utilidad de imágenes hiperespectrales satelitales obtenidas por PRISMA, una misión satelital de la Agencia Espacial Italiana (ASI, por sus siglas en inglés), en la identificación de zonas de alteración hidrotermal y supergéna asociadas a depósitos IOCG (Proyecto de Cobre Marimaca) y pórfidos Cu-Mo (distrito Río Blanco - Los Bronces) en el rango VNIR-SWIR. La metodología del estudio se basa en la combinación de índices de extracción de características de absorción diagnósticas en un flujo de trabajo multifacético para mapear alteraciones a gran escala y detectar variaciones composicionales en minerales claves como micas blancas, cloritas, sulfatos (alunita y jarosita), minerales del grupo de la caolinita, óxido e hidróxidos de hierro (hematita y goetita). Lo anterior no sólo permite la identificación de áreas amplias de alteración mineral

sino también la detección de sutiles variaciones composicionales que pueden ser cruciales para vectorizar hacia cuerpos mineralizados y definir nuevos *targets* de exploración. Finalmente, los resultados revelan una zonificación de sulfatos proximales (natroalunita a alunita) y mica blanca rica en Al a mica blanca pobre en Al distal y una zona externa de clorita-epidota, evidenciando patrones de alteración hidrotermal ascendente y externa bien definidos. Los autores destacan la utilidad de la espectroscopía de imágenes hiperespectrales espacial como un valioso apoyo a la exploración mineral al habilitar un mapeo no invasivo sobre grandes áreas.

CAPÍTULO 3

ESPECTROSCOPIA

La espectroscopía es el estudio de la luz como función de su longitud de onda que ha sido emitada, reflejada o dispersada por un sólido, líquido o gas. Cuando un material es sometido a un haz de luz, algunos fotones son reflejados, otros atraviesan y otros son absorbidos (Clark, 1999). Esta interacción haz de luz-material se manifiesta en forma de ondas electromagnéticas que pueden ser medidas en función de su longitud de onda. Para ilustrar mejor, la figura 3.1 muestra el comportamiento de un haz de luz al incidir sobre un material, este puede ser absorbido, propagarse de forma difusa, reflejarse y/o transmitirse.

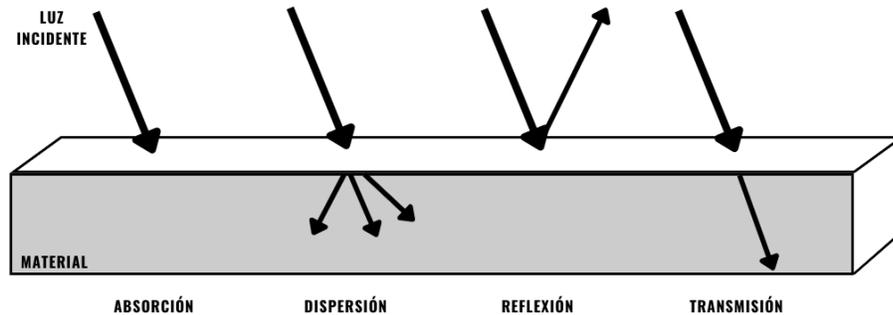


Figura 3.1: Comportamiento de un rayo de luz al incidir sobre un material. Ilustración propia.

Como se mencionó anteriormente, cuando los fotones emitidos por una fuente interactúan con un mineral, pueden ser reflejados, absorbidos o transmitidos. En la espectroscopía de reflectancia, los fotones que son reflejados son detectados y medidos por espectrómetros obteniendo curvas de reflectancia o espectros, las cuales indican la cantidad de energía reflejada en rangos de longitud de onda. Más aún, la cantidad de energía reflejada depende de las características químicas y físicas del mineral. Por lo tanto, cada mineral posee un espectro característico que entrega información sobre su química y permite su identificación (Clark, 1999), ya que distintos minerales producen distintos

espectros.

Para analizar la química mineral, la espectroscopía de reflectancia típica considera los rangos Visible (VIS), Infrarrojo cercano (NIR, por sus siglas en inglés) e Infrarrojo de onda corta (SWIR, por sus siglas en inglés). Sin embargo, existen instrumentos especializados que permiten medir rangos mayores, como los rangos Infrarrojo de onda media (MWIR, por sus siglas en inglés), de onda larga (LWIR, por sus siglas en inglés) y termal (TIR, por sus siglas en inglés). Así, el universo de todos los rangos espectrales conforman el espectro electromagnético de la Figura 3.2, donde VIS corresponde al rango entre 0.35 a 0.75 micrómetros (μm), NIR entre 0.75 y 1.3 μm , SWIR entre 1.3 y 2.5 μm , MWIR entre 3.0 y 5.0 μm , LWIR entre 7.0 y 13.0 μm y TIR entre 6.0 y 1000 μm .

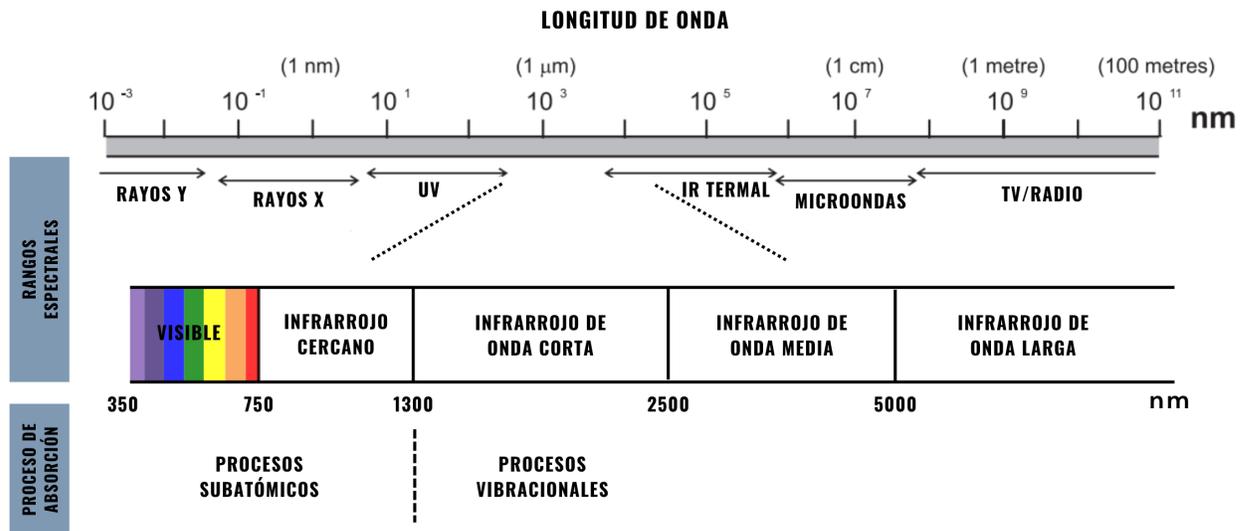


Figura 3.2: Espectro electromagnético con sus respectivos rangos espectrales en nanómetros, se destacan las zonas de interés para la espectroscopía (VIS, NIR, SWIR, MWIR y LWIR). Además, se incluyen los principales procesos de absorción en los respectivos rangos. Modificado de Kerr et al., 2011.

Dependiendo del rango de detección, ocurren distintos procesos de absorción y reflectancia. En el rango VNIR (0.35-1.3 μm), los procesos de absorción y reflectancia se dan producto de las transiciones electrónicas, que consisten en la liberación de energía cuando un electrón cambia de un nivel a otro. Estos procesos controlan el color del mineral y pueden asociarse a la presencia de elementos de transición, por ejemplo, la presencia de hierro férrico o ferroso permite identificar minerales con Fe como la hematita, goetita, jarosita, clorita y turmalina (Kerr et al., 2011). Más aún, según Clark (1999), las transiciones electrónicas y sus efectos en la absorción y reflectancia en minerales se deben a varios factores clave, como:

- Efecto de campo cristalino: Influencia en los niveles de energía de los electrones de elementos de transición en un cristal, afectando la absorción de luz y el color visible

del mineral.

- **Absorciones por Transferencia de Carga:** Involucra el movimiento de electrones entre diferentes iones o entre iones y ligandos, resultando en absorciones significativas y colores distintivos en ciertos minerales.
- **Bandas de Conducción:** Diferencias entre las bandas de conducción y valencia en ciertos minerales, que pueden influir en su capacidad para absorber y reflejar luz, afectando su color.
- **Centros de Color:** Defectos en la red cristalina que crean niveles de energía discretos, afectando la absorción de luz y provocando colores característicos en ciertos minerales.

Por otro lado, en el rango SWIR, los fotones son absorbidos por procesos vibracionales. Cuando los fotones interactúan con un mineral, generan vibraciones en los enlaces moleculares. Cada enlace tiene modos de vibración únicos del mineral, conocidos como *stretching* (estiramiento) y *bending* (flexión), por lo tanto, un enlace individual puede absorber en múltiples frecuencias de longitud de onda del rango infrarrojo. En este rango, grupos de minerales como filosilicatos, sorosilicatos, inosilicatos, carbonatos y zeolitas exhiben espectros característicos con rasgos diagnósticos de absorción de ciertos grupos moleculares (por ejemplo, OH, H₂O y NH₄) y enlaces cationes hidroxilo (por ejemplo, Al-OH, Fe-OH y Mg-OH) dentro de la red cristalina del mineral (Hunt, 1977). En cambio, en el rango LWIR, es posible identificar grupos minerales como tectosilicatos (por ejemplo, feldespato, cuarzo, albita), inosilicatos (por ejemplo, diópsido, anfíboles), fosfatos (por ejemplo, apatito) y sulfatos (Kraal y Ayling, 2019). En la tabla 3.1 se muestran las principales franjas espectrales por grupo molecular y mineral asociado, mientras que en la figura B.0.1 se muestra a mayor detalle las principales absorciones en posiciones de longitud de onda de una variedad de minerales.

Tabla 3.1: Principales rangos de absorción en longitudes de onda con grupos moleculares y minerales asociados a estas. Modificado de Ni et al., 2020.

POSICIÓN (nm)	GRUPO MOLECULAR	GRUPO MINERAL
1350 - 1560	Hidroxilo (OH)	Arcillas, Zeolitas, Sulfatos, Clorita, Biotita, Anfíboles, Talco
1900 - 2000	Agua (H ₂ O)	Esmectita, Illita, Yeso, Zeolitas
2160 - 2220	Hidróxido de Aluminio (AlOH)	Caolinita, Pirofilita, Illita, Montmorillonita, Alunita
2240 - 2300	Hidróxido de Hierro (FeOH)	Clorita, Biotita, Jarosita, Epidota
2260 - 2320	Carbonato (CO ₃)	Calcita, Dolomita
2300 - 2360	Hidróxido de Magnesio (MgOH)	Anfíboles, Ankerita, Biotita, Calcita, Clorita, Dolomita, Serpentina, Talco

3.1. Espectroscopía Infrarroja por Transformada de Fourier (FTIR)

La espectroscopía de infrarrojos por transformada de Fourier (FTIR, por sus siglas en inglés) corresponde a una técnica de la espectroscopía que difiere de los métodos convencionales. En lugar de analizar las muestras irradiándolas con longitudes de onda únicas de manera secuencial, la técnica FTIR captura datos espectrales de todas las longitudes de onda simultáneamente en una sola pasada. Esto se logra utilizando una fuente de luz infrarroja que emite un amplio rango de longitudes de onda, las cuales pasan a través de un interferómetro y luego inciden en la muestra (Figura 3.3). Posterior a esto, se obtiene un interferograma, que es una señal bruta representando la intensidad de la luz no en función de la longitud de onda. Luego, mediante la transformación de Fourier, se convierte este interferograma en un espectro en función de la longitud de onda. Así, las ventajas de esta técnica sobre los métodos dispersivos son su rapidez, precisión y presentar una mejor relación señal-ruido (Bruker, s.f.).

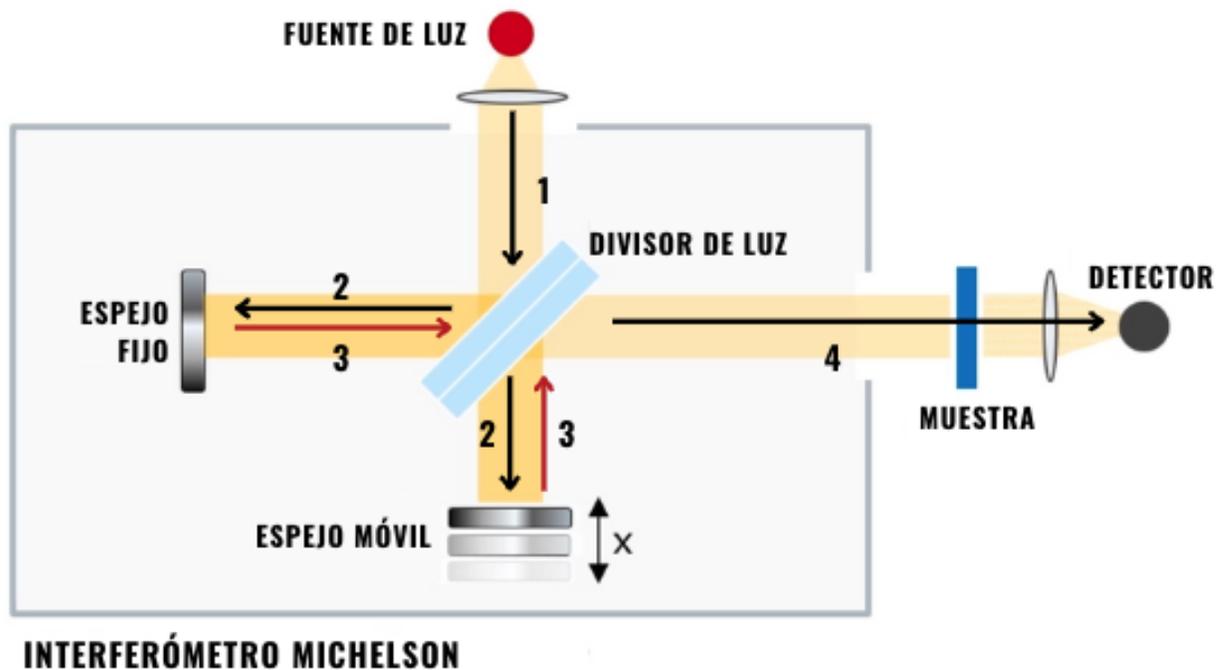


Figura 3.3: Funcionamiento de un interferómetro tipo Michelson. 1) Durante la medición, el haz de luz entra en el interferómetro y se dirige hacia el divisor de haz. 2) El haz se divide y se proyecta respectivamente hacia un espejo fijo y un espejo móvil. Luego, 3) los haces de luz son reconducidos al divisor de haz donde se recombinan causando interferencia y 4) finalmente son dirigidos hacia la muestra. De esta manera, se adquiere simultáneamente la información espectral de todas las longitudes de onda. Modificado de Bruker, 2019.

3.2. ASD Terraspec

El espectrómetro ASD TerraSpec es una herramienta avanzada para la exploración minera que proporciona un análisis rápido y preciso de los minerales. Trabaja en el rango VNIR-SWIR (350-2500 nm), permitiendo capturar datos espectrales precisos de manera rápida, incluso en muestras con características minerales más oscuras. Inclusive, se puede combinar con el software *The Spectral Geologist* (TSG) de *Spectral Geoscience* para analizar los espectros y producir productos de salida mineralógicos. Dentro de sus aplicaciones clave destacan análisis de arenas bituminosas y lutitas bituminosas; producción minera y procesamiento de minerales para evaluar y monitorear propiedades minerales; análisis de corte de perforación, registro principal, determinación de gradientes geoquímicos y delineación de especies de arcilla; exploración de minerales, asignación de depósitos, identificación del ensamblador de minerales y más (Malvern Panalytical, s.f.). Una desventaja de este instrumento es el rango espectral que abarca, ya que minerales comunes como cuarzo y feldespatos no son distinguibles dentro del rango VNIR-SWIR. Por lo anterior, se puede complementar con instrumentos como FTIR para proporcionar datos aún más útiles, brindando mayor detalle al rango VNIR-SWIR dentro del espectro capturado por la técnica FTIR. Por ejemplo, FTIR puede estimar la cantidad total de mica blanca, mientras que los resultados de ASD Terraspec pueden contribuir para indicar el tipo y cristalinidad de la mica blanca identificada. Más aún, empresas como Bureau Veritas han incorporado el servicio de análisis espectral y *Machine Learning* (ML), que consiste en la captura de información de muestras mediante FTIR y espectroscopía ASD, combinación de los espectros e interpretación por medio de modelos matemáticos de ML. Lo anterior, es empaquetado en la aplicación BV Prediction, que permite obtener una huella digital de cada una de las muestras y predecir condiciones de procesamiento futuras, parámetros metalúrgicos, parámetros mineralógicos, y propiedades químicas y físicas (Bureau Veritas Chile, s.f.).

CAPÍTULO 4

MACHINE LEARNING

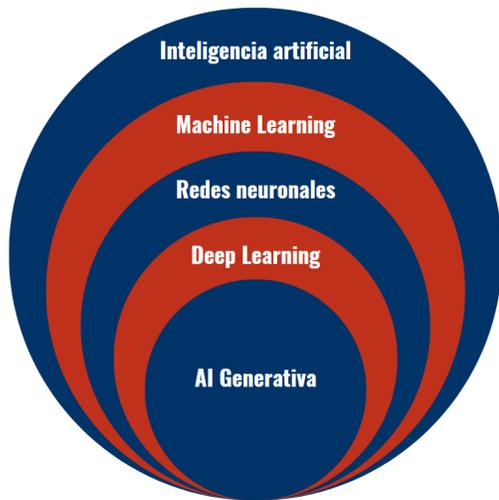
La inteligencia artificial (IA) es una disciplina de la informática que se centra en la creación de sistemas capaces de emular el comportamiento humano, utilizando máquinas para aprender y realizar tareas que, normalmente, requieren inteligencia humana (Wagh, 2023). La IA se divide en varios subcampos, cada uno con enfoques y aplicaciones específicas, entre los más destacados se pueden mencionar (Figura 4.1):

- ***Machine Learning (ML) o Aprendizaje de máquina:*** es un subconjunto de la inteligencia artificial que entrena a los sistemas o máquinas para aprender de los datos y tomar decisiones o predicciones basadas en patrones.
- ***Deep Learning (DL) o Aprendizaje profundo:*** es una subdisciplina de ML que se basa en redes neuronales artificiales profundas, es decir, con múltiples capas de entrada, salida y ocultas. Cada capa contiene unidades que transforman los datos de entrada en información que la siguiente capa puede usar para una tarea predictiva específica. Los modelos de aprendizaje profundo pueden ser discriminativos o generativos. Los modelos discriminativos se entrenan típicamente en un gran conjunto de datos de muestras etiquetadas, con el objetivo de descubrir la relación entre las características de las muestras y las etiquetas dadas. Una vez que un modelo discriminativo está entrenado, se puede utilizar para predecir la etiqueta de cualquier nueva muestra. Un modelo generativo, por otro lado, crea nuevas instancias de muestra basándose en una distribución de probabilidad aprendida de las muestras existentes (Hadid et al., 2024).
- ***Generative AI (GAI) o Inteligencia artificial generativa:*** es un subconjunto de modelos de DL capaces de generar contenido nuevo y original, como imágenes, música, vídeos o texto que no existía previamente (Hadid et al., 2024). Estos sistemas están diseñados para aprender e imitar los patrones y características de un tipo particular de datos, y luego usar ese conocimiento para crear nuevo contenido similar. Modelos como las redes adversariales generativas (GANs), los transformadores, los autoencoders variacionales y los Modelos de lenguaje de gran tamaño, utilizan grandes

conjuntos de datos y técnicas de aprendizaje profundo para generar nuevas muestras que imitan los datos de entrenamiento originales. Uno de los principales avances en GAI fue la introducción de las GANs en 2014, en las cuales una red genera contenido (modelo generativo) y la otra ayuda a determinar si es una muestra auténtica o no (modelo discriminativo). Otro avance significativo fue la introducción de los Transformadores, que han revolucionado el procesamiento de lenguaje natural y visión por computadora. Mientras que, más recientemente, modelos como ChatGPT han demostrado amplias capacidades en tareas que van desde la generación de texto hasta la programación, con aplicaciones en áreas como biología, medicina, educación, legislación y finanzas (Hadid et al., 2024).

En el campo de la minería y las geociencias, estos subtipos de IA están revolucionando la manera en que las y los científicos interpretan y entienden los procesos complejos. Desde la generación y aumento de datos hasta la simulación y toma de decisiones, sin duda, la IA está proporcionando herramientas innovadoras que mejoran los procesos convencionales. Por ejemplo, pueden identificar patrones y relaciones en los datos que los humanos podrían malinterpretar, lo que lleva a interpretaciones más precisas y fiables, como predecir actividades sísmicas analizando patrones en datos históricos, ayudando así en la gestión de desastres; e incluso asistiendo en la exploración de recursos al identificar sitios óptimos para la extracción, aumentando la eficiencia y reduciendo el impacto ambiental de estas actividades (Hadid et al., 2024). Sin embargo, la integración de la IA en la geociencia no está exenta de desafíos, como el acceso a datos de alta calidad y sin sesgos, e incluso falta de datos etiquetados para el entrenamiento de modelos.

En particular, *Machine Learning* (ML) es una disciplina científica dentro del campo de la Inteligencia Artificial (IA) que permite a las máquinas aprender y mejorar a partir de la experiencia sin estar explícitamente programadas (Géron, 2019). Este aprendizaje se logra a través de algoritmos que identifican patrones en grandes volúmenes de datos, permitiendo realizar predicciones o tomar decisiones de manera autónoma (Petrelli, 2021). Según Mitchell (1997), un programa informático se considera que aprende de una experiencia E en relación con una tarea T y una medida de rendimiento P , si su desempeño en T , medido por P , mejora con E . Así, ML integra estadística, inteligencia artificial e informática. La aplicación de modelos de ML es bastante popular, aplicándose en áreas como el reconocimiento de voz, visión computacional, control de robots y procesamiento de lenguaje natural (Petrelli, 2021). En minería, los algoritmos de ML son utilizados para exploración, análisis mineralógico, optimización de la extracción y modelamiento, contribuyendo a la mejora de la seguridad y la eficiencia operativa (Azhari et al., 2023). Por otro lado, los procesos de aprendizaje en ML se pueden dividir en: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo (Figura 4.2). A continuación, se explica cada uno de ellos.



Inteligencia artificial (IA)

Es un campo de estudio que abarca el desarrollo de técnicas que equipan a las máquinas para emular el comportamiento humano, siendo capaces de realizar tareas que típicamente requieren de inteligencia humana.

Machine Learning (ML)

Es un subcampo de la IA que utiliza algoritmos avanzados para detectar patrones en largos conjuntos de datos, permitiendo a las máquinas aprender y adaptarse.

Redes neuronales (NN)

Es un tipo de algoritmo de ML que emula la estructura y funcionamiento del cerebro humano, permitiendo a los sistemas de IA aprender y procesar información compleja.

Deep Learning (DL)

Es un subcampo de ML que utiliza NNs con múltiples capas para tareas de procesamiento y análisis, con el propósito de extraer características de alto nivel, simulando la manera en que el cerebro humano percibe y entiende el mundo.

AI Generativa

Es un subconjunto de modelos de DL que generan contenido como texto, imágenes o códigos, basados en entradas proporcionadas.

Figura 4.1: Desglose del campo de Inteligencia Artificial. Modificado de Jhean (2024).

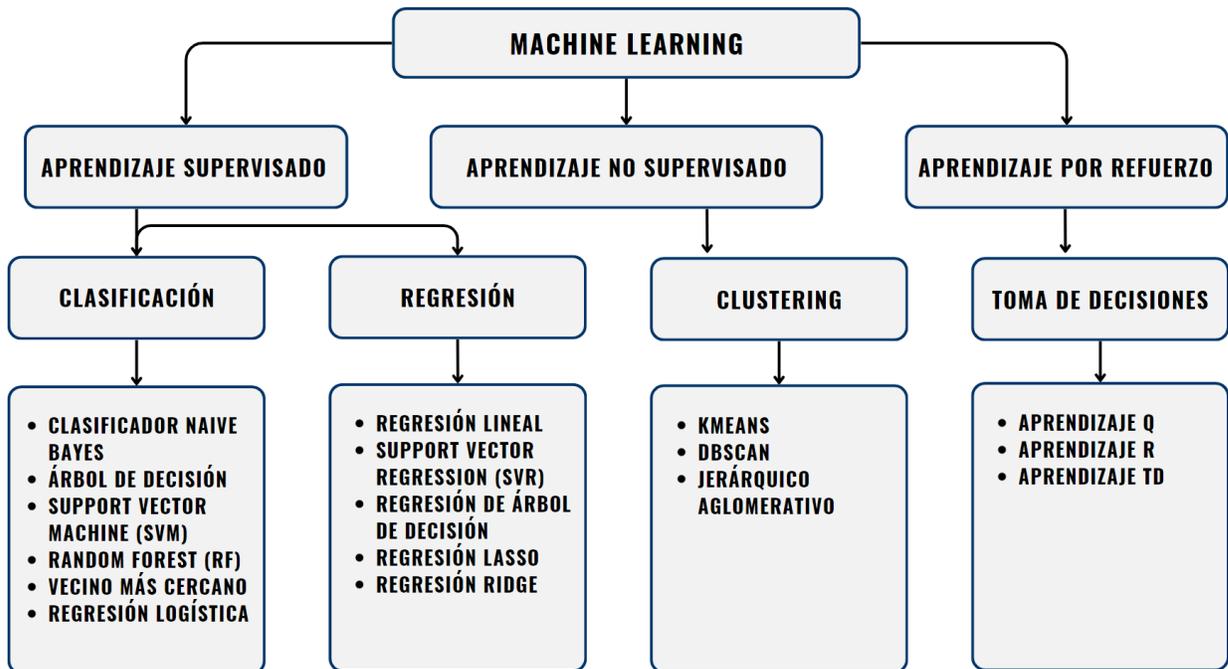


Figura 4.2: Tipos de Machine Learning. Modificado de Rajbanshi (2021).

4.1. Aprendizaje supervisado

En el aprendizaje supervisado, el modelo se entrena con *inputs* y sus respectivos *outputs*, por lo tanto, el conjunto de entrenamiento incluye soluciones deseadas, conocidas como etiquetas (Géron, 2019). En base al tipo de problema, puede catalogarse como clasificación o regresión. Ambas categorías utilizan un conjunto de datos etiquetados para el entrenamiento, permitiendo al modelo aprender y hacer predicciones o clasificaciones precisas sobre nuevos datos. A continuación, se presentan a detalle.

4.1.1. Clasificación

Se utiliza para tareas donde se requiere clasificar elementos en categorías discretas, intenta definir el límite de decisión que mejor divide los datos en clases. Comúnmente utilizado en automatización y generalización. Por ejemplo, la detección de spam en correos electrónicos, donde los correos se clasifican como spam o no spam. En el campo de la espectroscopía, tiene como objetivo agrupar los espectros en diferentes clases, basado en sus patrones y características distintivas. Se utilizan variadas técnicas y algoritmos, cada uno con sus características y aplicaciones específicas.

- Árboles de decisión: utilizan una estructura de árbol para tomar decisiones secuenciales y dividir los datos en grupos más pequeños basados en atributos específicos.
- K-Vecinos más cercanos (K-NN): clasifica los datos basándose en la proximidad o similitud con los datos de entrenamiento más cercanos.
- Máquinas de vectores de soporte (SVM): buscan el mejor hiperplano que separa las clases con un margen máximo, siendo eficaces en espacios de alta dimensión.
- *Naive Bayes*: utiliza el teorema de Bayes para clasificar datos, asumiendo independencia entre los atributos.
- Redes neuronales: modelos complejos capaces de aprender patrones complicados a través de una estructura inspirada en el cerebro humano.
- *Gradient Boosting*: combina múltiples modelos débiles para crear un modelo más fuerte y preciso.

4.1.2. Regresión

Se aplica en situaciones donde se debe predecir un valor continuo, intenta obtener la mejor línea o curva que prediga los *outputs* de forma precisa. Por ejemplo, determinar las temperaturas pre-eruptivas y las profundidades de almacenamiento en sistemas de conductos volcánicos mediante la termobarometría de piroxeno, basándose en variables de entrada como la química de los clinopiroxenos y las relaciones entre las variaciones de entropía y volumen (Petrelli, 2021). Las regresiones deben cumplir 5 supuestos:

1. Linealidad: los datos deben poder ajustarse a una recta.
2. No endogeneidad: no deben existir relaciones entre los errores y las variables independientes.
3. Normalidad y homocedasticidad de los residuos: los errores de medición presentan una distribución normal y los residuos tienen igual varianza.
4. No autocorrelación de los residuos: no deben existir relaciones entre los errores.
5. No multicolinealidad: si dos o más variables tienen alta correlación, si se puede, se debe representar una variable usando la otra.

Con respecto a algoritmos más relevantes, estos son:

- Regresión lineal: establece una relación lineal entre las variables independientes y la variable dependiente. Utiliza mínimos cuadrados para estimar los coeficientes.
- *Support Vector Regression (SVR)*: plantea la regresión como un problema de optimización, buscando un hiperplano que maximice el número de puntos sujetos a un margen de tolerancia.
- *Decision Tree Regression*: construye un árbol de decisión que divide el espacio de atributos en regiones pequeñas, asignando un valor promedio a cada una para realizar predicciones.
- Regresión Lasso: aplica regularización L1, útil para eliminar variables irrelevantes y simplificar el modelo.
- Regresión Ridge: utiliza regularización L2, reduciendo la influencia de variables altamente correlacionadas y mejorando la estabilidad del modelo.

4.2. Aprendizaje no supervisado

En el aprendizaje no supervisado, los algoritmos analizan datos no etiquetados, buscando patrones o estructuras sin instrucciones específicas. En este contexto, el sistema intenta aprender sin un 'profesor' y descubre patrones ocultos en los datos. Una tarea común es el agrupamiento (*clustering*), donde los datos se agrupan en subconjuntos o clústeres basados en similitudes.

4.2.1. *Clustering*

Se utiliza para dividir datos, encontrar patrones o características comunes, definir estructura natural de los datos, encontrar grupos homogéneos (reducción de dimensionalidad) y encontrar valores atípicos. Los clústeres o grupos pueden ser excluyentes, traslaparse o probabilísticos. Cuando un individuo pertenece a un único grupo, se dice que

es un clúster excluyente, si los grupos se traslapan uno a otro, es un clúster traslapado y si un individuo pertenece a un grupo con cierta probabilidad, es un clúster probabilístico. Para validar los clústeres se utilizan medidas internas y/o externas. Dentro de las medidas internas se encuentra el índice de silueta, índice de Dunn y la compactacidad. El índice de silueta es una medida de la cohesión y separación de los clústeres, ecuación 4.1 , donde valores cercanos a 1 indican buena división de los clústeres. Mientras que el índice de Dunn es una medida de la separación entre clústeres y de compactación dentro de los clústeres, ecuación 4.2, por lo que un valor alto indica que los clústeres están bien definidos. Por otro lado, la compactacidad indica qué tan cercanos son los puntos dentro de un clúster, así, un valor bajo indica que los clústeres están bien divididos.

$$SI = \frac{\text{distancia intracluster} - \text{dist.intercluster}}{\max(\text{dist.intracluster}, \text{dist.intercluster})} \quad (4.1)$$

$$DI = \frac{\min(\text{distancia intercluster})}{\max(\text{dist.intracluster})} \quad (4.2)$$

Existen distintos tipos de algoritmos de *clustering*. Los particionales que dividen el espacio sin traslape de clústeres, por ejemplo, técnicas como *K-means*, método isodata y nubes dinámicas. Los jerárquicos que generan conjunto de clústeres anidados definidos por la fusión o subdivisión de individuos, por ejemplo, técnicas aglomerativas y divisivas. Y, los basados en densidad que generan clústeres basados en la densidad del área, por ejemplo, la técnica DBSCAN. Dentro de las técnicas mencionadas, destacan:

- *K-means* : método iterativo que agrupa los datos en un número k predefinido de clústeres basándose en la minimización de una métrica de distancia de los datos a cada clúster. Para ejemplificar su implementación, se define como métrica de distancia la suma de la distancia cuadrática. Primero, se seleccionan k datos como centroides iniciales. Luego, se calcula la distancia al centroide de cada dato y forma k clústeres considerando la menor distancia al centroide. Posterior a esto, se recalcula el centroide de cada clúster. Después, se itera hasta converger, es decir, hasta que los centroides no cambien.
- DBSCAN: identifica clústeres basados en la densidad de los datos, agrupando puntos cercanos y marcando puntos aislados como anomalías. Su procedimiento inicia seleccionando un punto arbitrario. Se visita la vecindad de este y si tiene suficientes puntos, se marca como clúster 1. Se van agregando puntos si pertenecen a la vecindad, esta depende de un radio especificado. Luego, se itera hasta visitar todos los puntos.
- Agrupamiento Jerárquico: construye jerarquías de clústeres, ya sea agrupando progresivamente puntos de datos individuales (enfoque aglomerativo) o dividiendo un clúster inicial en subgrupos (enfoque divisivo). El agrupamiento jerárquico aglomerativo cuenta con el siguiente procedimiento, primero, se calcula la matriz de distancias.

Luego, se combina el par de clústeres más cercano y se actualiza la matriz de distancias. Posterior a esto, se itera hasta que quede sólo un clúster. Finalmente, se selecciona el número de clústeres.

En la tabla 4.1 se enlistan algunos pros y contras de cada técnica a tener en consideración a la hora de utilizar una u otra.

Tabla 4.1: Pros y contras de las técnicas *K-means*, Agrupamiento jerárquico y DBSCAN. Extraído de Mery y Morales (2023).

	K-means	Agrupamiento jerárquico	DBSCAN
Pros	Simple de entender, fácil de implementar y ampliamente disponible. Resultados interpretables.	No requiere definir nº. clústeres. Fácil visualización de resultados (dendograma).	No requiere definir nº. clústeres. Permite identificar outliers. Permite generar clústeres de múltiples formas.
Contras	Sensible al nº. de clústeres. Sensible a la configuración de los centroides iniciales. Sensible a los outliers. Genera soluciones esféricas. Estandarización necesaria.	No adecuado para largas bases de datos. Computacionalmente demandante. Una vez unidos dos clústeres, no se pueden dividir. No hay función objetivo a minimizar.	No adecuados para bases de datos con densidades variables. No adecuados para bases de datos de alta dimensionalidad.

4.3. Aprendizaje por refuerzo

En el aprendizaje por refuerzo, un agente interactúa con su entorno realizando acciones y recibiendo recompensas o penalizaciones. El objetivo es aprender la mejor estrategia o política para maximizar las recompensas acumuladas a lo largo del tiempo (Géron, 2019). Este tipo de aprendizaje se diferencia del supervisado y no supervisado en que no hay datos de entrenamiento etiquetados ni se proporcionan respuestas correctas. En su lugar, el agente debe descubrir por sí mismo, a través de la prueba y error, qué acciones le llevan a resultados más favorables.

4.4. Bosque aleatorio

La técnica bosque aleatorio o *Random Forest* (RF) se basa en árboles de decisión y se puede utilizar tanto para regresión como para clasificación. Utiliza una estructura de árbol

para tomar decisiones secuenciales y dividir los datos en grupos más pequeños (subnodos) basados en atributos específicos. Esencialmente, crea una cascada de decisiones binarias basadas en subconjuntos de variables, tal como se ilustra en la figura 4.3. *Random Forest* es una combinación de árboles predictivos, una modificación del *Bagging*, en la cual se trabaja con una colección de árboles incorrelacionados y los promedia. Estos árboles de decisión se entrenan dividiendo iterativamente los datos para descubrir el método óptimo para agruparlos, con el objetivo de reducir la impureza de Gini ($g(\tau)$), donde la impureza de Gini de un nodo se calcula según la ecuación 4.3. En la ecuación anterior, n es el número total de espectros en el nodo, y n_A y n_B son los números de espectros pertenecientes a la clase A o B, respectivamente (primario o secundario). El proceso se itera hasta lograr un índice de Gini de 0, lo que indicaría que las muestras dentro de un nodo pertenecen a una sola clase (Mokari et al., 2023).

$$g(\tau) = 1 - \left(\frac{n_A}{n}\right)^2 - \left(\frac{n_B}{n}\right)^2 \quad (4.3)$$

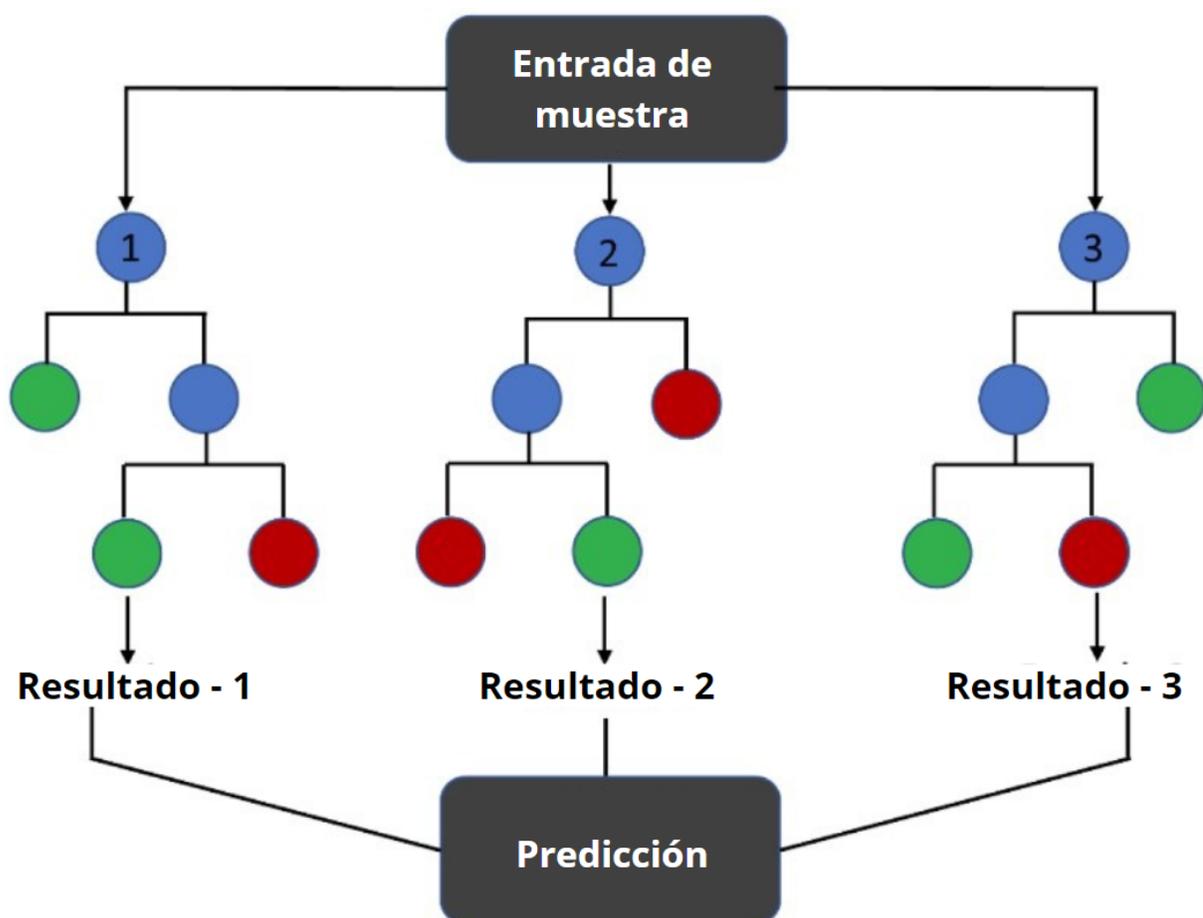


Figura 4.3: Visualización de un bosque aleatorio con tres árboles de decisión, donde los círculos azules corresponden a los nodos de decisión, mientras que los círculos verdes y rojos corresponden a los nodos hoja. Modificado de Mokari et al. (2023).

Cada árbol se desarrolla usando un conjunto de entrenamiento y un vector aleatorio específico (θ_k), lo que resulta en un clasificador donde $h(x, \theta_k)$ es un vector de entrada. Como se menciona anteriormente, se basa en una muestra que entra al árbol y se somete a una serie de pruebas binarias en cada nodo, llamadas divisiones o *split*, hasta llegar a una hoja que contiene la respuesta. En la etapa de entrenamiento, el algoritmo optimiza los parámetros de las funciones de división a partir de las muestras de entrenamiento. Esta técnica puede utilizarse para dividir un problema complejo en problemas más simples.

4.5. Regresión Logística

La técnica de Regresión Logística o *Logistic Regression* (LR) busca estimar la probabilidad de un evento, en el caso de la clasificación, busca predecir si un elemento pertenece a cierta clase. Las variables de entrada deben ser independientes, discretas o continuas, mientras que la variable dependiente es una probabilidad que se convierte a una variable binaria mediante un umbral o corte. Esta técnica requiere que los residuos sean normales y que la varianza sea constante en el tiempo. Así, el modelo se entrena entregándole datos de *inputs* y una clase binaria a la cual estos pertenecen. Una vez entrenado, el modelo toma la forma de una ecuación de regresión logística:

$$y = \frac{1}{1 + \exp\{-(w_o + w_1 * x)\}} \quad (4.4)$$

Donde y corresponde a probabilidad de pertenecer a una clase, $\frac{1}{1 + \exp\{-z\}}$ es la función sigmoideal y $w_o + w_1 * x$ es el modelo lineal dentro de la regresión logística. La función sigmoideal toma cualquier valor real y entrega un valor entre 0 y 1. Por otro lado, el modelo lineal representa una relación lineal entre las características del *input* y el *output* estimado, donde x es la variable *input* o independiente, w_1 es el peso de esta y w_o es el error asociado. Una vez calculado el valor de y , este se clasifica en 0 o 1 dependiendo de un umbral, por ejemplo, si la probabilidad estimada de y es menor que 0.5 se clasifica como 0 y si es mayor, se clasifica como 1 (Keboola, 2020).

El objetivo de entrenar el modelo es determinar los mejores pesos para el modelo lineal dentro de la regresión logística, donde estos pesos se calculan optimizando la función de costos *LogLoss*, buscando que sea lo más bajo posible. Para lo anterior, se puede utilizar el método de descenso de gradiente, que ajusta los pesos de un modelo según la función de pérdida de cada punto de datos. Primero, se calcula la función de costos ($J(\theta)$) en cada punto de datos y se obtiene la pendiente de $J(\theta)$ con respecto a los pesos y el sesgo. Luego, se actualizan los valores de los pesos y el sesgo en función de la pendiente, y se repite el proceso hasta que se alcance un error mínimo. Por otro lado, la tasa de aprendizaje (*learning rate*) controla la rapidez con la que se alcanza el mínimo óptimo, pero debe elegirse con cuidado para evitar que sea demasiado alta o demasiado baja (Keboola, 2020).

Consideraciones sobre LR:

- Entrenar el modelo y usarlo para predicciones es muy simple y no requiere muchos gastos de ingeniería para el mantenimiento.
- A diferencia de los modelos de aprendizaje profundo (redes neuronales), la regresión logística es sencilla de interpretar. Aunque no es tan interpretable como la regresión lineal, la regresión logística puede ayudarnos a evaluar qué variable de entrada es responsable del mayor cambio en el valor predicho.
- Los tiempos de entrenamiento rápidos combinados con bajos requisitos computacionales hacen que la regresión logística sea fácil de escalar, incluso cuando el volumen y la velocidad de los datos aumentan.
- Se puede utilizar en entornos en línea, lo que significa que el modelo se puede volver a entrenar con cada nuevo ejemplo y generar predicciones casi en tiempo real. Esto contrasta con los enfoques computacionalmente pesados, como las redes neuronales o las máquinas de vectores de soporte, que requieren una gran cantidad de recursos informáticos o largos tiempos de espera mientras se reentrenan nuevos datos.

4.6. Máquina de vectores de soporte

La técnica de Máquina de vectores de soporte o *Support Vector Machine* (SVM) es un algoritmo de aprendizaje supervisado que se puede utilizar para clasificación o regresión. En el contexto de la clasificación binaria, SVM construye un hiperplano en un espacio de dimensionalidad alta o incluso infinita, con el objetivo de separar de manera óptima todas las clases. La característica fundamental de SVM radica en buscar el hiperplano que tenga la máxima distancia con los puntos más cercanos al mismo. Estos puntos de datos más cercanos, llamados vectores de soporte, desempeñan un papel vital en la determinación de la posición y orientación de este hiperplano. Además de su capacidad para manejar datos linealmente separables, SVM puede manejar eficazmente datos no lineales mediante el uso del "truco del kernel", que consiste en mapear los datos a un espacio de mayor dimensión donde la separación lineal es posible. Esto les permite abordar problemas de clasificación más complejos y mejorar su capacidad de generalización.

Consideraciones sobre SVM (Betancourt, 2005):

- Son principalmente clasificadores para 2 clases, pero se pueden adaptar para clasificación multiclase.
- Son relativamente fáciles de entrenar y no tienen óptimos locales como las redes neuronales.

- Se escalan bien para datos en espacios dimensionales altos y el compromiso entre la complejidad del clasificador y el error puede ser controlado explícitamente.
- Se necesita una buena función kernel y metodologías eficientes para sintonizar los parámetros de inicialización de SVM.

4.7. Redes neuronales

La técnica de redes neuronales o *Neural Networks* (NN) se basa en procesar los datos de una manera que está inspirada en la forma en que lo hace el cerebro humano, utilizando nodos o neuronas interconectadas en una estructura de capas, similar a una red neuronal biológica, que trabajan juntas para resolver un problema (AWS, s.f.). Estas redes pueden tener diferentes topologías, como las *single-layer feedforward*, *multilayer feedforward* y *recurrent*. La topología *multilayer feedforward* o perceptrón multicapa (figura 4.4), que consta de capas de entrada, capas ocultas y capa de salida interconectadas, solo admiten como entradas las salidas de las neuronas de la capa anterior, a diferencia de las *recurrent* que presentan lazos de retroalimentación entre neuronas. La salida de cada nodo es la suma ponderada de todos sus nodos de entrada más un sesgo. Además, se utilizan funciones de activación para seleccionar qué neuronas se activan y se transmiten a la siguiente capa oculta. Dentro de las funciones de activación se tienen las de tipo lineal y no lineal (umbral, sigmoide, tangente hiperbólica, reLu, softmax). Con respecto a la clasificación final, se determina por la clase con la probabilidad más alta (salida). El clasificador de NN se entrena de manera supervisada a través de múltiples iteraciones basadas en algoritmos de retropropagación. En cada iteración, se calcula el error entre las etiquetas verdaderas y la predicción, que luego se retropropaga a través de las capas para ajustar los pesos de cada capa y mejorar la salida (Mokari et al., 2023). En teoría, cuanto más datos se suministren a una red neuronal, mayor será su potencial predictivo, sin embargo, las redes neuronales a veces requieren un gran volumen de datos de entrada para lograr una alta precisión de predicción.

Se distinguen tres tipos de capas:

- Capa de entrada: corresponde a la vía de entrada de la información externa a la red neuronal. Los nodos de entrada procesan los datos, los analizan o los clasifican y los pasan a la siguiente capa.
- Capas ocultas: analizan la salida de la capa anterior, la procesan aún más y la pasan a la siguiente capa.
- Capa de salida: proporciona el resultado final de todo el procesamiento de datos realizado por la red neuronal artificial. Puede tener uno (clasificación binaria) o varios nodos de salida (clasificación multiclase).

Debido a su constitución y a sus fundamentos, las redes neuronales artificiales presentan características semejantes a las del cerebro. Por ejemplo, utilizan aprendizaje profun-

do para aprender de los errores y mejorar continuamente, son capaces de aprender de la experiencia, de generalizar casos anteriores a nuevos casos y de abstraer características esenciales a partir de entradas que representan información relevante (Matich, 2001). Más aún, dentro de sus aplicaciones más importantes destacan la visión artificial, reconocimiento de voz, procesamiento de lenguaje natural y motores de recomendaciones (Amazon Web Services, s.f.).

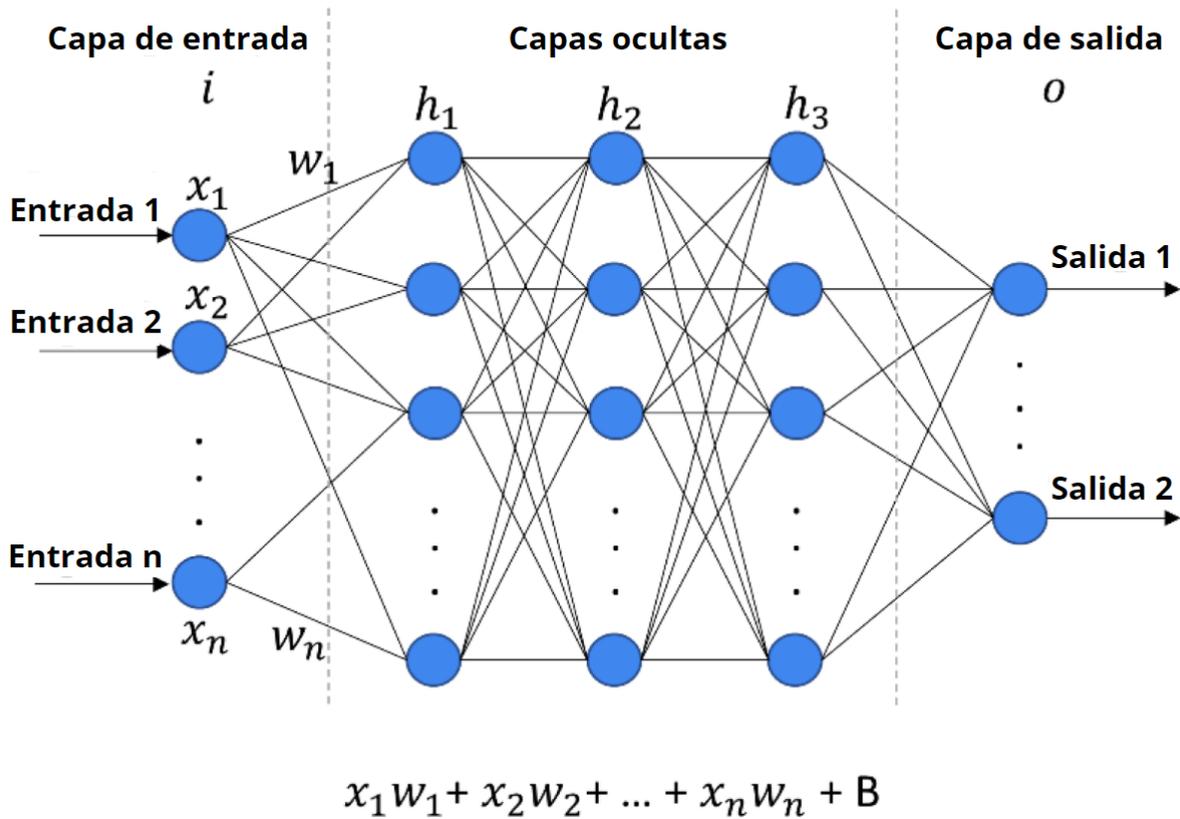


Figura 4.4: Representación de una red neuronal artificial *multilayer feed-forward* con tres capas: entrada, ocultas y salida. Modificado de Mokari et al. (2023).

CAPÍTULO 5

GEOLOGÍA REGIONAL

Este capítulo condensa información bibliográfica relacionada a la caracterización geológica de la zona de estudio, incluye estratigrafía, geología estructural, geotectónica, marco metalogénico y geomorfología.

5.1. Generalidades

El distrito Río Blanco-Los Bronces se ubica en la franja metalogénica del Mioceno tardío a Plioceno temprano de Chile central, al sur de la franja del Neógeno. Su localización coincide con la transición del *flat-slab* a una subducción normal. Esta transición es concordante con la subducción de la dorsal de Juan Fernández, que coincide con los mega eventos de mineralización en El Teniente (Skewes & Stern, 1994).

En el sector afloran rocas andesíticas pertenecientes a las formaciones Abanico y Farrellones de edad oligocena a miocena medio, instruidas por rocas plutónicas pertenecientes al batolito San Francisco (mioceno inferior a medio), cuerpos menores subvolcánicos y complejos de brechas hidrotermales de edad Plioceno inferior (Serrano et al., 1996).

5.2. Estratigrafía

En esta sección se presenta una breve reseña de las formaciones de las rocas estratificadas en la zona de estudio.

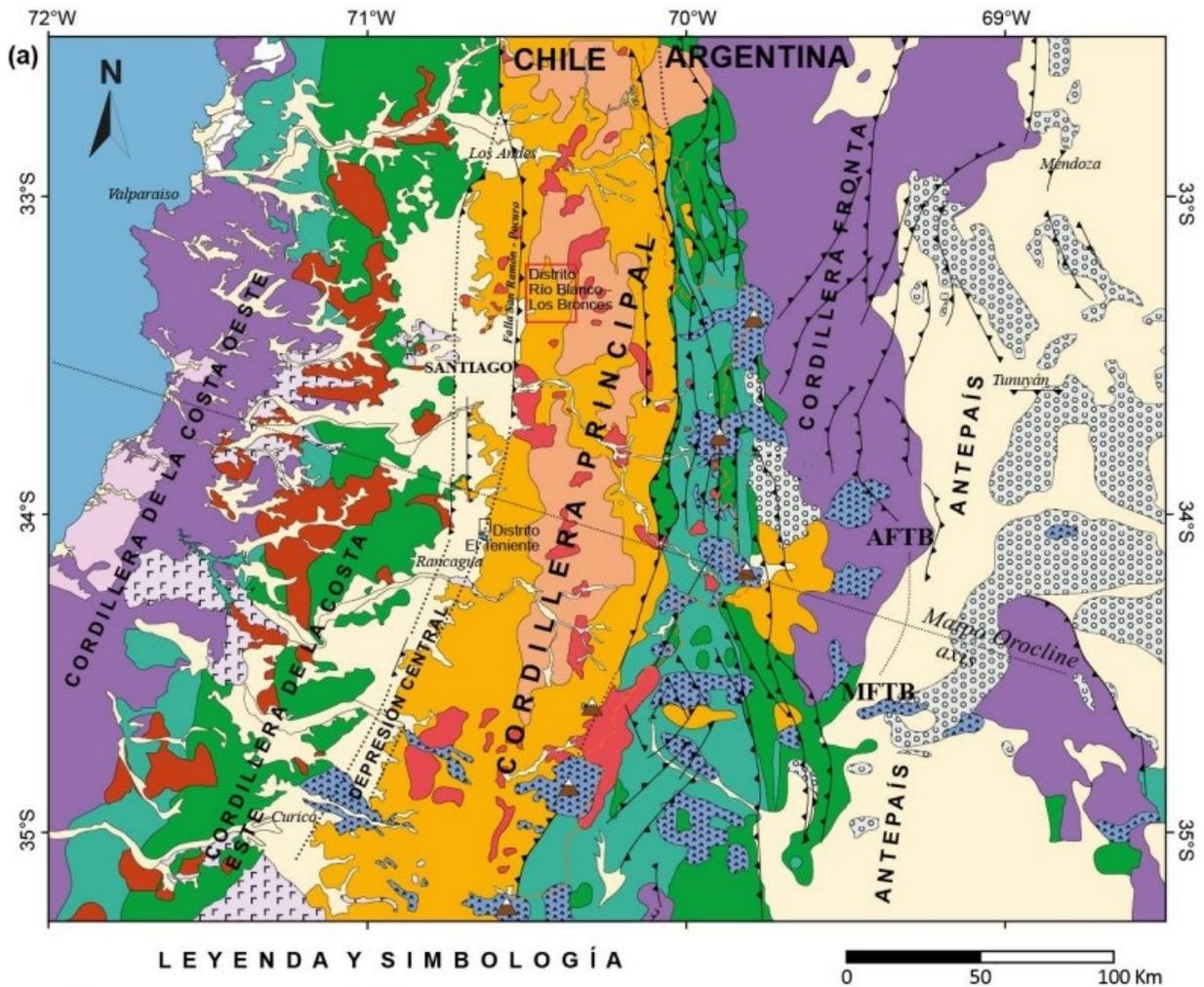


Figura 5.1: Mapa geológico regional de Chile central. En un recuadro color rojo se visualiza la ubicación del distrito Río Blanco - Los Bronces, junto con las formaciones Abanico y Farellones, unidades intrusivas y principales fallas activas durante el Neógeno. Modificado de Rojas (2022), basado en Farías et al., 2010, donde AFTB y MFTB corresponden a la faja plegada y corrida de Aconcagua y Malargüe, respectivamente.

5.2.1. Rocas estratificadas

5.2.1.1. Formación Abanico (Eoceno - Mioceno)

La formación fue definida por Aguirre (1960) como una secuencia formada por tobas y brechas volcánicas de colores principalmente violáceos, púrpura y gris, con un espesor

aproximado de 3000 m. Posee intercalaciones de lavas andesíticas y riolíticas, además de sedimentitas volcanoclásticas consistentes en areniscas de grano fino a medio, lutitas y limos finamente estratificados, con flora fósil escasa en relación al resto de la secuencia (Thiele, 1980). Está expuesta entre 32.5° y más al sur de 36°S. Más aún, entre 33 y 35°S se dispone en dos franjas NS paralelas en la Cordillera Principal, separadas por la Formación Farellones (figura 5.1).

Le subyace en concordancia la Formación Colimapu y le sobreyace la Formación Farellones en discordancia angular (Thiele, 1980). De acuerdo a diversas dataciones radiométricas, se acota su edad entre 36 Ma en los depósitos cenozoicos de Las Termas del Flaco en el valle del Río Tinguiririca (Charrier et al., 1996) y 16.1 Ma en las lavas altamente deformadas en el valle del Río Las Leñas (Kay & Kurtz, 1995). Así, considerando las mínimas edades establecidas, la Formación Abanico se interpreta como producto de varios eventos volcánicos o uno continuo de largo plazo, entre el eoceno medio a tardío al mioceno temprano tardío.

5.2.1.2. Formación Farellones (Mioceno inferior - medio)

La Formación Farellones fue definida por Klön (1960) como una secuencia de lavas, tobas e ignimbritas con intercalaciones de brechas, siendo las lavas predominantes. Además, posee alternancia de rocas volcanoclásticas más finas, lo que produce una marcada estratificación diferenciable de la Formación Abanico. Las tobas presentes en la formación son de color gris claro y presentan clastos de riolitas y andesitas, mientras las ignimbritas son de color gris claro y poseen clastos de andesitas y obsidiana. Localmente, las rocas se presentan muy alteradas a arcilla, limolita, hematita y sílice. El espesor de la formación se estima en 2500 m, aumentando por diversos mantos intrusivos y lacolitos que se intercalan en la secuencia. Más aún, Rivano et al. (1990) redefinen la formación, describiendo dos miembros, un miembro inferior formado por tobas e ignimbritas riolíticas con poca meteorización, con intercalación de lavas delgadas y se encuentra sobre la formación Abanico en discordancia angular; mientras que el miembro superior está compuesto por lavas andesíticas y basálticas, con tobas y aglomerados andesíticos intercalados.

Entre 32 y 35°S, se distribuye como una franja NS en la Cordillera Principal (Vergara et al., 1988), manteniendo una inclinación máxima de 25° (figura 5.1). Su límite inferior está marcado por la discordancia angular que la separa de la Formación Abanico (Thiele, 1980), mientras que su techo es levemente discordante con la Formación Colorado-La Parva, aunque gran parte de su techo corresponde a la actual superficie de erosión. No posee fósiles para acotar su edad. Sin embargo, dataciones de U/Pb en circones del sector Río Blanco, con rangos de edad entre 17.20 ± 0.05 Ma y 16.77 ± 0.25 Ma (Deckart et al. 2005), permiten asignarle una edad miocena inferior a medio.

5.2.2. Rocas intrusivas

5.2.2.1. Intrusivos Plutónicos (Mioceno)

Representados por el Batolito San Francisco, corresponde a un conjunto de intrusivos que afloran en un área de 200 km² (Warnaars et al., 1985) y se disponen como una franja NS limitada por fallas de desplazamiento dextral (Carrizo et al., 2012; Piquer, 2015). Esta unidad fue definida como Unidad I por Thiele (1980), plutón Disputada por Cepeda (1974), Plutón Río Blanco-San Francisco por Stambuk et al. (1982) y, posteriormente, como batolito Río Blanco-San Francisco por Piquer (2015).

Sus tipos litológicos corresponden a diorita cuarcífera, granodiorita, monzonita cuarcífera, monzodiorita cuarcífera, granitos y sienitas (Serrano et al., 1996; Warnaars et al., 1985), con texturas que varían de porfídica a faneríticas. Es fuertemente peraluminosa y con composición calcoalcalina (Warnaars et al. 1985). Más aún, la unidad intruye a la Formación Farellones y subordinadamente a la Formación Abanico. En cuanto a su edad, se atribuye al Mioceno con base en edades que varían entre 20.1 y 8.4 Ma para rocas pertenecientes a la unidad (Warnaars et al., 1985; Serrano et al., 1996; Deckart et al., 2005) .

5.2.2.2. Intrusivos hipabisales a subvolcánicos (Mioceno Superior - Plioceno Inferior)

Complejo subvolcánico de composición esencialmente dacítica (Serrano et al., 1996), cuyas rocas se ubican en contacto con plutones más antiguos y brechas mineralizadas. Corresponden a pórfidos dacíticos y cuerpos subvolcánicos dacíticos, pórfidos andesíticos, domos riolíticos y diatremas, y domos dacíticos. Con respecto a su edad, dataciones de U/Pb en circones indican edades entre 7.12 ± 0.19 y 4.31 ± 0.05 Ma (Deckart et al., 2012), lo que permite asignarle una edad miocena superior a pliocena inferior.

5.2.2.3. Brechas hidrotermales (Mioceno Superior - Plioceno Inferior)

En el distrito Río Blanco-Los Bronces se han reconocido una serie de brechas hidrotermales, caracterizadas según sus minerales, tipo, textura y mineralización presente (Serrano et al., 1996). Se distinguen tres tipos de brechas principales: brechas de turmalina, brechas con matriz de polvo de roca y brechas ígneas.

Las brechas de turmalina se caracterizan por su color gris medio, dado por la presencia de polvo de roca y turmalina fina. Los clastos son subredondeados, variando de milimétricos a métricos, e incluyen fragmentos de andesitas y pórfidos granodioríticos-dacíticos alterados a sericita. La matriz esta constituida por polvo de roca y minerales hidrotermales que constituyen la secuencia paragenética: turmalina, especularita, menor cuarzo, calcopirita, pirita, carbonatos (dolomita-ankerita) y anhidrita. El cuerpo brechoso es cortado por vetas tipo D y vetillas rellenas de pirita, calcopirita y molibdenita (Hémera, 2019).

Por otro lado, las brechas de polvo de roca presentan fragmentos de monzonita cuarcífera de grano fino junto con mineralización de tipo pórfido cuprífero diseminada, en una

matriz de polvo de roca. Comúnmente, tanto los fragmentos como la matriz, se presentan alterados homogéneamente a cuarzo sericita (Warnaars, 1985). En cuanto a las brechas ígneas, originadas por la intrusión del complejo de pórfidos Paloma y Sulfatos, presentan clastos de pórfidos granodioríticos y dacíticos, y en menor medida clastos de andesitas muy alteradas, sub angulosos en una matriz aplítica. Estas brechas presentan zonas profundas de intensa alteración biotítica, con presencia de biotita, cuarzo, feldespato potásico y albita, con intensa disseminación de bornita. En algunas porciones altas, la alteración sericítica es dominante y está acompañada de clorita, cuarzo, turmalina, anhidrita y especlarita. Finalmente, se han obtenido edades entre 7.4 y 4.2 Ma que permiten asignarles a estas brechas edades miocena superior a pliocena inferior (Serrano et al., 1996).

5.2.3. Depósitos no consolidados

5.2.3.1. Depósitos glaciares (Pleistoceno - Holoceno)

Depósitos morrénicos con morfologías variadas asociadas a morrenas laterales, frontales o de fondo. Estos depósitos son polimícticos, macizos, mal seleccionados y con porcentajes entre 20-50 % de matriz. Se componen de fragmentos mayores (0.8 a 2.5 m), comúnmente redondeados y con estrías glaciares; y fragmentos medios (0.01 a 0.8 m), angulosos a subangulosos; inmersos en una matriz tamaño arena a limo (Hémera, 2019).

5.2.3.2. Depósitos aluviales y aluvio-coluviales (Holoceno)

Depósitos recientes, mal seleccionados, macizos a incipientemente estratificados, constituidos por fragmentos tamaño bloque a limo. Estos depósitos se disponen como relleno de quebradas, principalmente en los valles de los esteros Yerba Loca y Cepo, cubriendo áreas de hasta 210,000 m². Morfológicamente presentan superficies suavizadas con pendientes inferiores a 15° en sus facies proximales y entre 5 y 8° en las facies distales. Los depósitos de relleno de fondo de valle alcanzan extensiones de hasta 200 m de largo, mientras que los de morfología de abanico alcanzan 300 m desde el pie al ápice (Hémera, 2019).

5.2.3.3. Depósitos coluviales (Pleistoceno - Holoceno)

Depósitos de origen gravitacional, presentan conos de deyección que pueden extenderse hasta 700 m. Son monomícticos, con estratificación y pendientes de hasta 35°. Son clastosportados y formados por fragmentos angulosos a subangulosos, con tamaño variable desde bloque hasta arena gruesa.

5.2.3.4. Depósitos de flujo de detritos (Holoceno)

Depósitos macizos, de muy mala selección, localmente clastosportados y polimícticos, con fragmentos mayores angulosos a subangulosos, inmersos en una matriz fina de tamaño arena y limo, con menor arcilla. Estos depósitos se presentan con morfologías superficiales lobuladas levées y frentes abruptos suavizados por erosión, parcialmente

retrabajados y/o disectados por cursos actuales de ríos. Se ubican como rellenos de valles en el sector de Los Bronces, cubriendo áreas de hasta 280,000 m² y presentando extensiones que superan los 1.5 km de largo (Hémera, 2019).

5.2.3.5. Depósitos antrópicos (Holoceno - actual)

Corresponde a todo tipo de modificaciones a la superficie natural del terreno como consecuencia de la actividad humana o reciente. En el sector Los Bronces están ampliamente distribuidos, cubriendo un área total de aprox. 660 km². Los depósitos se asocian a las faenas de explotación de los yacimientos de cobre Río Blanco y Los Bronces.

5.3. Geología Estructural

En los Andes de Chile Central y Argentina (32-35°S), se identifican dos dominios estructurales principales, divididos por fallas de rumbo NS a NW como Río Olivares, El Fierro y Altos de Juncal, que marcan la separación entre rocas sedimentarias marinas y continentales altamente deformadas del Jurásico al Cretácico temprano en el dominio oriental, y las Formaciones Abanico y Farellones, que datan del Eoceno al Plioceno en el dominio occidental, donde se sitúa el distrito Río Blanco-Los Bronces (Piquer, 2015; Ramos, 1996; Charrier, 2002). Este último está definido por fallas inversas al este y limitado al oeste por la falla San Ramón y sus equivalentes, presentando una serie de pliegues asimétricos y estructuras orientadas principalmente norte-sur a noreste (Charrier, 2002; Fock, 2005; Piquer & Skarmeta, 2012).

Dentro del distrito, se destacan estructuras específicas como las fallas de orientación NS (Falla Río Olivares, El Quempo y El Fierro-Altos de Juncal) y las estructuras de orientación NE (Fallas Saladillo, El Salto y Flores), estas últimas muestran un movimiento predominantemente dextral con componente inversa menor (figura 5.2). Si bien las fallas regionales de orientación NS no afloran en el distrito, son ciegas y actúan deformando a las rocas de las formaciones Abanico y Farellones (Carrizo et al., 2012; Piquer, 2015).

En cuanto a la deformación superficial en la zona, se caracteriza por la presencia de estructuras frágiles de corta extensión y corresponden a vetas, vetas falla y fallas menores, estratificación y fallas intracutáneas. No superan los 6 metros de espesor de relleno hidrotermal y los 4 km de persistencia. Asimismo, la cinemática de estas indica tensión y movimientos de rumbo, dominando estructuras dextrales de orientación NE-SW (Carrizo et al., 2014).

5.4. Geotectónica

La zona del distrito Río Blanco-Los Bronces se ubica en la Cuenca del Abanico, en la Cordillera Principal de Chile Central entre los 33° y 34°S. Esta región refleja la transición de un régimen tectónico extensional a uno compresivo, marcando la evolución desde la

formación de cuencas extensionales durante el Oligoceno hasta su inversión tectónica en el Mioceno. Dicha inversión contribuyó a la formación del relieve actual que delimita la Depresión Central con el frente cordillerano occidental (Charrier et al., 2005).

Con respecto a la Cuenca del Abanico (CA), esta limita al este con la Faja Plegada y Corrida del Aconcagua (AFTB en figura 5.1) mediante la falla inversa El Coironal y la falla Laguna Negra (Ramos et al., 1991, Fock; 2005), mientras que hacia el oeste está delimitada por el sistema de falla San Ramón (Fock, 2005; Charrier et al., 2009). Por otro lado, las fallas principales de CA, de orientación predominantemente norte-sur, con variaciones hacia NNE y NNW, manifiestan un patrón de deformación compresiva bivergente que ha afectado a la Formación Abanico en ambos flancos del dominio (Fock, 2006; Charrier et al., 2005).

En cuanto a la evolución tectónica del dominio CA en el distrito Río Blanco-Los Bronces, se destaca el control ejercido por un sistema de fallas oblicuas pre-mineral de orientación NW-NNW y NE. Estas estructuras, que inicialmente se formaron como fallas normales durante un periodo de tectónica extensional, experimentaron una fase de reactivación durante un subsiguiente episodio compresivo. Dicha fase de reactivación sería contemporánea a la depositación de la formación Farellones y el emplazamiento del Batolito San Francisco junto a la serie de pórfidos asociados el sistema Río Blanco-Los Bronces (Piquer & Skarmeta, 2012).

5.5. Marco Metalogénico

El distrito Río Blanco-Los Bronces se localiza en el sector sur de la franja metalogénica del Mioceno al Plioceno inferior, la que se extiende desde el suroeste de Colombia hasta el centro de Chile y Argentina, abarcando alrededor de 6000 km. En particular, el distrito se encuentra en la sub-franja de Chile Central, que se extiende por aproximadamente 400 km a lo largo de la Cordillera Principal, entre los 32° y 35°S (Sillitoe & Perelló, 2005) (figura 5.3).

La ubicación geográfica y geológica del distrito sobre una zona de subducción normal, al sur de los 33°S, lo distingue de la parte norte de la sub-franja que se caracteriza por una subducción plana (*flat-slab*) sin volcanismo activo (Stern, 2004; Yáñez et al., 2002). La transición entre estos dos regímenes tectónicos, marcada por la influencia de la subducción de la Dorsal de Juan Fernández, ha sido determinante en la génesis y desarrollo del distrito, especialmente entre los 6 y 3 Ma, un período de intensa actividad magmático-hidrotermal y rápida exhumación que condujo a la formación de los yacimientos de Río Blanco-Los Bronces y El Teniente (Maksaev et al., 2009). Más aún, los yacimientos de Los Pelambres, El Teniente y Río Blanco-Los Bronces son pilares de esta sub-franja, con recursos para la franja que superan los 300 Mt de cobre fino (Toro et al., 2009). En el caso del distrito Río Blanco-Los Bronces, descubrimientos como Los Sulfatos y San Enrique Monolito, permiten estimar recursos de más de 200 Mt de cobre fino contenido (Toro et al., 2012).

5.6. Geomorfología

En la región central de Chile, la geomorfología se distingue por la secuencia de la Planicie Litoral, la Cordillera de la Costa, los Valles Centrales, la Precordillera y la Cordillera Principal, dispuestas predominantemente en orientación norte-sur como fajas elongadas, con excepción de los valles que se encuentran separados por cordones montañosos este-oeste que bajan desde la precordillera (Barros, 2007). El área de Los Bronces, situada en la Cordillera Principal Occidental, se caracteriza por su topografía accidentada, con valles profundos en forma de U y escarpadas laderas, resultado de la erosión glacial, y donde actualmente predominan los procesos erosivos por agua y las variaciones térmicas. Este relieve abrupto ha sido esculpido en gran parte por la retirada de las masas de hielo, con diferencias de cota que pueden alcanzar los 500 m entre valles y cumbres (Vargas, 1985).

En particular, la mina Los Bronces se ubica dentro de la Quebrada La Disputada, parte de la subcuenca del Río San Francisco, que pertenece a la subcuenca Mapocho Alto y que, a su vez, está dentro de la cuenca del Río Maipo (Demierres, 2021). Con respecto a los glaciares presentes en Quebrada La Disputada, se tienen los glaciares rocosos Infernillo, Agua Los Machos y La Perla, situados entre las cotas 3,600 a 4,400 m.s.n.m. Cabe mencionar que los glaciares rocosos se definen como *"Glaciares compuestos predominantemente por fragmentos de roca, sedimentos o materiales finos e intersticialmente por hielo, mostrando evidencias de flujo presente o pasado"* (Demierres, 2021). En relación con la topografía en Los Bronces, esta varía desde 3,200 m.s.n.m hasta 4,800 m.s.n.m, con una media aproximada de 3700 m.s.n.m, lo que refleja el relieve montañoso y las condiciones periglaciares dominantes.

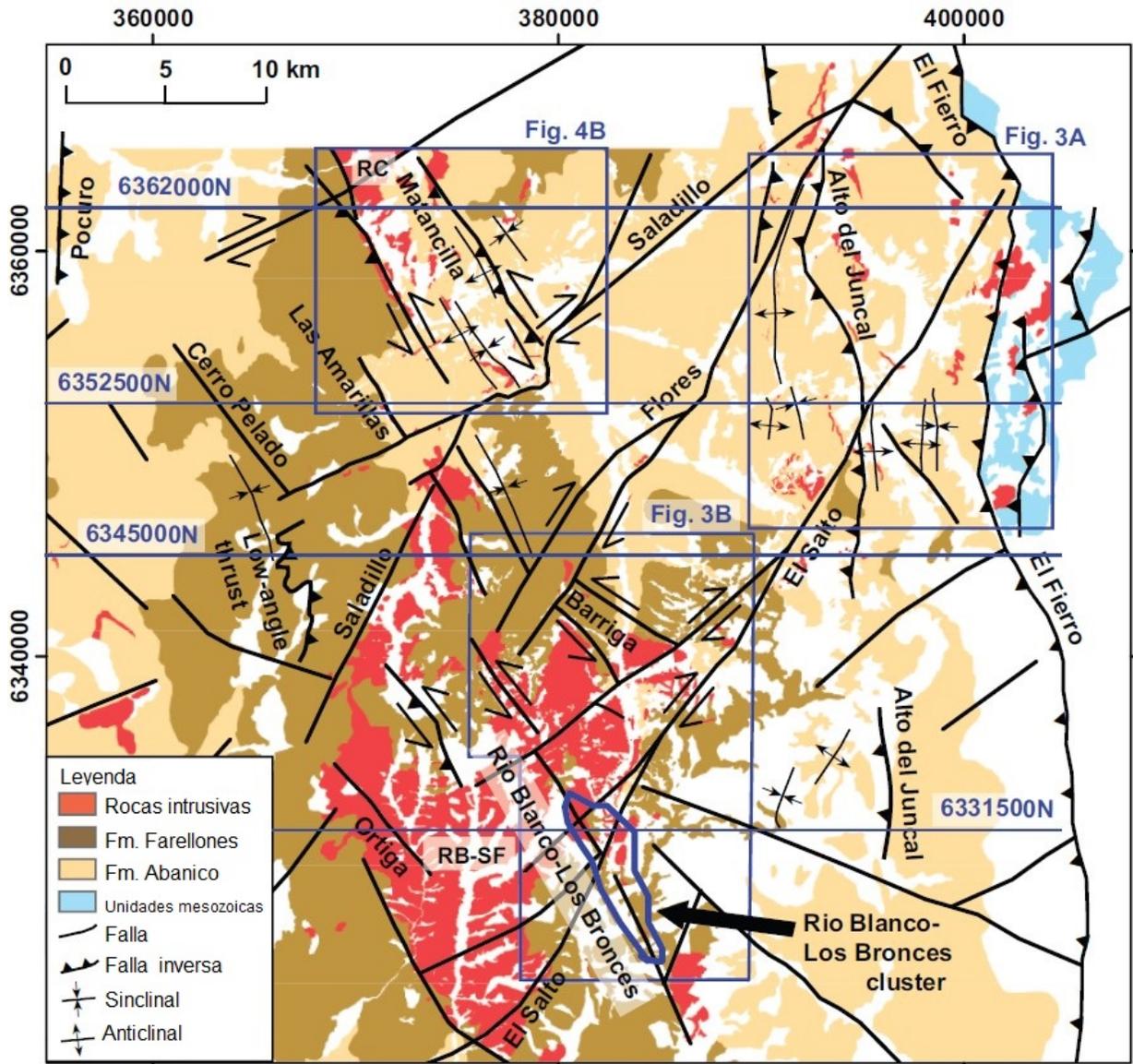


Figura 5.2: Principales Estructuras Regionales en el Distrito Río Blanco-Los Bronces. En el mapa se observan las tendencias de sistemas de fallas principales y pliegues, donde RB-SF y RC corresponden a abreviaciones de Batolito San Francisco-Río Blanco y Complejo plutónico Río Colorado, respectivamente. Modificado de Piquer, 2015.

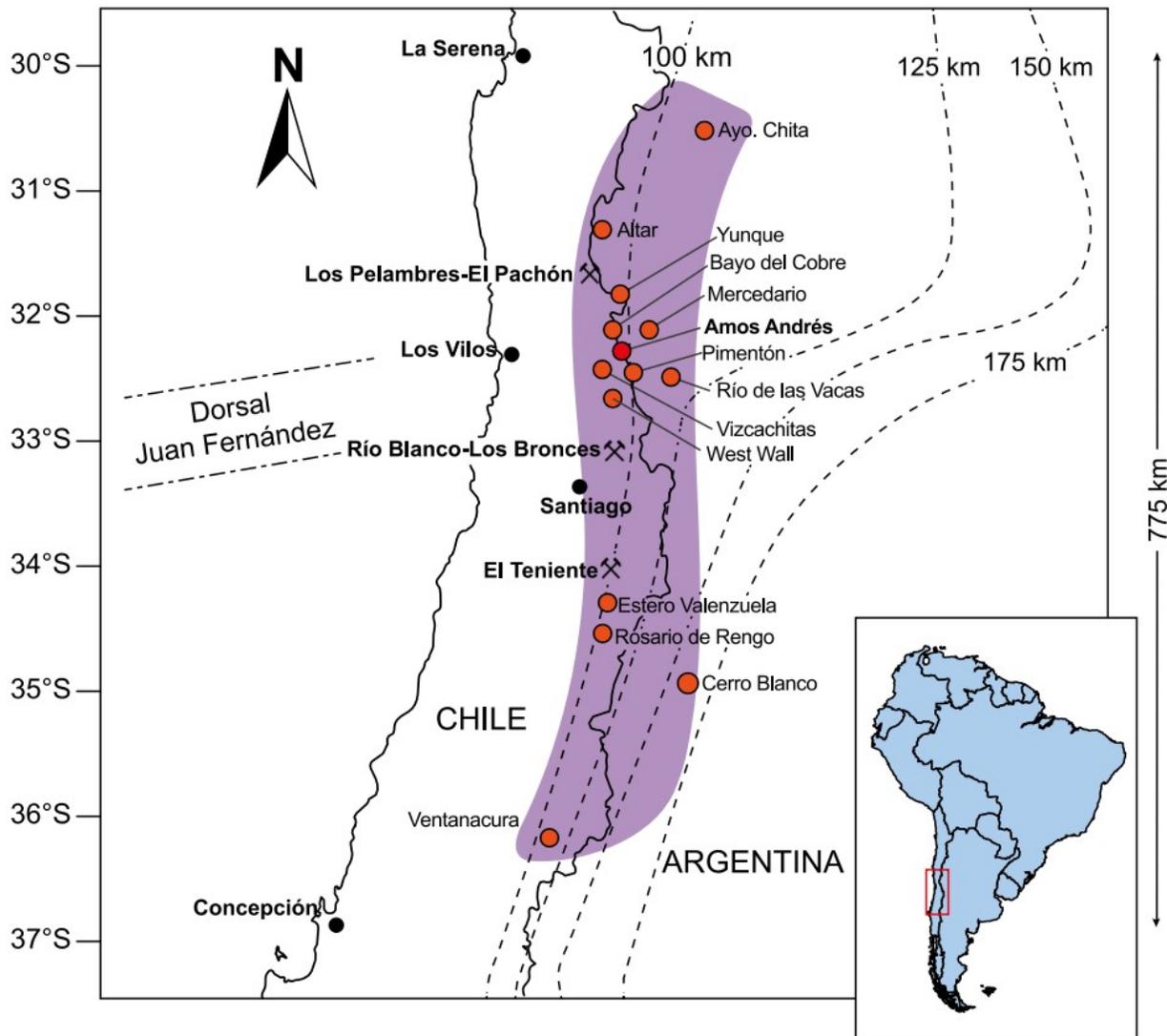


Figura 5.3: Mapa de los Andes de Chile central, La franja violeta indica la sub-franja de pórfidos cupríferos de Farellones, donde se distribuyen las minas y prospectos mineros de edades Mioceno Medio-Plioceno Inferior. En líneas grises segmentadas se muestra la posición de la Dorsal de Juan Fernández y los contornos de profundidad de la placa oceánica. Extraído de Rojas (2022), modificado de Cortés (2011) y basado en Gígola (2004) y Yañez et al. (2002).

CAPÍTULO 6

GEOLOGÍA DEL YACIMIENTO

6.1. Generalidades

El sistema porfídico Río Blanco-Los Bronces se extiende aproximadamente por 5 km desde Río Blanco y Los Bronces en el norte, pasando por Don Luis, Sur Sur, La Americana y Los Sulfatos en el sur, ubicándose al lado este del Batolito San Francisco, incluyendo centros hidrotermales y al menos cinco cuerpos sub-volcánicos (Deckart et al., 2013). Las rocas más antiguas del yacimiento corresponden a secuencias volcánicas y volcano-sedimentarias de las formaciones Abanico y Farellones, emplazadas en unidades intrusivas por el lado oriental del Batolito San Francisco durante el Mioceno. Luego, sobre estas unidades se generan una serie de brechas de origen hidrotermal y magmático. Finalmente, un conjunto de cuerpos porfídicos intruyen las unidades preexistentes durante el Mioceno Superior-Plioceno (Serrano et al., 1996). Por último, cabe mencionar que, localmente, se pueden clasificar las litologías dependiendo de su período de formación con respecto al período estimado que dio lugar a la mineralización de mena en la mina, lo que deriva en la clasificación pre-mineral (Batolito San Francisco y las andesitas de la formación Abanico y Farellones), intra-mineral (Pórfido Cuarzomonzonítico, sistema de brechas hidrotermales e ígneas y pórfidos tardíos) y post-mineral (complejo riolítico La Copa) (Abarzúa, 2014; AngloAmerican, 2021).

Por otro lado, el yacimiento exhibe, en general, las alteraciones propílica (prograda y retrograda), sericítica, fílica y potásica, donde el reemplazo de los minerales máficos por especularita y/o turmalina caracteriza a la zona propílica. También, los sulfuros de hierro, cobre y molibdeno abarcan cerca de 12 km², presentándose diseminados y en *stockworks* (Warnaars et al., 1985). Estas alteraciones fueron redefinidas localmente, teniendo alteración potásica, albita-turmalina-epidota, sericita-gris verde (o sericita), clorita-epidota, cuarzo-sericita y clorita-sericita-arcillas (AngloAmerican, 2023).

Finalmente, la formación de la mineralización tipo pórfido cuprífero, alteración y brechas de turmalina portadoras de cobre se asocia a un período de al menos 2.5 Ma entre los 7.4 y 4.9 Ma (Warnaars et al., 1985). En particular, a escala de distrito, el complejo principal de brechas está orientado N10° a N15°W, con una distancia de 9 km entre cuer-

pos de brecha mayores, Los Bronces y Los Sulfatos. A escala de mina Los Bronces, el contenido de mena se asocia principalmente a las brechas hidrotermales, las que crean un cuerpo elíptico de 2 km de largo, 8 km de ancho y 1 km de profundidad. Con respecto a la mineralización hipógena, se caracteriza principalmente por pirita, calcopirita, molibdenita, con menor bornita, donde 70 % de la mineralización económica se da en la matriz de las brechas, mientras que el 30 % restante se encuentra diseminado y en *stockworks*. Por otra parte, la mineralización supérgena se caracteriza por sulfuros de cobre secundarios, principalmente calcosina con menor covelina, donde el fracturamiento y la presencia de pirita en las rocas son las variables principales que controlan la ocurrencia de la zona supérgena, alcanzando profundidades de 700 metros bajo la superficie (AngloAmerican, 2021).

6.2. Litologías

Las litologías del yacimiento (figura 6.1), según el documento interno "Geología de Los Bronces" del departamento de Geología Mina (AngloAmerican, 2023), se dividen en 2 grandes grupos: Rocas ígneas (intrusivas y extrusivas) y Brechas.

6.2.1. Rocas ígneas intrusivas

6.2.1.1. Batolito San Francisco

El Batolito San Francisco es parte del conjunto de cuerpos intrusivos desarrollados en la región cordillerana de Chile Central y cuyas edades abarcan desde el Mioceno al Plioceno. Este cuerpo cubre un área de cerca de 200 km², 20 km en dirección N-S y 10 km en dirección E-W. Además, intruye a rocas de las formaciones Abanico y Farellones y, a su vez, es intruida por cuerpos porfídicos subvolcánicos de edades Mioceno Superior–Plioceno (Abarzúa, 2014). De acuerdo a Thiele (1980), estos cuerpos intrusivos se caracterizan por poseer poca extensión en superficie y una menor edad progresiva hacia el oriente.

- **Cuarzomonzodiorita (QMD):** Compuesta por cristales inequigranulares de plagioclasa (50 % total de minerales félsico), cuarzo (40 %) y feldespato potásico (10 %), con un contenido menor de biotitas secundarias y hornblenda, y un color preferentemente grisáceo. Además, puede presentar mayor contenido de máficos que las cuarzomonzonitas (QM), alcanzando aproximadamente el 20 % del total de la roca, destacando la mayor presencia de anfíbolos. Por otro lado, es afectada por todos los eventos de alteración, observándose en algunos casos una superposición de ellos. En particular, se observa una alteración propilitica dominante, expresada en el reemplazo selectivo de los máficos por la asociación clorita-epidota, esta alteración se superpone a una alteración potásica caracterizada por la presencia de biotita secundaria como reemplazo selectivo de los máficos. Otro rasgo de esta litología,

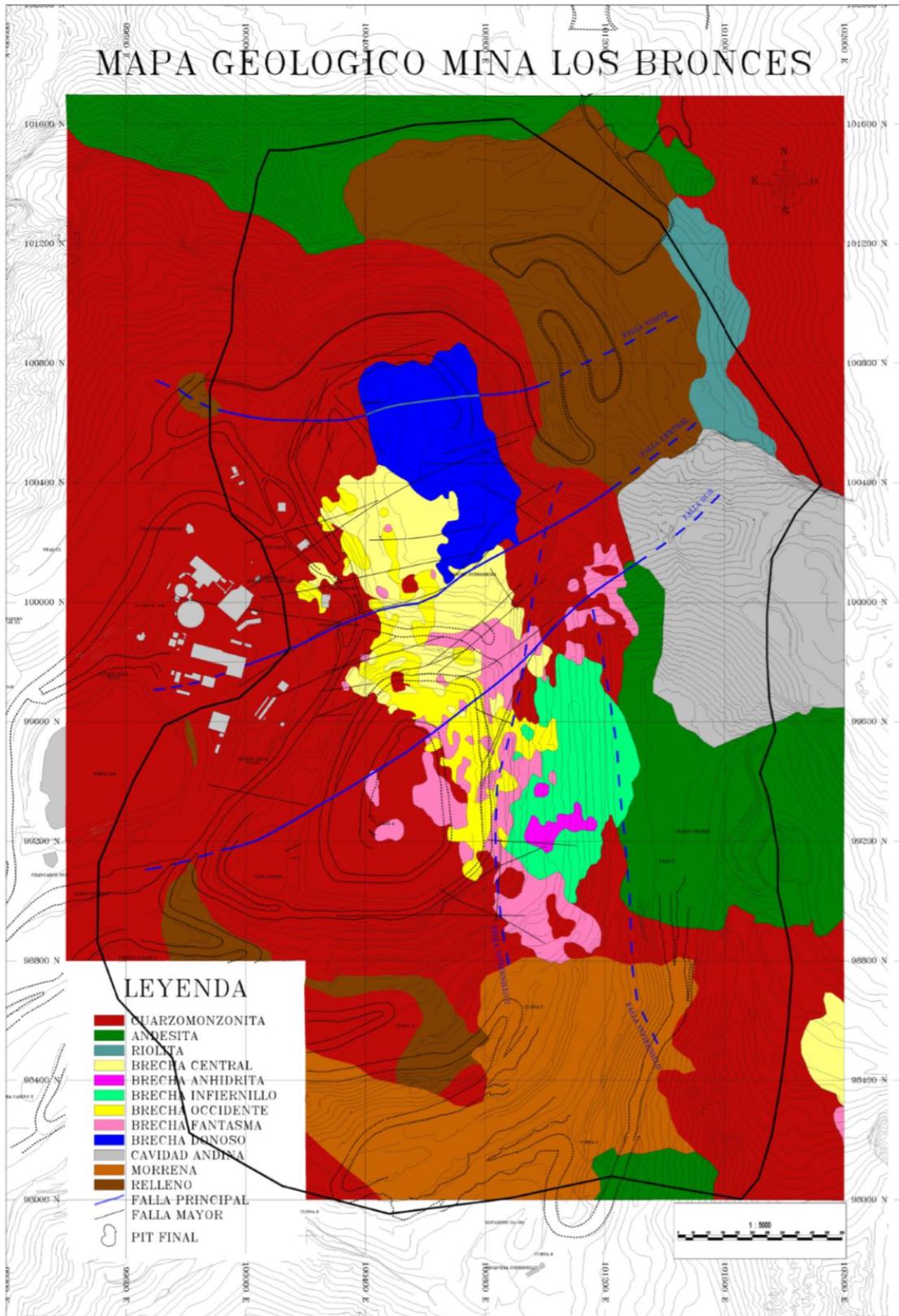


Figura 6.1: Mapa litológico de la mina Los Bronces. Extraído de Ortiz (2020), modificado por Vargas (2015) y basado en Abarzúa (2014).

es que se puede subdividir por el tamaño de sus granos en cuarzomonzodiorita fina (QMDF) y cuarzomonzodiorita gruesa (QMDG).

- **Diorita (DIO):** Compuesta casi sólo por plagioclasa, dado que el contenido de cuarzo y feldespatos alcalinos no superan el 5 % total de la roca. Además, los máficos más comunes a simple vista son hornblenda verde y biotita, tomando un color de gris a verde. Por otro lado, la textura es hipidiomórfica, grano fino y equigranular, pero el contenido de plagioclasas empañan la equigranularidad. En general, son cuerpos restringidos cortando intrusivos más félsicos y con una escasa mineralización, mientras que en profundidad son susceptibles a una alteración potásica rica en biotita.
- **Cuarzomonzonita (QM):** Compuesta por cristales de plagioclasas, cuarzo, hornblenda y feldespato potásico. Suele presentar un color rosado pálido, textura hipidiomórfica granular y fanerítica de grano grueso a fino. Más aún, al igual que QMD, puede verse afectada por más de un tipo de alteración. En reiteradas ocasiones, se observa la ocurrencia de biotita secundaria reemplazando máficos y un débil vetilleo A remanente, lo que indica la presencia de una alteración potásica moderada a débil. A la alteración anterior, se sobreimpone una alteración cuarzo-sericita y, a su vez, una alteración propilítica retrógrada, caracterizada por la cloritización de los máficos. Con respecto a su clasificación QAP de Streckeisen (1979), está compuesta por 50 % de feldespatos alcalinos, 30 % de feldespatos plagioclasa anhedral y 20 % de cuarzo anhedral (Abarzúa, 2014). Por último, se diferencia con QMD por el porcentaje mayor de feldespato potásico intersticial.
- **Pórfido Aplítico (PA):** Roca ígnea hipabisal leucocrática (de colores claros), caracterizada por una textura microcristalina granular de grano fino. Con respecto a su composición, la aplita se encuentra en el rango del granito-sienita-monzonita, por lo que el contenido de cristales de cuarzo varía de 20-60 %, plagioclasa de 10-90 % y feldespato alcalino de 90-10 %. Se caracteriza por su color rosa blanquecino, con contactos normalmente claros con la roca caja en la que cortan. No poseen mineralización significativa siendo estériles en la mayoría de ellos, no obstante, ocasionalmente se observan vetillas D cuyos halos de alteración son restringidos, serícíticos y de color gris.
- **Pórfido Cuarzomonzodiorítico (PQMD):** Intrusivo de grano medio, textura porfídica y color gris claro a blanco (Warnaars et al., 1985), presenta fenocristales (25 %) de plagioclasa, cuarzo, feldespato potásico y máficos, junto con una masa fundamental (75 %) compuesta por cuarzo, anhidrita y feldespato potásico. Por un lado, los minerales que componen la masa fundamental se encuentran parcialmente alterados a sericita, clorita, calcita y yeso, mientras que los fenocristales presentan una débil alteración potásica remanente con biotita, feldespato potásico y en menor grado anhidrita-biotita, evidenciada por el reemplazo de biotita secundaria sobre los máficos y la presencia de escasas vetillas A remanentes. También, se observa la so-

breimposición de una alteración fílica de sericita-cuarzo-pirita, de moderada a fuerte, junto a una alteración propilítica moderada, compuesta por clorita, montmorillonita, calcita y epidota (Abarzúa, 2014).

6.2.1.2. Pórfidos intra-mineral a tardi-mineral

- **Pórfido Dacítico (PDA):** Intrusivo de textura porfídica con fenocristales de plagioclasa (90 %), micro-fenocristales de biotita y hornblenda y masa fundamental de cuarzo, feldespato y biotita. Se reconocen como cuerpos acotados y de dimensiones cortas, texturalmente, los fenocristales son bien definidos, pueden presentar “ojos de cuarzo” y la masa fundamental es muy fina casi afanítica.
- **Pórfido Cuarzodiorítico (PQD):** Roca de textura porfídica, grano fino a medio y de color gris oscuro. Presenta fenocristales (30 %) de plagioclasa, biotita, anfíbol y magnetita, además de una masa fundamental (70 %) compuesta por plagioclasa, cuarzo, feldespato potásico y clorita. Con respecto a su mineralización, es escasa, presentándose como calcopirita diseminada y calcopirita-bornita en vetillas tipo A. Por otro lado, se observa una alteración potásica remanente e intensa de biotita secundaria, albita, anhidrita y feldespato potásico, sumado al desarrollo de vetillas tipo A. Además, exhibe una alteración propilítica moderada de clorita, biotita y montmorillonita, junto a una alteración fílica con parches de sericita, andalucita y cuarzo secundario (Contreras, 2023). Por último, se reconocen cuerpos restringidos cortando QM y QMD.
- **Pórfido Cuarzomonzonítico (PQM):** Roca de textura porfídica, holocristalina y de color gris verdoso intramineral. Presenta fenocristales de albita \pm ortoclosa subhedrales a anhedrales, más una masa fundamental compuesta por feldespatos potásico, escasas plagioclasas y cuarzo. En los lugares en que el PQM se emplaza genera un halo cuarzo-sericítico, caracterizado por tener alteración sílica y cuarzo-sericita, donde no se reconoce biotita hidrotermal (Serrano et al., 1996). Hay que mencionar que este pórfido aumenta su espesor en profundidad y se cree que se une a un stock cuarzomonzonítico en profundidad.
- **Pórfido Don Luis (PDL):** Pórfido riodacítico tardimineral con fenocristales de plagioclasa, cuarzo, feldespato potásico y menor contenido de biotita de grano fino, todo en una masa fina aplitica. La menor cantidad relativa de biotita se corresponde con un desarrollo débil de alteración y a los bajos contenidos de cobre (<0.5 %) dentro y alrededor del pórfido. El PDL es cortado por vetillas de cuarzo \pm sericita \pm anhidrita \pm sulfuros y vetillas de feldespato potásico, indicando que una menor actividad hidrotermal continuaba luego del emplazamiento del PDL (Frikken, 2003; Serrano et al., 1996). En terreno, se reconoce en contacto con brechas, débilmente alterado, con una escasa mineralización pirita dominante y “ojos de cuarzo” característicos.

- **Pórfido Granodiorítico (PGR):** Pórfido intramineral de color gris a rosado claro, textura porfídica desde *crowded* hasta con mayor desarrollo de masa fundamental, la que varía entre 30-50 % y esta compuesta principalmente por cuarzo, feldespato potásico y en menor cantidad plagioclasas. Al menos el 80 % de fenocristales de plagioclasa presenta un tamaño entre 3-6 mm, acompañado en menor medida de biotita con textura laminar, cuarzo y feldespatos. Más aún, suele presentar ojos de cuarzo de hasta 2 mm. Con respecto a las alteraciones, en algunas zonas presenta alteración potásica preservada o relictas debajo de alteración fílica posterior cuarzo-sericita o clorita-sericita-arcillas. En terreno, suelen confundirlo con una QM fina en la que no se presenta cuarzo, o bien un PA cuyos cristales no son finos.

6.2.2. Rocas ígneas extrusivas

6.2.2.1. Andesita (AND)

Representan la roca caja pre-mineralización y corresponden a rocas andesíticas de las formaciones Abanico y Farellones, con un rango de edad entre 22.7 ± 0.4 y 16.8 ± 0.3 (Toro et al., 2012). El conjunto perteneciente a la formación Farellones afloran como un *roof pendant* en la parte superior del sistema, exhibiendo principalmente alteración propilítica y sericitización de plagioclasas. Estas rocas se observan frecuentemente como componentes clásticos en la mayoría de las brechas, siendo particularmente abundantes en la Brecha de Anhidrita. Múltiples fases del Batolito San Francisco, variando en composición y textura, intruyen a estas unidades andesíticas.

Las andesitas del sector Los Bronces se describen como rocas de color gris oscuro a negro, en ocasiones negro-verdoso a negro-azulado. También, presenta textura porfídica o afanítica, observándose un aumento de la textura porfírica cercana al contacto con otras unidades. Asimismo, exhibe un 10-35 % de fenocristales de plagioclasa, epidota, clorita, hornblenda, turmalina y magnetita. Además, una masa fundamental compuesta por plagioclasa, clorita, sericita, cuarzo, especularita y magnetita. Con respecto a los eventos de alteración, suele presentar alteración propilítica débil a moderada y alteración cuarzo-sericita débil. Vale mencionar que la mineralización de las andesitas es escasa a inexistente, observándose comúnmente vetillas con piritita total o parcialmente lixiviada.

6.2.2.2. Riolita (RIO)

Perteneciente al Complejo La Copa, corresponde a una chimenea de brecha subvolcánica de composición riolítica, que en su parte superior está conformada por flujos de brechas tobáceas soldadas riolíticas y pórfido riolítico. Parte de su extensión se apoya a través de discordancia de erosión sobre estratos brechizados de la Formación Farellones. Por otro lado, la roca suele presentar un color oscuro, clastos de composición andesítica y fiammes (AngloAmerican, 2023).

6.2.3. Brechas

Warnaars et al. (1985) identificaron al menos siete cuerpos de brechas que son, desde la más antigua a la más reciente, Fantasma, Central, Occidente, Infiernillo, Anhidrita, Gris Fina y Donoso. Estas se distinguen por su ubicación, matriz/cemento, tipo de clastos y grado de alteración-mineralización. Se reconoce que el contenido de cobre hipógeno aumenta progresivamente desde la brecha más temprana (antigua) a la más tardía (reciente), donde brechas más tempranas (Fantasma, Central y Occidente) alcanzan contenidos de cobre hipógeno de hasta 0.2 % Cu, caracterizadas por un cemento de asociación cuarzo-turmalina-pirita-calcopitira. En cuanto a las fases intermedias (Infiernillo, Anhidrita y Gris Fina), alcanzan contenidos de cobre hipógeno de hasta 0.6 %Cu, con mayores contenido de calcopirita. Notablemente, la brecha Donoso presenta el mayor contenido de cobre hipógeno, alcanzando leyes promedio de 1 % Cu, cuyo cemento corresponde a una asociación de cuarzo-turmalina, pirita-calcopirita y calcopirita-bornita (Toro et al., 2012).

En general, las brechas pueden presentar clastos de andesita, cuarzomonzonita y diorita, mientras que la matriz y/o cemento puede ser cuarzo, turmalina, especularita, anhidrita, pirita, calcopirita, bornita, molibdenita, sericita, clorita y polvo de roca, con relaciones de contacto comúnmente gradacionales (Toro et al., 2012). En definitiva y teniendo en cuenta las definiciones anteriores, a nivel mina, se han redefinido las brechas, clasificándolas de acuerdo a la naturaleza de brechización y emplazamiento (AngloAmerican, 2023).

6.2.3.1. Brechas magmático-hidrotermales de emplazamiento profundo

- **Brecha de Feldespato Potásico (BXFK):** Corresponde a una brecha con matriz de feldespato potásico + albita + cuarzo, de emplazamiento en ambiente frágil-dúctil observado en sectores profundos del trend principal de brechas de Los Bronces. Asimismo, se relaciona con el núcleo potásico y se superpone a brechas magmáticas y de matriz de pórfido, presentando contenido variable de calcopirita-pirita y algunas veces con bornita.
- **Brecha de Pórfido (BXP):** Corresponde a una brecha cuya matriz es porfírica, generalmente de composición intermedia. Se produce durante la intrusión y emplazamiento de pórfidos (PGR, PQM, PDA y PDL), variando la matriz desde grano fino, cuando hay escaso desarrollo de matriz, a presentar textura porfídica cuando hay mayor porcentaje de matriz (hasta 35 %). Los fragmentos se observan subredondeados a subangulosos, de tamaño fino, comúnmente son rocas félsicas del Batolito San Francisco y algunas veces BXF.
- **Brecha Fantasma (BXF):** Corresponde a una brecha de matriz magmática-hidrotermal que suele ser monomítica. Los clastos se observan de tamaño fino a grueso princi-

palmente de QM, varían de angulosos a subredondeados, con mineralización diseminada y ocasionalmente en *stockwork*. Su matriz esta caracterizada por polvo de roca y cuarzomonzonita con pequeños cristales diseminados de cuarzo, turmalina, especularita y sulfuros. Tanto clastos como matriz, pueden presentar una alteración cuarzo-sericita de moderada a intensa, lo que dificulta la distinción entre ambos. Asimismo, los clastos pueden presentar contactos desde nítidos a difusos, lo que se interpreta como emplazamiento en ambiente frágil-dúctil. Por otro lado, la matriz varía de magmática a hidrotermal, presentándose como roca intrusiva asignable al Batolito San Francisco y en algunas zonas incorpora biotita-cuarzo-feldespatos potásico hidrotermal subordinado a la matriz magmática descrita. Cabe mencionar que se considera como la brecha más temprana, ya que se han reconocido clastos de ella en todas las otras bechas y remanentes que son cortados por el emplazamiento de brechas más tardías (Abarzúa, 2014; AngloAmerican, 2023).

6.2.3.2. Brechas hidrotermales de emplazamiento medio

- **Brecha de Biotita (BXXM):** Brecha polimíctica formada por clastos subangulosos de QM, QMD y pórfidos. Su matriz esta compuesta por biotita, magnetita, clorita, cuarzo, anhidrita, polvo de roca y, en menor medida, sericita verde y sulfuros. En cuanto a los sulfuros, diseminados y en matriz, predominan calcopirita y pirita, con presencia de bornita y molibdenita. Domina la alteración potásica expresada por la abundancia de biotita secundaria en matriz, así como en fragmentos, albitización y alteración a feldespatos potásico de las plagioclasas. Esta brecha se presenta siempre en ambiente potásico, siendo unas de sus propiedades diagnósticas su matriz negra grisácea con biotita secundaria junto a sílice.
- **Brecha de Turmalina (BXXC):** Compuesta por clastos de QM, angulares a subredondeados, tamaño variable entre 5 y 50 cm, cabe mencionar que sus clastos angulosos suelen presentarse con poco retrabajo, otorgándole un aspecto de rompecabezas. Su matriz también varía en porcentaje de 10 a 30 % y está compuesta por turmalina masiva, cuarzo, anhidrita, especularita y calcopirita. Se observa nula presencia de polvo de roca y gran desarrollo cristalino de anhidrita, turmalina y cuarzo. Se presenta como cuerpos vetiformes discontinuos principalmente al SW del tren principal de brechas de Los Bronces, siendo su zona de ocurrencia elemento clave a la hora de reconocerla, ya que no se presenta en otros sectores del yacimiento.
- **Brecha Donoso (BXXD):** Brecha con matriz de turmalina masiva (15-35 %). Su matriz puede alcanzar hasta 10 % en volumen de sulfuros, especularita y escasamente anhidrita y bornita. Con respecto a sus clastos, angulosos a subangulosos y de tamaño fino a intermedio, estos se componen por QM, menor QMD, sienita y raramente AND. De emplazamiento somero (ambiente frágil, fragmentos angulosos), se interpreta como el último pulso en el trend principal de brechas de Los Bronces, al contener clastos de BXXC y BXXO en sus márgenes. Constituye uno de los cuerpos de

mejor ley del yacimiento, presentando sectores de hasta 5 %CuT, con leyes promedio de 1.2 % Cu (Skewes et al., 2003). Cabe mencionar que esta brecha comúnmente se presenta con una alteración fílica moderada. Una de sus características diagnósticas es la turmalina masiva acompañada de la ocurrencia de sulfuros masivos de mena en su matriz, principalmente calcopirita.

- **Brecha Anhidrita (BXA):** Brecha con matriz (30-60 %) de grano grueso de anhidrita. Además, sus clastos más comunes son fragmentos de brechas hidrotermales (BXC, BXI), lo que evidencia su emplazamiento posterior a estas, junto a andesitas débilmente cloritizadas a biotitizadas en profundidad. También puede presentar minerales opacos como especularita, pirita, calcopirita y molibdenita. Puesto que se suele confundir brechización de anhidrita con brecha de anhidrita, se fijó que si la matriz alcanza un 30 % se considera brecha, mientras que si es menor se considera brechización.
- **Brecha Central (BXC):** Brecha con matriz (30-40 %) de turmalina, polvo de roca, menor cuarzo, especularita, sericita, sulfuros y raramente anhidrita, lo que le confiere un color negro a gris oscuro. Sus clastos se componen principalmente por QM, los que están hidrotermalmente afectados por alteración cuarzo-sericita y/o clorita-sericita mayoritariamente. Estos clastos son en su mayoría finos a muy finos, angulosos y fuertemente obliterados. Más aún, a diferencia de BXD, los sulfuros pueden presentarse diseminados tanto en la matriz como en los clastos, con un contenido mayor de pirita que calcopirita.
- **Brecha Hidrotermal Indiferenciada (BXH):** Brecha polimíctica, matriz soportada y con clastos angulosos, finos, difusos y generalmente obliterados. La matriz se compone principalmente de sílice, turmalina diseminada, sulfuros (pirita y molibdenita), especularita, clorita y escaso polvo de roca. Por otro lado, los clastos que la componen corresponden a litologías de rocas del Batolito San Francisco (QM, QMD, PA y pórfidos félsicos), brechas de turmalina preexistentes y fragmentos de BXI y BXA. Con respecto a la alteración, predomina la alteración cuarzo-sericita moderada a intensa. Por último, la mineralización hipógena corresponde mayoritariamente a pirita, molibdenita y calcopirita, diseminada tanto en clastos como en matriz. Más aún, las zonas donde la ley de cobre aumenta se debe principalmente a la presencia de calcosina como mineralización secundaria.

6.2.3.3. Brechas freático-hidrotermales de emplazamiento somero

- **Brecha Infiernillo (BXI):** Brecha con matriz (2-15 %) verdosa de clorita, cuarzo y en menor cantidad especularita, turmalina, epidota, pirita, calcopirita y magnetita. Por otro lado, sus clastos son angulosos a subangulosos y mal preservados, compuestos por fragmentos de AND y menores cantidades de QM y QMD. Más aún, los minerales máficos de los clastos están fuertemente cloritizados y en ciertas áreas silicificados

y argilizados. Con respecto a la mineralización, en general, es escasa predominando pirita.

- **Brecha Occidente (BXO):** Brecha de matriz verdosa con abundante clorita y polvo de roca. Sus clastos son angulosos y se preservan de buena manera, suelen ser fragmentos de QM o QMD que presentan alteración clorítica y de reemplazo de minerales máficos por clorita, especularita y/o turmalina. Por otro lado, la mineralización es escasa, tanto en la matriz como en los clastos, predominando pirita diseminada. Esta brecha sería posterior a BXC y simultánea a BXI con una orientación NNW-SSE (Abarzúa. 2014).
- **Brecha Gris Fina (BXGF):** Brecha con matriz (10-40 %) microcristalina de color gris, formada por polvo de roca, en menor cantidad finos granos de turmalina, sericita, cuarzo, clorita, especularita y sulfuros. En cuanto a sus clastos, están compuestos por fragmentos de PQM y AND. Además, presenta una fuerte alteración cuarzo-sericita y, localmente, silicificación y cloritización. Con respecto a los clastos de sulfuros, se presentan en vetillas y finamente diseminados. Por último, esta brecha corta y contiene fragmentos de BXC, BXI y BXO (Abarzúa, 2014).
- **Brecha Monolito (BXMN):** Brecha polimíctica con clastos redondeados a subredondeados, mientras que su matriz presenta variaciones en la vertical, pasando de clorita primaria a clorita derivada de biotita, biotita, turmalina y biotita-magnetita. Adicionalmente, su matriz puede presentar polvo de roca.
- **Brecha Polvo de Roca (BXPR):** Brecha con matriz grisácea de polvo de roca (mayor al 30 %), sericita, clorita, cuarzo hidrotermal, especularita y sulfuros. Además, sus clastos son angulosos y variados, principalmente de QM, QMD, pórfidos y brechas en menor medida. En cuanto a su ocurrencia, se observa en distintos sectores del yacimiento, sin embargo, su presencia en la fase Infiernillo 5 se interpreta como la continuidad en profundidad de la BXI.
- **Brecha de Turmalina Bronces Sur (BXTS):** Brecha con matriz compuesta por turmalina negra y polvo de roca, con escaso cuarzo hidrotermal, pirita, calcopirita y especularita, raramente anhidrita y bornita. Sus clastos son fragmentos de QM fina y en menor cantidad QMD. Más aún, Esta brecha se encuentra en Bronces Sur (San Enrique Monolito) al oeste del pórfido Don Luis, el que la corta y puede presentar sectores por sobre el 3 % de CuT.

6.3. Alteraciones Hidrotermales

Las alteraciones hidrotermales en Los Bronces se relacionan con los distintos cuerpos de brechas con una zonación vertical, donde la alteración potásica biotítica es predominante y está presente en todos los sectores, desde Donoso por el norte a Los Sulfatos

por el sur. En los niveles más superficiales, particularmente en las brechas periféricas, se observa una alteración inicial de clorita-caolinita en el centro de los clastos, que progresa hacia sericita en los bordes. El cemento hidrotermal de la matriz incluye cuarzo, turmalina, especularita, pirita y calcopirita. A profundidades intermedias, las brechas presentan una composición de cuarzo-turmalina, biotita, feldespato potásico, anhidrita, magnetita, calcopirita y bornita, donde los sulfuros se observan principalmente diseminados y en *stockwork*. En los niveles más profundos, predomina un cemento hidrotermal de biotita, acompañado en menor medida por pirita, calcopirita y bornita (Toro et al., 2012).

Abarzúa (2014), basado en estudios de las asociaciones minerales y estudios petrográficos, clasifica tres grandes grupos de alteración, Potásico, Fílico y Propilítico, desarrollando modelos 3D con la distribución espacial de estas en la mina (figura 6.2). En cuanto a las relaciones temporales de las alteraciones, se tiene que la alteración potásica y fílica se traslapan con los rangos de edad en molibdenitas entre 5.89 y 4.24 Ma (edades $40\text{Ar}/39\text{Ar}$ en Deckart et al., 2005, 2013), siendo la alteración fílica el evento más joven (Deckart et al., 2014). Mientras que la alteración propilítica se superpone a la alteración potásica y es superpuesta por la alteración fílica, existiendo zonas de transición entre todas. Para concluir, a nivel operación y siendo la clasificación más reciente, se han definido al menos seis tipos de alteración en Los Bronces basados en el estudio de Abarzúa (2014), las que se detallan a continuación.

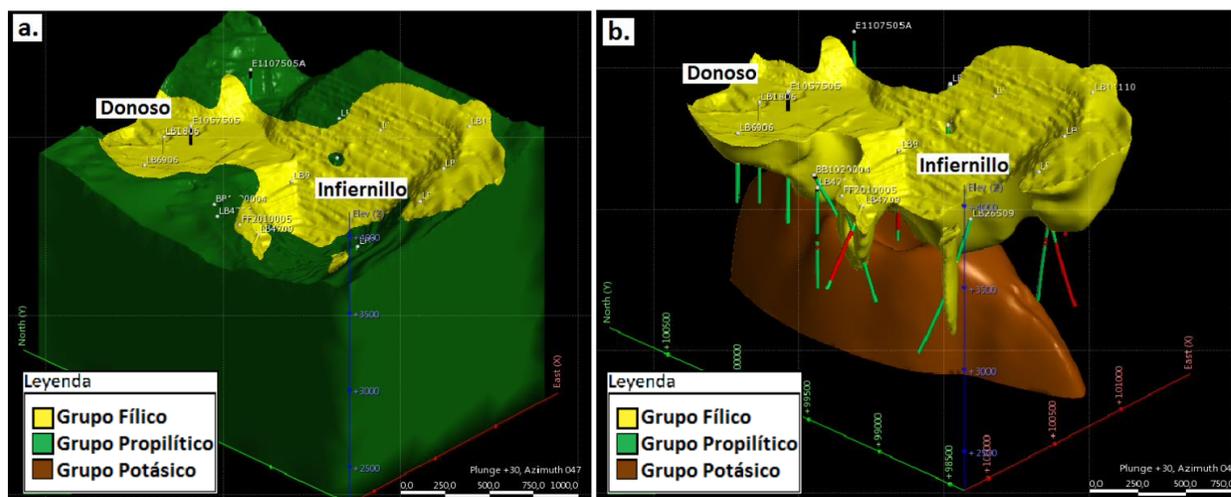


Figura 6.2: Modelo 3D de sólidos de alteración hidrotermal en la mina Los Bronces. En a) se observa el halo de alteración propilítica (verde), la cual es cortada y superpuesta por la alteración fílica (amarillo), mientras que en b) se visualiza el núcleo potásico profundo (pardo), de orientación NNW-SSE, cubierto y sobre impuesto por las alteraciones del tipo fílicas. Extraído de Abarzúa (2014).

6.3.1. Alteración Potásica (Grupo Potásico)

La alteración potásica afecta tanto a las rocas granodioríticas del Batolito San Francisco como a las rocas volcánicas que conforman la roca caja, posee una mineralización diseminada y en *stockwork* de alta ley (Abarzúa, 2014). Comprende minerales de alta temperatura, ocurriendo en ambiente frágil-dúctil, incluyendo biotita diseminada, como sobrecrecimiento, en vetillas y matriz; albita en halos de vetillas y/o alterando plagioclasas en roca de caja; cuarzo-anhidrita en vetillas A y B; y magnetita diseminada y en vetillas. Más aún, esta alteración incluye la zona del núcleo potásico con presencia de calcopirita, menor bornita, ocasional presencia de pirita y molibdenita, con desarrollo claro de *stockwork* de vetillas A, B, EB y sobreimposición de vetillas de sericita-gris verde, principalmente C y C2. Dicha alteración también se expresa en el sector distal del núcleo potásico, el que es representado por biotitización metasomática de máficos del batolito San Francisco, siendo sobreimpuesta por alteración clorítica (AngloAmerican, 2023).

6.3.2. Alteración Sericita-Gris Verde (Grupo Potásico)

Corresponde a una alteración hidrolítica de alta temperatura, relacionada en tiempo y espacio a la alteración potásica. Sus minerales principales son sericita-anhidrita-biotita y subordinadamente clorita-cuarzo-epidota-feldespato potásico (Abarzúa, 2014). Se expresa en las vetillas C, C1 y C2, las que exhiben halo sericítico con presencia de clorita, magnetita-especularita y sílice, además de calcopirita y pirita diseminadas en menor medida. En sectores más profundos y de mayor temperatura, se observa biotita relictas en equilibrio en el halo, así como magnetita y bornita. El relleno lo constituye principalmente calcopirita-pirita con presencia de cuarzo y anhidrita, variando desde calcopirita en el sector proximal a la fuente de calor hasta predominar la pirita en sectores más distales. Texturalmente, el halo es de grano fino y se oblitera casi por completo la textura de la roca original, a diferencia de la alteración cuarzo-sericítica o clorita-sericita-arcillas, en donde sí se preserva al menos el cuarzo intersticial de las rocas del batolito San Francisco. A este evento se sobreimponen alteraciones hidrolíticas más tardías, frías y someras (cuarzo-sericitaz o clorita-sericita-arcillas), haciendo difícil su identificación en sectores transicionales (AngloAmerican, 2023). Es difícil distinguirla de una alteración cuarzo-sericita, sin embargo la alteración sericita-gris verde es más temprana y por lo tanto es cortada por las vetillas tipo D de cuarzo con halo sericítico blanco, sumado a que tiende a mostrar un color modal más gris a verde y una mineralogía de mayor temperatura (biotita, feldespato potásico, cuarzo y anhidrita). Finalmente, su importancia radica en el alto contenido de calcopirita, con menor o nula presencia de molibdenita, y en su proximidad y relación con sectores con alteración potásica intensa (marginales y cercanos a núcleo potásico).

6.3.3. Alteración Albita - Turmalina - Epidota (Grupo Propilítico)

Se presenta en cúmulos, diseminaciones y vetillas de albita (en halo), turmalina y epidota (vetillas C3A y cúmulos o parches). Al mismo tiempo, se interpreta como una asociación en contexto propilítico proximal a alteración potásica. Por lo que es común encontrar esta alteración coexistiendo con biotitas secundarias cloritizadas. Por otro lado, se observa frecuentemente con mineralización de pirita, con vetillas de pirita gruesa y subordinadamente calcopirita (AngloAmerican, 2023). Esta asociación es reflejo de un metasomatismo de tipo sódico-cálcico, caracterizado por los minerales albita-epidota-turmalina-cuarzo y subordinadamente clorita-sericita-arcillas-anhidrita (Abarzúa, 2014).

6.3.4. Alteración Clorita - Epidota (Grupo Propilítico)

Se presenta predominantemente en vetillas C3B y alteración de máficos a clorita, ocasionalmente epidota reemplazando selectivamente a las plagioclasas y máficos. También se puede observar un halo difuso con presencia o ausencia de albita en vetillas, débilmente alterando plagioclasas de roca caja en sectores cercanos a la fuente de calor (AngloAmerican, 2023). Esta asociación corresponde al halo más distal y prograda del propilítico, está caracterizado por la asociación clorita-epidota y subordinadamente albita-cuarzo-anhidrita-yeso-calcita (Abarzúa, 2024).

6.3.5. Alteración Cuarzo - Sericita (Grupo Fílico)

Alteración hidrolítica característica de niveles someros del yacimiento, se manifiesta dentro y alrededor de las brechas de turmalina y polvo de roca, y ocurre principalmente como reemplazo selectivo de los feldespatos, variando de parcial a total. Se observa en vetillas D, con pirita como mineral predominante tanto en relleno como diseminada, y en ocasiones, calcopirita y especularita subordinadas. También, se identifican dos halos de alteración sericítica: un halo interno, donde la alteración sericítica es más intensa, y un halo externo, más débil. Además, se ha reconocido asociada con turmalina en brechas, donde la turmalina ocurre relleno de espacios y formando la matriz, mientras que los clastos presentan un reemplazo parcial o total por cuarzo y/o sericita (Abarzúa, 2014; AngloAmerican, 2023).

6.3.6. Alteración Clorita - Sericita - Arcillas (Grupo Fílico)

Corresponde a una alteración hidrolítica tardía, incluye condiciones de sulfuración desde intermedia a baja. Se superpone a todas las demás alteraciones hipógenas y ocurre en zonas de debilidad estructural, asociado a vetillas, vetas y vetas DT. Compuestas por carbonatos férricos-yeso-especularita-calcopirita en un contexto de baja sulfuración hasta encontrar presencia de especularita con presencia de enargita, tennantita, calcopirita en condiciones de sulfuración intermedia. También se observa baritina localmente y se observan anomalías de arsénico en vetillas y vetas DT (AngloAmerican, 2023). Más aún,

presenta halos principalmente de sericita, clorita en menor medida y ocasionalmente argilización débil. Estos halos son de grosor variable (cm hasta m), reflejando procesos de intensa interacción agua-roca, de modo que, cuando la alteración no es tan intensa, el feldespato potásico se tiende a preservar por sobre las plagioclasas, mientras que si la alteración hidrolítica persiste, los feldespatos son totalmente reemplazados a arcillas. En general, la alteración ocurre de forma penetrativa y pervasiva, pero siempre restringida a estructuras y/o zonas de mayor permeabilidad (Abarzúa, 2014).

6.4. Vetillas

La clasificación de vetillas utilizada actualmente fue redactada por Holmgren & Núñez (2011) para el sector Los Bronces Sur. Dicha clasificación esta basada en la realizada al yacimiento El Salvador por Gustafson y Hunt (1975), posteriormente modificada por Gustafson y Quiroga (1995), quienes definen la mineralogía, aspectos texturales y relaciones de temporalidad de las vetillas.

6.4.1. Vetillas tempranas

6.4.1.1. Vetillas Clase M

Se caracterizan por sus bordes irregulares y difusos, con espesor variable que generalmente no excede de 1 mm (Rojas, 2022). No presentan halo, ya que se producen en un ambiente frágil-dúctil. En cuanto a los minerales de ganga presentes en su relleno, predomina la magnetita y, en menor medida o ausentes, cuarzo y anhidrita. Además, la mineralización del relleno es principalmente calcopirita y pirita, los que están subordinados o incluso ausentes. Estas vetillas se asocian con zonas profundas y de alta temperatura, donde predomina la alteración potásica. Cabe mencionar que se han reconocido de forma muy escasa en la mina Los Bronces (Abarzúa, 2014).

6.4.1.2. Vetillas Clase EB

Se caracterizan por ser vetillas finas, con bordes irregulares y continuas, destacando su halo proporcional a su anchura, veánse en figura 6.3. Además, se reconocen en zonas profundas normalmente formando *stockwork*. En cuanto a su relleno, está compuesto predominantemente por cuarzo, con cantidades menores de anhidrita, albita, feldespato potásico, magnetita y biotita (Abarzúa, 2014). Con respecto a su mineralización de mena, se distingue una presencia predominante de calcopirita, seguida por pirita y bornita, distribuidas de forma diseminada. Por último, en su halo se observa biotita y magnetita en menor concentración, junto a sericita gris verde subordinada.

6.4.1.3. Vetillas Clase A

Se caracterizan por ser continuas, con espesor entre 1 a 10 mm, sin halos y con bordes regulares y bien definidos, veánse en figura 6.6. No obstante, pueden presentar un

halo muy fino y discontinuo. En cuanto a los minerales ganga en su relleno, se reconoce predominancia de cuarzo sobre anhidrita, con feldespato potásico y albita subordinados o ausentes. Asimismo, la mineralización de mena en su relleno corresponden a calcopirita, pirita y muy escasa bornita. Es normal encontrarlas siendo cortadas por otro tipo de vetillas, evidenciando su carácter temprano y profundo (AngloAmerican, 2023). En general, ocurren en zonas profundas asociado a los dominios de alteración potásica, o bien, con una menor frecuencia, como un halo externo de la alteración potásica (Abarzúa, 2014).

6.4.1.4. Vetillas Clase B

Corresponden a vetillas planares y continuas, con bordes regulares y bien definidos, cuyo espesor varía entre 5 a 15 mm (AngloAmerican, 2023). No obstante, pueden presentar un halo fino, penetrativo y continuo. En cuanto a los minerales de ganga, cuarzo y anhidrita son predominantes, con menor cantidad de feldespato potásico y albita. Por otro lado, sus minerales de mena principales son molibdenita y calcopirita, con pirita subordinada y bornita muy escasa o ausente (Abarzúa, 2014). En cuanto a su halo, puede componerse de feldespato potásico, albita, cuarzo, sericita gris verde, calcopirita, pirita y, ocasionalmente, biotita. Estas vetillas se ubican típicamente en zonas intermedias a profundas, asociándose con la alteración potásica profunda. En particular, la molibdenita se puede disponer de tres formas en este tipo de vetillas, 1) de manera discontinua en los bordes, 2) como sutura en el centro y 3) como cúmulos discontinuos a través de la vetilla.

6.4.2. Vetillas transicionales

6.4.2.1. Vetillas Clase C

Corresponden a vetillas continuas y sinusoidales, con bordes irregulares, halo simétrico y espesor relativamente constante entre 2 a 3 mm, veánse en figura 6.4. El relleno está compuesto principalmente por cuarzo, seguido de biotita y magnetita, con presencia ocasional de anhidrita. En cuanto a la mineralización presente en el relleno, la calcopirita es el mineral principal, acompañado de pirita y trazas de bornita que pueden ser escasas o inexistentes. El halo de estas vetillas es continuo, difuso y gradacional con bordes poco definidos, proporcional al ancho del relleno, incluyendo cuarzo, biotita, sericita gris verde, feldespato potásico y, en menor frecuencia, albita. Con respecto a la mineralización en el halo, está compuesta principalmente por calcopirita y en menor medida pirita y bornita. Cabe mencionar que todos los minerales en el halo ocurren de forma diseminada. Más aún, este tipo de vetillas se asocian comúnmente con zonas profundas donde domina la alteración potásica, pero su zona de ocurrencia principal es una zona transicional entre el ambiente potásico y fílico (Abarzúa, 2014).

6.4.2.2. Vetillas Clase C1

Corresponden a vetillas planares y continuas, con bordes regulares y bien definidos, cuyo espesor varía entre 2 y 3 cm. En cuanto a su halo, es restringido y puede ser difuso,

manteniéndose proporcional en espesor al del relleno, con un promedio de 4 mm. Los minerales de ganga que componen el relleno son cuarzo y anhidrita subordinada, mientras que calcopirita y pirita son los principales minerales de mena, con presencia ocasional de bornita y molibdenita. Con respecto a los minerales de ganga en el halo, predomina el feldespato potásico y albita, con cuarzo de forma subordinada y escasa sericita temprana; también se reconoce biotita y clorita. Por otro lado, calcopirita, pirita y bornita subordinada componen la mineralización del halo. En particular, todos los minerales del halo ocurren de forma diseminada. Estas vetillas se asocian a zonas profundas, superiores a 500 m, donde predomina la alteración potásica. Aunque, también se identifican en entornos de alteración fílica temprana y alteración de tipo sódica.

6.4.2.3. Vetillas Clase C2

Corresponden a vetillas continuas, de bordes regulares y bien definidos, manteniendo un espesor que varía entre 2 y 4 mm, aunque en ocasiones pueden alcanzar los 5 cm, veánse en figura 6.5. En cuanto a su halo, es simétrico, continuo y proporcional en espesor al del relleno. Los minerales presentes en el relleno son principalmente cuarzo y anhidrita subordinada, mientras que en la mineralización, predomina la calcopirita, seguida por pirita y, de manera más escasa, bornita y molibdenita. Por otro lado, presenta un halo compuesto principalmente por biotita y sericita gris verde, con clorita y magnetita subordinadas. En cuanto a la mineralización, se tiene calcopirita dominante y pirita subordinada. Cabe mencionar que la predominancia de calcopirita sobre pirita es característico de este tipo de vetillas. No obstante, en zonas de transición entre la alteración potásica y fílica, específicamente donde predomina la alteración fílica, la relación se invierte y la pirita se vuelve más dominante que la calcopirita. Estas vetillas se asocian con zonas profundas del yacimiento, representando la transición entre un ambiente potásico tardío o un fílico temprano.

6.4.2.4. Vetillas Clase C3

Este grupo de vetillas se subdivide en dos, las que difieren básicamente en el relleno, veánse en figura 6.8. Son vetillas ampliamente reconocibles en el distrito asociándose a la alteración propilítica. Dicha división es la siguiente:

- **Vetillas Clase C3A**

Corresponden a vetillas continuas, con bordes irregulares y un espesor que se mantiene constante en promedio entre 2 y 3 mm, aunque pueden encontrarse de mayores anchuras ocasionalmente. En cuanto a su halo, es simétrico y continuo, con bordes irregulares y espesor proporcional al del relleno. Los minerales presentes en el relleno corresponden a turmalina principalmente, seguida de cuarzo y, de forma subordinada, anhidrita, aunque es posible reconocer biotita y magnetita. Por otro lado, los minerales de mena principales son pirita y calcopirita, cuyas concentraciones varían según la profundidad y el ambiente específico. Con respecto a la composi-

ción mineral del halo, predomina feldespatos potásico y albita, con cuarzo, clorita y epidotas subordinados. Si bien la pirita es el mineral dominante en el halo, puede presentar sectores con mayor contenido de calcopirita, no obstante, ambos minerales no son muy abundantes en estas vetillas. Finalmente, se debe agregar que se ubican típicamente en zonas superficiales a intermedias, asociadas a la alteración propilítica extensa.

- **Vetillas Clase C3B** Corresponden a vetillas discontinuas, de bordes irregulares, finas a muy finas y de espesor variable. En cuanto a su halo, es discontinuo, difuso y también de bordes irregulares y espesor variable. Los minerales de ganga presentes en el relleno corresponden principalmente a epidota y clorita, con cuarzo, yeso y anhidrita subordinados o ausentes. Además, la mineralización es pirita dominante con calcopirita subordinada. Con respecto a la composición mineral del halo, se reconocen albita y cuarzo subordinado, con clorita, epidota y pirita en menor cantidad o ausentes.

6.4.3. Vetillas tardías

6.4.3.1. Vetillas Clase D

Corresponden a vetillas continuas y bordes regulares bien definidos, el espesor varía considerablemente desde 1 mm hasta 8 cm, veánse en figura 6.7. En cuanto a su halo, es simétrico y pervasivo, obliterando la roca de caja con límites claros y rectos. En el relleno, el cuarzo se presenta como el mineral de ganga principal, acompañado de anhidrita o yeso de manera subordinada; y la pirita domina la mineralogía de mena, con calcopirita y molibdenita presentes de forma escasa y ocasional, respectivamente. Por otro lado, está compuesto principalmente por cuarzo, sericita y arcillas, con biotita y escasa clorita. Más aún, la pirita es el mineral de mena predominante en el halo, con calcopirita casi ausente, ambos distribuidos de forma diseminada. Estas vetillas son típicas de zonas someras dominadas por la alteración cuarzo-sericita, aunque también pueden encontrarse escasamente, en ambientes de alteración propilítica y, excepcionalmente, en zonas más profundas, posiblemente debido a la presencia de estructuras.

6.4.3.2. Vetillas Clase DT

Son vetillas continuas, también conocidas como "D Tardías", veánse en figura 6.7, presentan bordes regulares y bien definidos, mostrando una variabilidad en sus espesores de entre 1 y 5 mm, aunque pueden alcanzar decenas de centímetros en localidades específicas (Abarzúa, 2014; Rojas, 2022). En cuanto a su halo, es penetrativo y destructivo, con un espesor proporcional al del relleno. En el relleno, cuarzo es el mineral de ganga dominante, acompañado de specularita, clorita, carbonatos de hierro (siderita y ankerita) y, en menor cantidad, yeso. Además, la pirita es el mineral de mena dominante en el relleno, seguida por calcopirita, presente en altas concentraciones en áreas particulares,

y molibdenita escasa o ausente. Por otra parte, en el halo predominan el cuarzo, con menor cantidad de clorita, sericita y arcillas, mientras que la calcopirita, presente como mineral de mena, exhibe concentraciones variables, generalmente bajas. Otro rasgo de estas vetillas, es que se ubican en zonas someras y están estrechamente asociadas con fluidos de bajas temperaturas y alteración hidrolítica intensa.

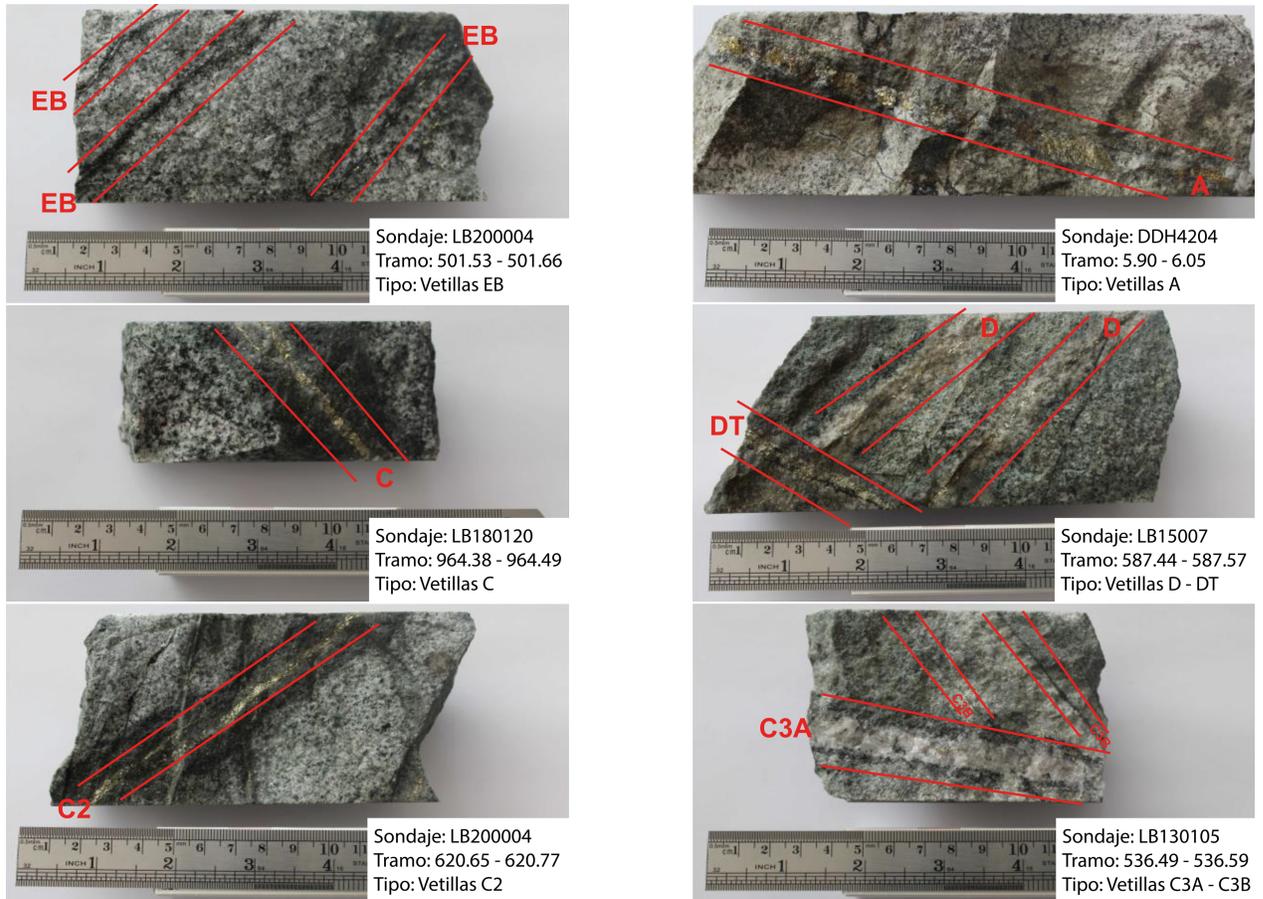


Figura 6.9: Fotografías de diferentes tipos de vetillas en la mina Los Bronces. Extraído de Rojas (2022).

CAPÍTULO 7

CONTACTO PRIMARIO-SECUNDARIO

El techo de sulfatos, llamado en este trabajo como "contacto primario-secundario", es un fenómeno que se ha observado en gran parte de los depósitos de tipo pórfido cuprífero, consiste en un límite o zona de transición entre roca primaria y secundaria.

Vela (2014) relaciona, localmente, este límite al comportamiento de la anhidrita presente en las alteraciones hidrotermales del depósito Los Bronces. En particular, al percolar las aguas meteóricas por las rocas y entrar en contacto con la anhidrita, que generalmente se encuentra cercana a la alteración potásica, esta se hidrata a yeso secundario, lo que produce un aumento de volumen e induce fracturamiento de rocas adyacentes. Posterior a esto, el yeso se disuelve por el paso del agua, quedando las estructuras sin relleno y abiertas. Lo anterior, permite distinguir dos unidades geotécnicas, roca primaria y roca secundaria. La roca primaria, más profunda y competente, puede presentar permeabilidad prácticamente nula, dado el sellamiento de estructuras, actuando como un piso impermeable de cara a los flujos de agua que la alcanzan por medio de infiltración a través de la roca secundaria, generando así que el agua escurra hacia cotas menores por medio de zonas preferenciales de circulación (fracturas, discontinuidades, etc.) (Demierres, 2021).

En particular, en el Distrito Los Bronces–Yerba Loca–Río Blanco hay presencia de abundante anhidrita asociada a la mineralización hipógena en zonas de alteración potásica y, por lo tanto, completamente ausente en zonas supérgenas. Dado lo anterior, el contacto primario-secundario se entiende como aquella superficie o zona donde dejan de tener efecto los procesos supérgenos de enriquecimiento de cobre y, bajo la cual, solo se evidencian los efectos de los procesos hipógenos de mineralización (JJa, 2019). En consecuencia, el techo de sulfatos en la mina Los Bronces (figura 7.1) se considera tanto como un horizonte de cambio en la calidad geotécnica como en la mineralogía presente.

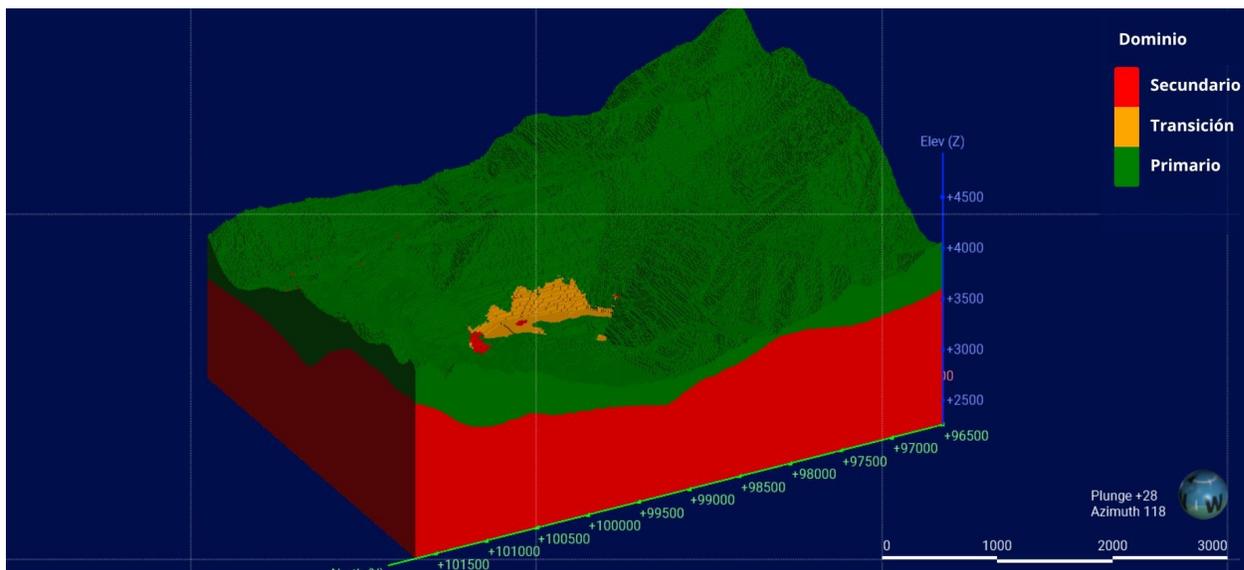


Figura 7.1: Modelo de bloques actual del contacto primario-secundario en la Mina Los Bronces, donde en color rojo se muestra el dominio primario, en verde el secundario y en naranja una transición entre estos.

CAPÍTULO 8

METODOLOGÍA

8.1. Base de datos

La base de datos corresponde a espectros combinados de ASD Terraspec y FTIR, y planillas de mapeo geológico de sondajes de la mina Los Bronces. Se entrega más detalle a continuación.

Con respecto a los espectros combinados de ASD Terraspec y FTIR, llamados de aquí en adelante como “Espectros ASD-FTIR”, estos se obtienen desde muestras de pulpa enviadas a los laboratorios de Anglo American ubicados en 1) Perth (Australia) y 2) Santiago (Chile). Los espectros analizados en Perth se agrupan en sólo un CSV ('Espectros ASD-FTIR PERTH') que compila 42 *labjobs*, totalizando 14,409 muestras con su espectro electromagnético asociado que abarca rangos de longitud de onda desde 350 nm hasta 44, 700.59 nm (figura 8.1). Para efectos prácticos, es un archivo con 14, 409 filas y 3,304 columnas, donde la primera columna corresponde al nombre de la muestra y las columnas restantes a los valores de reflectancia según longitud de onda. Por otro lado, los espectros analizados en las dependencias de Santiago (figura 8.2) se encuentran agrupados en el CSV 'Espectros ASD-FTIR SANTIAGO', totalizando 1,408 *labjobs* y 57, 562 muestras, lo que genera un archivo de 57, 562 filas y 3,301 columnas, con la misma estructura de Espectros ASD-FTIR Perth.

Por otro lado, las planillas de mapeo geológico corresponden a dos archivos. El primero de estos, llámese 'FTIR_Sondajes', compila la información del contacto primario-secundario de 992 sondajes de la mina Los Bronces (figura 8.3), así también como nombre del sondaje, ubicación del tramo (From, To), nombre de la muestra (SampleId), proyecto, *labjob* (LabJob Perth, LabJob Santiago) y coordenadas geográficas (X, Y, Z), y corresponde a un archivo CSV con 18, 572 filas y 11 columnas. En cuanto al siguiente archivo, llámese 'Lito_Comp', contiene la litología observada, por tramos, en 5385 sondajes (figura 8.4), además de su HoleId, ubicación (From, To), porcentaje de la litología en el tramo y coordenadas geográficas (X, Y, Z). Este último es un archivo CSV con 22, 017 filas y 8 columnas.

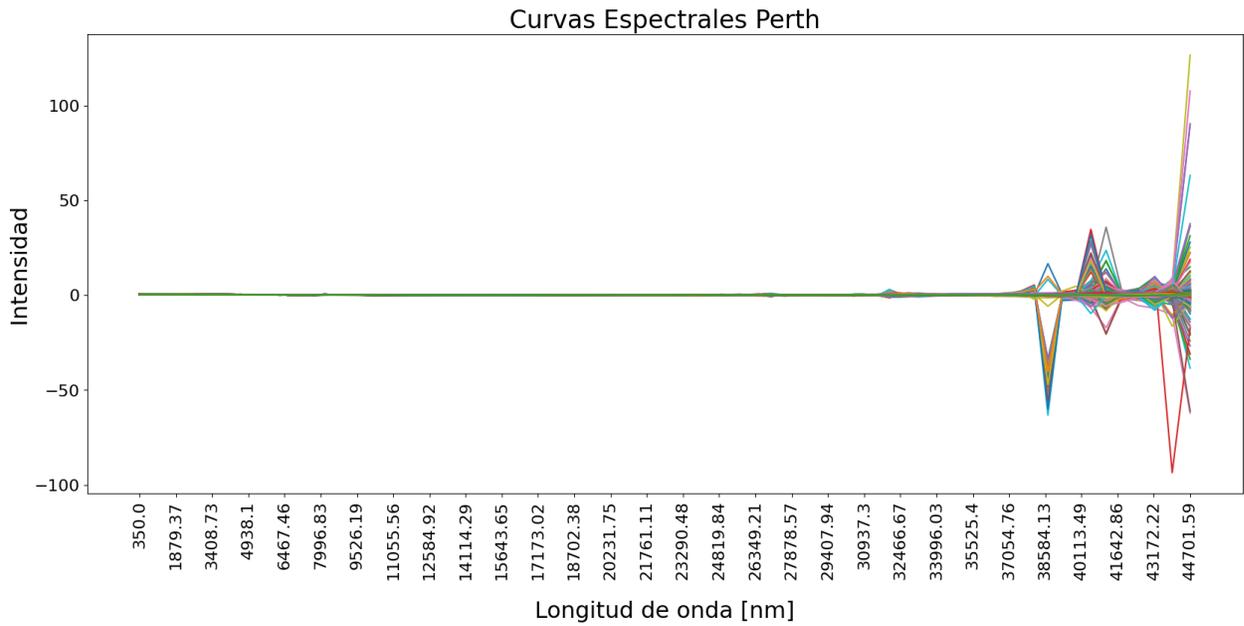


Figura 8.1: Curvas espectrales de muestras analizadas en los laboratorios de Anglo American ubicados en Perth, Australia. Datos sin ningún tipo de tratamiento.

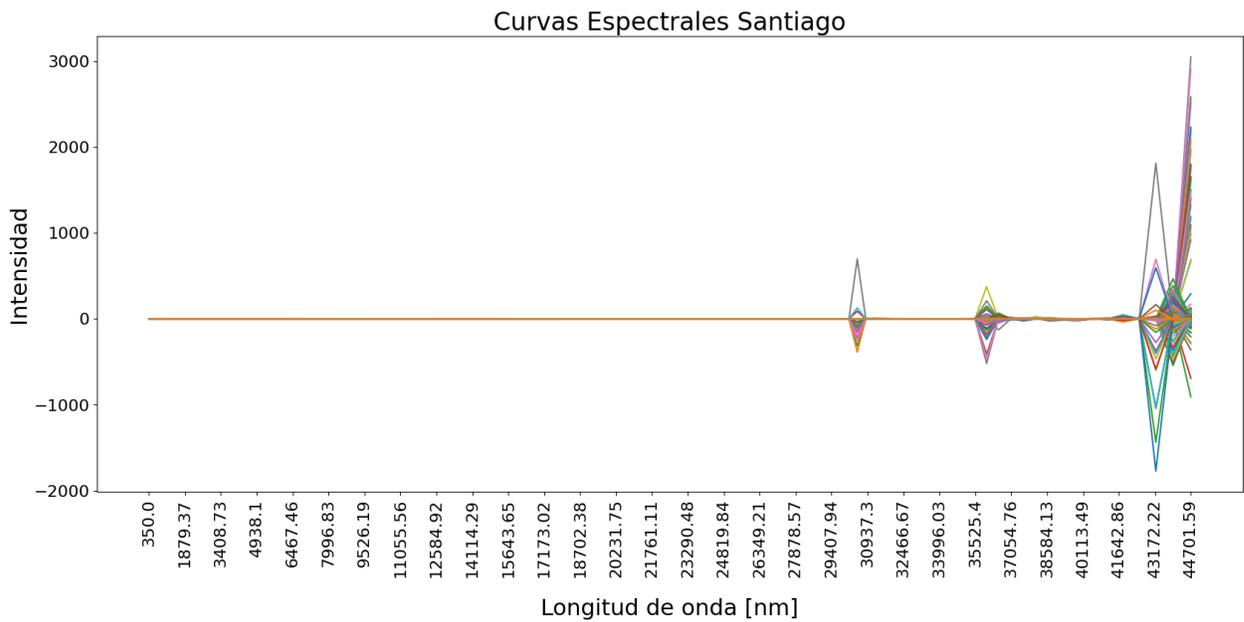


Figura 8.2: Curvas espectrales de muestras analizadas en los laboratorios de Anglo American ubicados en Santiago, Chile. Datos sin ningún tipo de tratamiento.

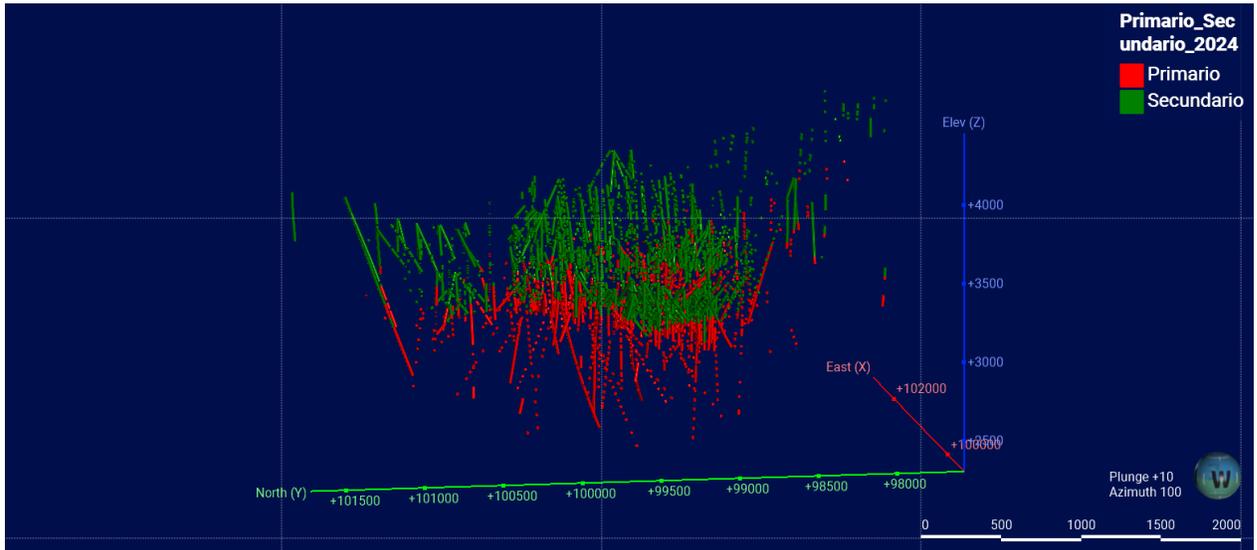


Figura 8.3: Visualización 3D de sondajes en *Leapfrog Geo* con información del contacto primario-secundario, en color rojo tramos mapeados como roca primaria y en color verde tramos mapeados como roca secundaria.

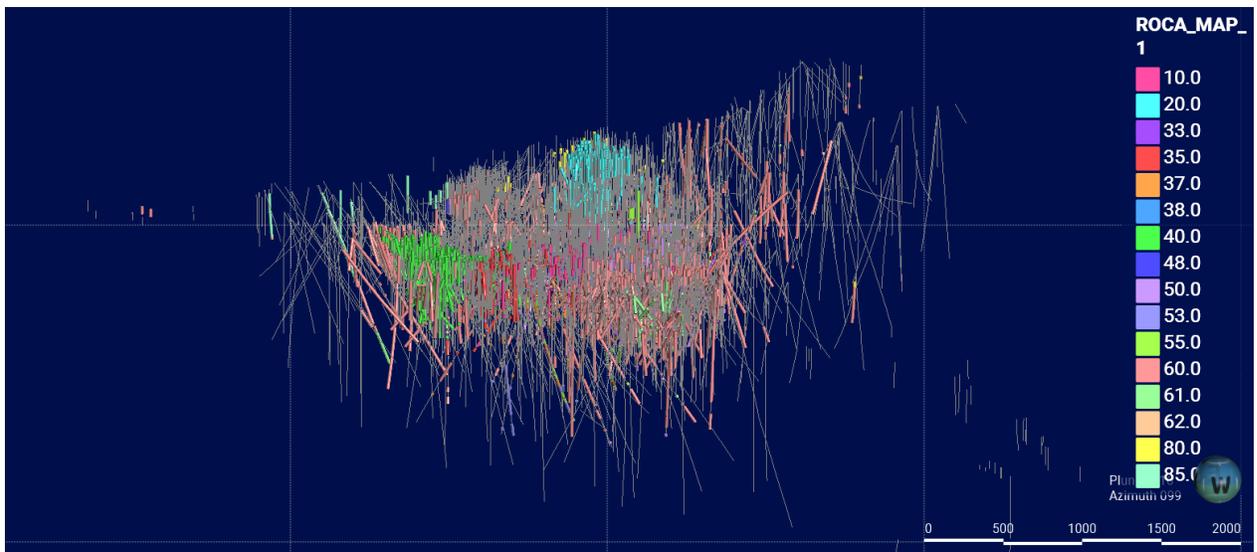


Figura 8.4: Visualización 3D de sondajes en *Leapfrog Geo* coloreados según litología reportada, donde los códigos corresponden tanto a litologías individuales como agrupadas. En color gris las trazas de los sondajes.

8.2. Preprocesamiento de los datos

Para el análisis de datos se utiliza *Google Colaboratory*, plataforma online que permite escribir y ejecutar códigos de Python. Posterior a la carga de los datos, se realiza el preprocesamiento de estos.

8.2.1. Sondajes

El tratamiento de 'FTIR_Sondajes' y 'Lito_Comp' se lleva a cabo mediante varios filtros, estos preparan los datos para ser integrados a los 'Espectros ASD-FTIR' y así, ser utilizados en la construcción de los modelos de clasificación. Mayor detalle a continuación:

1. **Filtro 1: Mantener litologías con 100%**, consiste en filtrar las columnas de 'Lito_Comp' que presenten mezclas de litologías y conservar sólo las que correspondan a 100% de una litología.
2. **Composición**, consiste en llevar al mismo soporte dos archivos. En este caso, se lleva a cabo para agregar la información de litología a 'FTIR_Sondajes'. Lo anterior se realiza en el software *LeapfrogGeo*, mediante la herramienta *Majority Composites* que permite llevar los datos categóricos a longitudes de intervalos de otra tabla. En este caso, los datos categóricos corresponden a los datos de litología y los intervalos se toman de 'FTIR_Sondajes', lo que genera un nuevo documento que posee las mismas columnas que 'FTIR_Sondajes' más dos columnas con información de litología.
3. **Filtro 2: Eliminar filas sin SampleID**, consiste en eliminar las muestras (filas) de un archivo que no tengan *SampleID* asociado, ya que la correlación con los 'Espectros ASD-FTIR' sólo es posible si se tiene el nombre de la muestra analizada, es decir, su *SampleID*.
4. **Filtro 3: Separar en muestras de Perth y Santiago**, consiste en crear 2 archivos, uno para contener las muestras de Perth y otro para las muestras de Santiago.
5. **Filtro 4: Ajustar nombre de las muestras a la nomenclatura solicitada**, consiste en ajustar la columna *SampleID* a la nomenclatura establecida, donde a cada muestra se le asigna un *SampleID* compuesto por un código referente al tipo de muestra seguido de un número de 5 dígitos. Si el número que sigue al código no tiene cinco dígitos, entonces se debe agregar un 0 delante. Por ejemplo, la muestra 'MET_04316' corresponde a una muestra metalúrgica (MET) de número 4316.
6. **Filtro 5: Separar muestras sin espectro**, consiste en separar las muestras de sondajes que no coinciden en *SampleID* con los espectros ASD-FTIR de Perth y Santiago.
7. **Filtro 6: Separar muestras sin contacto Primario/Secundario**, se crea un nuevo archivo para almacenar sólo las muestras que tengan información sobre el contacto primario-secundario, ya que el aprendizaje de los modelos será supervisado por estas etiquetas.

8.2.2. Espectros ASD-FTIR

El flujo comúnmente utilizado en espectroscopía se puede resumir en los cuatro siguientes pasos (Mokari et al., 2023) :

1. Limpieza (*Cleaning*), consiste en la detección y eliminación de *outliers*, así también como datos que se consideren de mala calidad por su baja relación señal-ruido, para lo anterior se pueden establecer umbrales tanto superiores como inferiores o a partir de zonas definidas como representativas de señal y de ruido (Gautam et al., 2015).
2. Filtrado o derivación (*Filtering*), consiste en una serie de técnicas diseñadas para refinar la señal mediante la eliminación de componentes no deseadas, estas técnicas pueden ser filtros derivativos, filtro Savitzky-Golay (SG) o autodesconvolución de Fourier (FSD).
3. Corrección de línea base (*Baseline correction*), consiste en ajustar o eliminar distorsiones en la línea base del espectro causadas por efectos instrumentales y de dispersión. Los métodos comúnmente utilizados son ajuste polinomial y diferenciación basado en filtros SG, corrección de dispersión multiplicativa (MSC) y MSC extendido (EMSC).
4. Normalización (*Normalization*), consiste en ajustar las diferencias de intensidad entre espectros para asegurar que los espectros del mismo material, medidos bajo condiciones ligeramente diferentes, se parezcan lo más posible. Los métodos más comunes de normalización incluyen la normalización min-max, la norma L1, la norma L2 y la normalización SNV.

En particular, como se puede observar en las figuras 8.1 y 8.2, el ruido de la señal comienza a volverse considerable hacia longitudes extremas (lado derecho del espectro), lo que se refleja en los gráficos como valores de reflectancia distorsionados (muy altos o negativos). Este fenómeno no es ajeno en espectroscopía y su origen se asocia a problemas del instrumento, factores ambientales o inconsistencias de las muestras (Mokari et al., 2023), que pueden ser ocasionados por 1) la sensibilidad del instrumento no es capaz de detectar con precisión las longitudes de onda extremas (energía más baja) debido a que la señal es muy baja, dominando el ruido sobre la señal detectada; 2) vibraciones externas durante la medición pueden causar distorsión de los espectros; 3) interferencias electromagnéticas, por ejemplo, dispositivos electrónicos cercanos al instrumento que pueden introducir ruido adicional; 4) la presencia de gases como CO₂ y vapor de agua que pueden absorber la radiación infrarroja y generar absorciones que se confunden con las características de la muestra; 5) inadecuada compensación de la línea de base (*baseline*) que puede alterar la posición y forma de las bandas espectrales. Es crucial realizar una calibración rigurosa y mantenimiento regular del instrumento para mitigar estos problemas y asegurar la precisión de los resultados.

No obstante, debido que se tienen limitaciones sobre el conocimiento pleno de la base de datos, se considera adecuado modificar los pasos de preprocesamiento. Primero, no se cuenta con espectros representativos de la señal ni del ruido, por lo que la relación señal-ruido no puede utilizarse para eliminar datos. Segundo, no se tiene un espectro representativo de un material en particular o un espectro de referencia, por lo que no se pueden normalizar las muestras, por ejemplo, por litología (se puede obtener una curva representativa de una litología al comparar todos los espectros de la misma, pero sigue siendo una estimación más que un estándar). Sumado a lo anterior, se desconocen las especificaciones técnicas de los instrumentos utilizados en la medición de los espectros, así también como las condiciones de laboratorio. En consecuencia, el preprocesamiento aplicado consiste en:

1. **Estadísticas**, consiste en calcular el promedio, desviación estándar, mínimo, máximo, conteo de valores y percentiles (25 %, 50 % y 75 %). Con lo anterior, se pueden detectar valores nulos, valores atípicos y estimar cuáles son las variables que aportan información. Cabe mencionar que se decide eliminar las columnas con valores nulos, ya que si bien las buenas prácticas evitan eliminar datos, no es adecuado estimar los valores nulos con muestras circundantes o longitudes cercanas, ya que se alteraría el espectro de la muestra, lo que podría llevar a una clasificación sesgada.
2. **Tratamiento de negativos**, consiste en detectar y eliminar los valores negativos de la base de datos, ya que, por definición, los valores de reflectancia deben mantenerse entre 0 y 1, donde 0 indica ninguna reflectancia (absorción total) y 1 indica reflectancia total (sin absorción). Considerando que no se pueden realizar correcciones adecuadas al no contar con información del instrumento ni sobre las condiciones de medición, se procede con la eliminación de estos datos.
3. **Tratamiento de reflectancia sobre 1**, consiste en detectar y eliminar valores con reflectancia sobre 1 (100 %), ya que, como se explicó anteriormente, debe mantenerse en valores desde 0 a 1 (0-100 %).
4. **Filtro 6: Separar muestras sin espectros**, las muestras a utilizar para el modelo deben poseer una curva desde 'Espectros ASD-FTIR', ya que es necesario para los modelos supervisados y redes neuronales contar con un input (qué necesito clasificar) y output (la clasificación en sí). Por lo tanto, se deben dejar fuera del análisis las muestras que no tengan espectro asociado, similar a lo realizado para las muestras sin *SampleID*. También, se deben separar de los archivos 'Espectros ASD-FTIR Perth' y 'Espectros ASD-FITR Santiago' las muestras que no tengan información de litología, contacto o ninguno de los anteriores. Por otro lado, para el modelo no supervisado se utilizará la base completa, ya que este tipo de modelos sólo necesita datos de entrada y la salida es interpretación del usuario. Finalmente, cabe mencionar que los espectros que no presentan coincidencias de *SampleID*, pueden ser utilizados en el modelo final seleccionado para estimar su clasificación de dominio.

5. **Integrar columna del contacto**, una vez separados los espectros que cuentan con *SampleID* coincidentes con los archivos de sondajes, se integra la columna del contacto primario-secundario. Finalmente, se cuenta con un archivo que contiene muestras de la mina Los Bronces que tienen un espectro ASD-FTIR, analizado en Santiago o en Perth, e información sobre a cual dominio (primario o secundario) pertenece.

Estos pasos dan como resultados una partición de la base original, donde, a grandes rasgos, se tiene un conjunto 1 ('Perth_1' y 'Santiago_1') que se genera al eliminar filas negativas, un conjunto 2 ('Perth_2' y 'Santiago_2') derivado de eliminar columnas negativas y un conjunto 3 ('Espectros_Perth' y 'Espectros_Santiago') donde sólo se eliminan columnas con valores nulos, este último permite estimar el impacto de la metodología de preprocesamiento aplicada en los resultados de los modelos.

8.3. Modelo supervisado

Anterior a la realización de los modelos supervisados, se deben normalizar los datos, codificar la variable categórica y dividir en conjuntos de entrenamiento, testeo y validación. Con respecto a la normalización de los datos, se utiliza la clase *StandardScaler* de la biblioteca *scikit-learn*, esta herramienta escala los datos de forma que se distribuyan normalmente, es decir, con media 0 y varianza 1. Por otro lado, codificar la variable categórica se refiere a convertir la columna categórica a valores 0 o 1, en este caso, se convierte la columna del contacto primario secundario, donde se asigna 0 para 'Secundario' y 1 para 'Primario'. Luego, se divide la base de datos en conjuntos de entrenamiento, testeo y validación para los modelos. Más aún, la práctica común es utilizar el set de entrenamiento para entrenar el modelo y obtener los hiperparámetros, luego acudir al set de validación para ajustar estos hiperparámetros y escoger el mejor modelo y finalmente, utilizar el set de testeo para evaluar el desempeño del modelo en datos no antes vistos. No obstante, para este proyecto se utiliza una metodología ligeramente distinta, primero, para la búsqueda de los mejores hiperparámetros se utilizan dos métodos (*RandomizedSearchCV* y *GridSearchCV* de la librería *scikit-learn*) y se evalúa el modelo en el conjunto de prueba. Posterior a elegir el mejor modelo según sus métricas, se procede a evaluar entre los distintos modelos (*Random Forest*, *Support Vector Machine*, *Logistic Regression* y Redes neuronales). Para lo anterior, se hace uso del conjunto de validación, que sería en sí un segundo conjunto de testeo.

En particular, cada modelo se entrena con los espectros analizados (en Perth o Santiago) y las etiquetas que 'supervisan' el entrenamiento, las que corresponden a la variable categórica del contacto primario-secundario. También, se realiza una primera búsqueda de los mejores hiperparámetros con la clase *RandomizedSearchCV*, esta técnica limita el tiempo de entrenamiento al reducir las combinaciones posibles de hiperparámetros. Con los resultados de esta primera búsqueda, se realiza una segunda búsqueda con la clase *GridSearchCV*, este es un método de búsqueda exhaustivo que toma en cuen-

ta todas las combinaciones de hiperparámetros y elige la combinación con el error más bajo. Lo anterior se realiza para comparar los mejores hiperparámetros obtenidos con cada método y estimar si es necesaria una búsqueda exhaustiva que implica un mayor tiempo de entrenamiento (caso al aplicar *GridSearchCV*) o si se obtienen resultados razonables con una búsqueda más arbitraria y que requiere menor tiempo computacional (caso al aplicar *RandomizedSearchCV*). Cabe mencionar que el orden de aplicación de algoritmos de búsqueda de hiperparámetros está pensado para que con *RandomizedSearchCV* sea una búsqueda más arbitraria y que, al aplicar una segunda búsqueda con *GridSearchCV*, se afine la búsqueda de los hiperparámetros encontrados en la primera búsqueda, ya que este método prueba todas las combinaciones posibles de los hiperparámetros ingresados. Por lo tanto, al encontrar los mejores hiperparámetros en la primera búsqueda con *RandomizedSearchCV*, estos se ingresan para una segunda búsqueda con *GridSearchCV*, tomando un rango de valores cercanos al óptimo propuesto para testar todas las combinaciones posibles, lo que a la vez reduce el tiempo computacional requerido para búsqueda finita con la clase *GridSearchCV*. Luego, una vez seleccionada la mejor combinación de hiperparámetros, se evalúa el modelo en el conjunto de prueba, de esto se obtienen las métricas a evaluar:

- **Matriz de confusión:** En el caso de clasificadores binarios, consiste en una matriz de 2x2 (figura 8.5), donde la primera columna proporciona el número de valores clasificados correctamente en la clase 'Primario' (Verdaderos Positivos o True Positive, TP) y valores clasificados erróneamente en la clase 'Secundario' (Falsos Positivos o False Positive, FP), mientras que la segunda columna proporciona el número de valores clasificados correctamente en la clase 'Secundario' (Verdaderos Negativos o True Negative, TN) y valores clasificados erróneamente en la clase 'Primario' (Falsos Negativos o False Negative, FN). A partir de esta matriz se pueden determinar exactitud, precisión, recall o sensibilidad.

		Clase predicha	
		Clase = Primario	Clase = Secundario
Clase real	Clase = Primario	TP	FN
	Clase = Secundario	FP	TN

Figura 8.5: Matriz de confusión para clasificación binaria. Elaboración propia.

- **Exactitud (*Accuracy*):** Es una medida de qué tan a menudo el modelo acierta sobre todas las clases y se calcula según la ecuación 8.1. Por lo tanto, una exactitud del

90 % significa que en 9 de cada 10 casos, el modelo clasifica correctamente.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (8.1)$$

- **Sensibilidad (*Recall*):** Mide cuántos elementos son clasificados en una categoría dentro del total de elementos en esa categoría, por lo tanto, indica la proporción de elementos de una categoría clasificados correctamente. Se puede calcular con la ecuación 8.2.

$$Recall = \frac{TP}{TP + FN} \quad (8.2)$$

- **Precisión (*Precision*):** Mide la proporción de identificaciones que fueron clasificadas correctamente en una categoría, es decir, el número de verdaderos positivos, dentro del total de clasificados en esa categoría. Puede interpretarse como la probabilidad de que, dado un elemento clasificado como 'Primario', esta clasificación sea correcta y se calcula con la ecuación 8.3.

$$Precision = \frac{TP}{TP + FP} \quad (8.3)$$

- **F1-score:** Permite comparar el rendimiento combinado de precisión y recall, se calcula como 2 veces la media armónica entre estos valores (ecuación 8.4). De lo anterior, se puede estimar que un clasificador perfecto tendrá $F1 = 1$, donde el modelo cuenta con alta precisión y alta sensibilidad, es decir, clasifica perfectamente esa clase. Si el modelo detecta bien la clase, pero incluye muestras de otra clase, entonces su sensibilidad es alta, pero su precisión baja. Si el modelo no detecta la clase muy bien, pero cuando lo hace es altamente confiable, entonces probablemente su precisión es alta a costas de una baja sensibilidad. En cambio, si el modelo no logra clasificar la clase correctamente, tanto su sensibilidad como precisión son bajas (Universidad Politécnica de Madrid, 2021).

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (8.4)$$

8.3.1. Bosque aleatorio

El método bosque aleatorio o *Random Forest* es un clasificador basado en árboles de decisión, de modo general, estos árboles de decisión se entrenan dividiendo iterativamente los datos hasta descubrir el método óptimo para agruparlos, mayor detalle en la sección 4.4. Para crear el modelo se utiliza la clase *RandomForestClassifier* de la librería *scikit-learn* de *Python*, con los siguientes hiperparámetros a ajustar:

- `n_estimators`: Especifica el número de árboles que se utilizarán en el bosque aleatorio. El valor predeterminado es 100.

- **criterion:** Determina la función de calidad que se utilizará para medir la calidad de una división, se tiene 'gini' para el índice Gini y 'entropy' para la ganancia de información. Por defecto, se utiliza 'gini'.
- **max_depth:** Controla la profundidad máxima de los árboles. Si se establece en 'None', los nodos se expandirán hasta que todas las hojas sean puras o hasta que contengan menos de 'min_samples_split' muestras. La profundidad máxima evita el sobreajuste. Por defecto, es 'None'
- **min_samples_split:** Especifica el número mínimo de muestras necesarias para dividir un nodo interno. Si una división da como resultado una hoja con menos muestras que 'min_samples_split', la división se cancela. Por defecto, es 2.
- **min_samples_leaf:** Especifica el número mínimo de muestras que debe tener una hoja. Por defecto, es 1.
- **max_features:** Determina el número máximo de características que se deben considerar al buscar la mejor división. El valor predeterminado es 'auto', que equivale a la raíz cuadrada del número total de características ('sqrt(n_features)').
- **class_weight:** Permite tratar de manera desigual a las clases, ideal para manejar conjuntos de datos desequilibrados. Las opciones son 'None' (todas las clases tienen el mismo peso), 'balanced' (ajusta automáticamente los pesos inversamente proporcionales a las frecuencias de clase en los datos de entrada) o un diccionario con pesos de clase específicos.
- **bootstrap:** Indica si se utilizará el muestreo con reemplazo durante la construcción de los árboles. Por defecto, es 'True'.

8.3.2. Máquina de vectores de soporte

El método máquina de vectores de soporte o *Support Vector Machine* es un clasificador que se basa en buscar el hiperplano óptimo que pueda separar todas las clases de datos etiquetados, donde los vectores de soporte son los puntos de datos más cercanos al hiperplano, mayor detalle en la sección 4.6. Para crear el modelo se utiliza la clase *SVC* de la librería *scikit-learn*, con los siguientes hiperparámetros a ajustar:

- **C:** Parámetro de regularización, controla cuánto quieres evitar clasificar erróneamente cada ejemplo de entrenamiento. Valores pequeños de 'C' permiten márgenes más amplios pero pueden llevar a más puntos mal clasificados (más regularización), mientras que valores grandes buscan clasificar todos los puntos correctamente a expensas de márgenes más estrechos y potencialmente a un sobreajuste (menos regularización).

- **kernel:** Define el tipo de kernel utilizado en el algoritmo SVC. Se tiene 'linear' para utilizar un kernel lineal, comúnmente utilizado para datos linealmente separables; 'poly' para kernel polinomial, útil para capturar relaciones lineales; 'rbf' es kernel de función de base radial, utilizado para relaciones no lineales y espacios de mayor dimensión; y 'sigmoid' basado en la función sigmoide. Por defecto, es rbf.
- **gamma:** Coeficiente que define hasta qué punto influye un solo punto de entrenamiento, valores bajos significan 'lejos' y valores altos significan 'cerca', así un 'gamma' alto puede llevar a un sobreajuste, ya que el modelo intentará ajustarse exactamente a los datos de entrenamiento. Utilizado para kernels no lineales ('rbf', 'poly' y 'sigmoid').
- **degree:** Indica el grado del polinomio en el kernel 'poly'. Un 'degree' más alto puede captar relaciones complejas, pero también incrementa el riesgo de sobreajuste.
- **class_weight:** Permite tratar de manera desigual a las clases, ideal para manejar conjuntos de datos desequilibrados. Las opciones son 'None' (todas las clases tienen el mismo peso), 'balanced' (ajusta automáticamente los pesos inversamente proporcionales a las frecuencias de clase en los datos de entrada) o un diccionario con pesos de clase específicos.

8.3.3. Regresión Logística

El método regresión logística o *Logistic Regression* es un clasificador que, utilizando una función logística, busca estimar la probabilidad de que una variable dependiente binaria pertenezca a una de las clases independientes, mayor detalle en la sección 4.5. Para crear el modelo se utiliza la clase *SGDClassifier* de la librería *scikit-learn*, con los siguientes hiperparámetros a ajustar:

- **penalty:** Especifica la norma utilizada en la penalización. Puede ser 'l2' para regularización L2 o ridge, 'l1' para regularización L1 o lasso o 'elasticnet' para una combinación de L2 y L1.
- **max_iter:** Número máximo de iteraciones durante el entrenamiento. Si no converge antes de alcanzar el número de iteraciones, se detendrá. Se debe ajustar para garantizar que el modelo converja sin tomar demasiado tiempo.
- **learning_rate:** Tasa de aprendizaje, controla el tamaño de los pasos que el algoritmo toma durante el proceso de optimización. Puede ser constante 'constant', adaptativo 'adaptive' u óptimo 'optimal', este último significa que la tasa se ajusta durante el entrenamiento para mejorar el rendimiento.
- **eta0:** Tasa inicial de aprendizaje, sólo se utiliza si 'learning_rate' se establece en 'constant'. Es un valor flotante positivo que especifica el tamaño del paso inicial.

- `class_weight`: Peso de clase, se utiliza para abordar el desequilibrio de clases en el conjunto de datos. Puede ser 'None', un diccionario con los pesos específicos por clase o 'balanced', este último ajusta automáticamente los pesos de clase inversamente proporcional a las frecuencias de clase.
- `alpha`: Parámetro de regularización que controla la fuerza de esta. Cuanto mayor sea su valor, mayor será la regularización aplicada al modelo. Es un valor flotante positivo.

8.3.4. Redes neuronales

El método de redes neuronales o *Neuronal Network* procesa los datos de una manera que emula la forma en que lo hace el cerebro humano. Estas redes se componen, básicamente, de capas: capas de entrada, capas ocultas y capas de salida, mayor detalle en la sección 4.7. Para construir los modelos, se utiliza la librería *Keras*, implementando una función generadora de modelos que configura cada red con varias capas: una capa de entrada que recibe datos, múltiples capas densas con activación ReLU (*Rectified Liner Unit*) y una capa de salida con activación sigmoidea. Con respecto a la capa de activación ReLU, es la función de activación más usada en la actualidad (Huet, 2023), ya que no tiene problemas de saturación como las funciones sigmoidea y tangente hiperbólica, y, por consiguiente, se recomienda el uso de la función sigmoideal únicamente en las capas de salida, siendo principalmente utilizada para clasificación binaria. En cuanto a la optimización de los hiperparámetros, se realiza con el sintonizador *Hyperband* de la librería *Keras Tuner*, esta técnica se basa en los algoritmos de tipo bandido multi-brazo y se enfoca en maximizar la eficiencia tanto en la asignación de recursos como en el rendimiento del modelo. El proceso comienza configurando el sintonizador para buscar la mejor combinación de hiperparámetros, incluyendo el número de neuronas y la tasa de aprendizaje del optimizador Adam. También, se establece un objetivo de maximizar la precisión de validación (`objective = 'val_accuracy'`), con un máximo de 10 épocas (`max_epochs = 10`) y un factor de reducción de 3 (`factor = 3`). Finalmente, la búsqueda de los mejores hiperparámetros se lleva a cabo a través de entrenamientos que duran hasta 50 épocas, utilizando los datos de entrenamiento y validación.

En relación con el entrenamiento del modelo, se lleva a cabo con un enfoque de detención anticipada (clase *EarlyStopping*) para prevenir el sobreajuste, monitoreando la precisión de validación (`monitor = 'val_accuracy'`) y deteniendo el entrenamiento si no hay mejoras después de un número predefinido de épocas (`patience = 10`, por ejemplo). Lo anterior con miras a asegurar que el modelo no sólo se ajuste a los datos de entrenamiento sino que también sea capaz de generalizar a datos no antes vistos. Finalmente, el modelo se evalúa en el conjunto de prueba para calcular las métricas de evaluación.

8.4. Modelo no supervisado

Para comparar con las técnicas de aprendizaje supervisado, se construyen modelos utilizando el método *K-means*, este un método iterativo de aprendizaje no supervisado que agrupa los datos en un número k predefinido de clústeres, basándose en la minimización de una métrica de distancia (por ejemplo, distancia euclidiana) de los datos a cada clúster, mayor detalle en la sección 4.2.1. Previa construcción de los modelos, se normalizan los datos con la clase *StandardScaler* de la biblioteca *scikit-learn* y se codifica la variable categórica del contacto primario-secundario, con tal de mantener la cohesión, se utilizan los datos obtenidos de estas etapas desde la sección anterior. Luego, para determinar el número óptimo de clústeres, se emplea el coeficiente de silueta, que evalúa cuán compactos y separados están los clústeres, donde valores cercanos a 1 indican buena división de los clústeres. También, se considera la homogeneidad y completitud de los mismos, buscando que ambos valores sean cercanos a 1. Así, se evalúa un rango de clústeres desde 2 a 10 y se escoge el número que tiene mejores resultados en cuanto a coeficiente de silueta, homogeneidad (*homogeneity*) y completitud (*completeness*). Una vez seleccionado el número óptimo de clústeres, se procede a construir el modelo con la clase *KMeans* de la librería *scikit-learn*. Finalmente, para evaluar el modelo se calculan las métricas de homogeneidad, completitud y *V-measure*, más la matriz de confusión, todo lo anterior para evaluar el rendimiento del modelo en función de los valores reales.

8.5. Comparación de modelos

Una vez generados los modelos supervisados (*Random Forest*, *Logistic Regression*, *Support Vector Machine* y Redes neuronales) y no supervisado (*K-Means*), estos se comparan en función de sus métricas de rendimiento, complejidad y costo computacional. En este paso de la metodología, se utiliza el conjunto de validación para evaluar la capacidad de generalización de los modelos al tratar con datos no antes vistos. Si los modelos rinden significativamente peor en el conjunto de validación, podría ser una señal de sobreajuste. En el caso de la red neuronal, dado que el conjunto de validación se utiliza anteriormente para ajustes del modelo, no sería correcto reutilizarlo para evaluaciones finales. Por lo tanto, se evalúan los modelos de *Random Forest*, *Logistic Regression* y *Support Vector Machine* en el conjunto de validación, para luego comparar las métricas del mejor modelo con las métricas de la red neuronal, ambos evaluados en el conjunto de prueba.

Cabe mencionar que las comparaciones se deben realizar de manera separada para los modelos desarrollados con datos de Perth y de Santiago. No es correcto comparar los modelos de muestras analizadas en diferentes laboratorios, ya que no sólo se desconocen los instrumentos utilizados y las condiciones de medición, si no que los modelos son entrenados con bases de datos distintas.

8.6. Implementación

Una vez seleccionado el mejor modelo, tanto para Perth como Santiago, se ingresan nuevos datos para su clasificación y se compara con el modelo actual del contacto primario-secundario. Esto implica integrar los datos al modelo 3D de *Leapfrog Geo* y modelar los dominios primario y secundario obtenidos con los modelos a partir de los 'Espectros ASD-FTIR'. Por último, se debe evaluar la desviación con el modelo actual que está basado en la descripción de sondajes.

CAPÍTULO 9

RESULTADOS

9.1. Preprocesamiento de los datos

9.1.1. Sondajes

9.1.1.1. Filtro 1: Mantener litologías con 100 %

Al eliminar los tramos de muestras con menos del 100 % de una litología, se crea una nueva base de datos que sólo posee tramos con litologías representativas a su muestra asociada, pasando de 21, 848 a 15, 673 filas, eliminándose aproximadamente el 28 % de tramos de la base original. El cambio en la distribución de los datos se puede ver en la figura 9.1, donde claramente se ve una disminución en la cantidad de tramos para cada litología.

9.1.1.2. Composición

Con el propósito de contar con información conjunta de litología y del contacto, se construye un nuevo archivo derivado de compositar el archivo de litologías, generado con el filtro anterior, y el archivo del contacto. Para lo anterior, se utiliza la herramienta *Majority Composites* del software *Leapfrog Geo* que permite llevar datos categóricos a longitudes de intervalos de otra tabla, es decir, permite llevar al mismo soporte dos archivos. En este caso, se toman los intervalos del archivo del contacto para compositar el archivo de litologías, luego la columna con la información de litología se añade al archivo original del contacto ('FTIR_Sondajes'), construyendo un nuevo archivo que llamaremos *Merged_table*. Finalmente, se cuenta con un archivo que tiene tanto información del mapeo de litologías como del contacto primario-secundario para las muestras de la mina Los Bronces, con un total de 6135 datos de primario y 8295 de secundario. Con esto, es posible visualizar la distribución de litologías en cada dominio, donde se observa 1) presencia dominante de QM en ambos dominios, 2) BXPR, BXA, BXBM mayormente presentes en el dominio primario que secundario (figura 9.2), 3) BXA con baja presencia en secundario, 4) RIO y AND más frecuentes en secundario que primario (figura 9.3), 5) baja presencia de BXI en

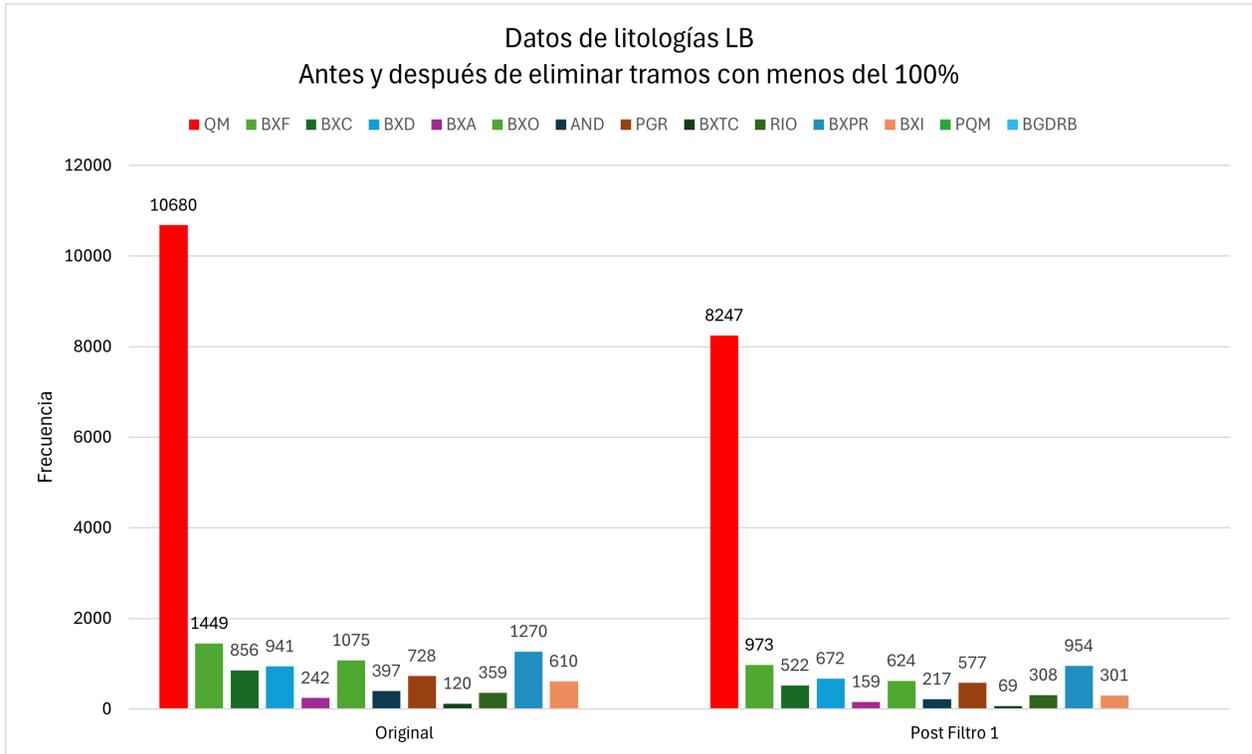


Figura 9.1: Distribución en la cantidad de tramos para cada litología presente, antes y después de eliminar las filas con menos del 100% de la litología asociada.

primario, y 6) baja frecuencia de PQM y BXGF en ambos dominios.

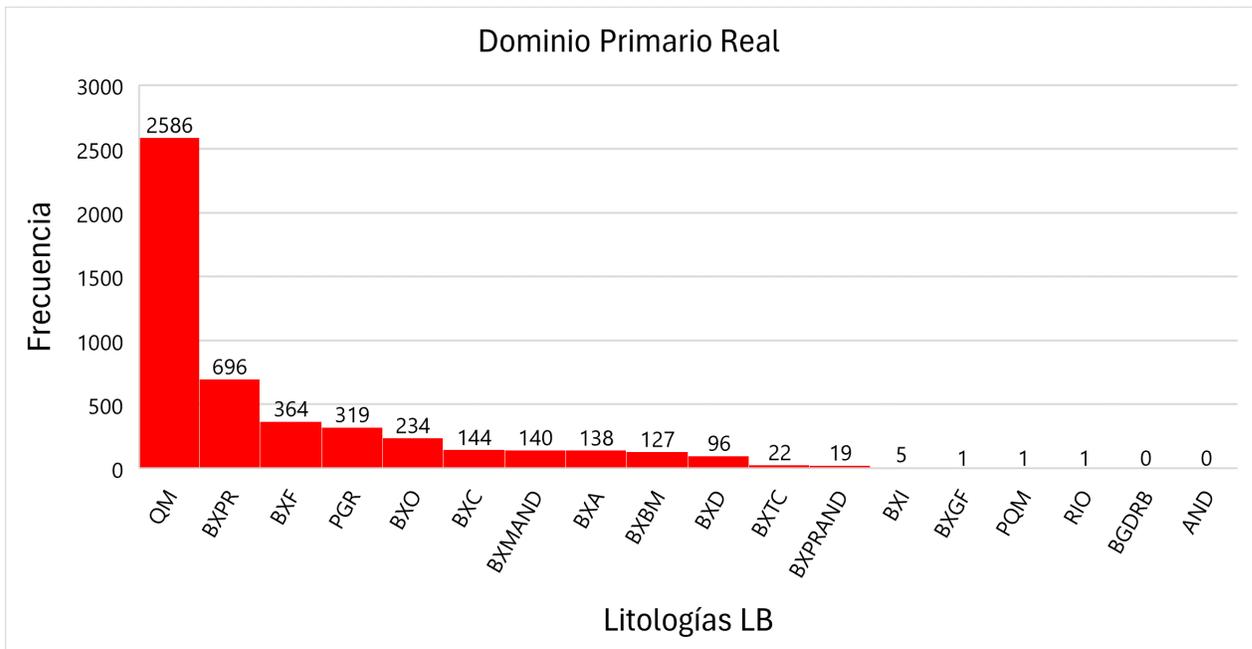


Figura 9.2: Distribución de las litologías presentes en el dominio primario, basado en los datos de mapeo de la mina Los Bronces.

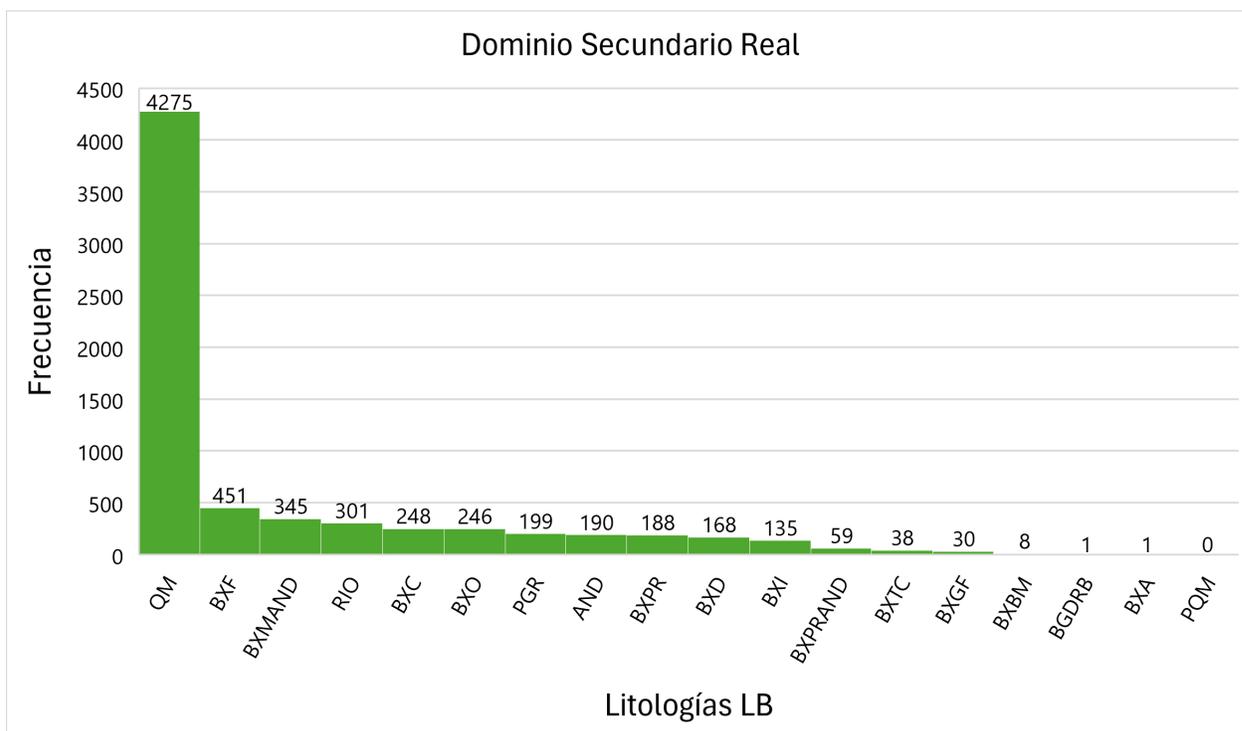


Figura 9.3: Distribución de las litologías presentes en el dominio secundario, basado en los datos de mapeo de la mina Los Bronces.

9.1.1.3. Filtro 2: Eliminar filas sin SampleID

La asociación con los espectros ASD-FTIR sólo es posible si se tiene el nombre de la muestra analizada (*SampleID*), por lo tanto, si algún tramo no presenta *SampleID* entonces no se podrá identificar su espectro. De esta manera, se eliminan 4 487 filas de *Merged_table* por no presentar *SampleID*.

9.1.1.4. Filtro 3: Separar en muestras de Perth y Santiago

Para la realización de los modelos, se deben separar las muestras analizadas en distintos laboratorios, ya que, si bien pertenecen al mismo yacimiento, las condiciones en las cuales se midieron y analizaron los espectros no son exactamente iguales. Por lo tanto, una buena práctica es realizar un modelo para las muestras analizadas en Perth y otro para las muestras analizadas en Santiago. Bajo la misma idea, se separa *Merged_table* en dos archivos: 'Perth' y 'Santiago', donde 'Perth' y 'Santiago' cuentan sólo con las muestras analizadas en los laboratorios de Perth y Santiago respectivamente. Así, se generan dos nuevas bases de datos, con 9 094 muestras en el caso de 'Perth', con 2 857 muestras de Primario y 6 092 muestras de Secundario, y un total de 5 506 muestras en el caso de 'Santiago', con 3 255 de Primario y 2 136 de Secundario.

9.1.1.5. Filtro 4: Ajustar nombre de las muestras a la nomenclatura solicitada

Para este proyecto se establece una nomenclatura para *SampleID*, esta consiste en un código referente al tipo de muestra, seguido de un número de 5 dígitos. Más aún, si el número que sigue al código no tiene 5 dígitos, entonces se debe agregar un 0 delante. Por ejemplo, la muestra 'MET_04316' corresponde a una muestra metalúrgica (MET) número 4316. Este proceso se realiza tanto para las muestras de sondajes como para los espectros ASD-FTIR, a través de la rutina 9.1, la que sobrescribe el nombre de cada muestra en la columna de muestras del archivo original.

Código 9.1: Función para ajustar el nombre de las muestras a la nomenclatura solicitada

```
1 import numpy as np
2
3 # Función para ajustar el nombre de las muestras
4 def ajustar_nombre_muestra(nombre):
5     partes = nombre.split('-')
6     if len(partes) > 1 and len(partes[1]) < 5:
7         partes[1] = '0' + partes[1]
8         return '-'.join(partes)
9     else:
10        return nombre
11
12 # Crear una lista con los archivos a ajustar el nombre de las muestras
13 variables_samples= [Perth, Santiago]
14
15 # Aplicar la función a la columna SampleID en cada variable
16 for df in variables_samples:
17     df.loc[:, 'SampleID'] = df['SampleID'].apply(ajustar_nombre_muestra)
```

9.1.1.6. Filtro 5: Separar muestras sin espectro

Tal como se mencionó anteriormente, las muestras cuyo nombre (*SampleID*) no esté presente en la columna de muestras de los archivos de Espectros ASD-FTIR, no tendrán un espectro asociado y, por consiguiente, no podrán ser utilizadas en los modelos, ni en el entrenamiento de estos o para predicciones posteriores con estos. Por lo tanto, se decide separar nuevamente la base de datos en función de la columna de muestras de los Espectros ASD-FTIR Perth y Espectros ASD-FTIR Santiago, según corresponda. De modo que se generan cuatro nuevas variables:

- **Perth_Sample_coincidentes** : 6 195 muestras de sondajes analizadas en Perth que tienen espectro en Espectros ASD-FTIR Perth.
- **Perth_Sample_no_coincidentes** : 3 202 muestras de sondajes analizadas en Perth

que **no** tienen espectro en Espectros ASD-FTIR Perth.

- **Santiago_Sample_coincidentes** : 5 050 muestras de sondajes analizadas en Santiago que tienen espectro en Espectros ASD-FTIR Santiago.
- **Santiago_Sample_no_coincidentes** : 521 muestras de sondajes analizadas en Santiago que **no** tienen espectro en Espectros ASD-FTIR Santiago.

9.1.1.7. Filtro 6: Separar muestras sin contacto Primario/Secundario

Para los modelos es esencial contar con muestras que tengan tanto espectro como contacto conocido, ya que permite 'supervisar' el aprendizaje y analizar el rendimiento de los modelos. Por lo anterior, se deben separar las muestras que sí presentan el contacto, creándose cuatro variables:

- **Perth_Sample_Contacto** : Conjunto de 6 176 muestras analizadas en Perth que presentan contacto y espectro asociado.
- **Perth_Sample_Sin_Contacto** : Conjunto de 19 muestras analizadas en Perth que **no** presentan contacto ni espectro asociado.
- **Santiago_Sample_Contacto** : Conjunto de 5 040 muestras analizadas en Santiago que presentan contacto y espectro asociado.
- **Santiago_Sample_Sin_Contacto** : Conjunto de 10 muestras analizadas en Santiago que **no** presentan contacto ni espectro asociado.

9.1.2. Espectros ASD-FTIR

9.1.2.1. Estadísticas

Se visualizan las estadísticas básicas de Espectros ASD-FTIR Perth y Espectros ASD-FTIR Santiago en la Tabla 9.1 y Tabla 9.2 respectivamente. Cabe mencionar que sólo se muestran las tres primeras y tres últimas columnas de cada archivo, ya que se tienen más de 3 000 columnas por archivo.

Por un lado, para Espectros ASD-FTIR Perth se tienen 14 409 muestras (filas) con valores de reflectancia en 3 303 longitudes de onda (columnas), de las cuales 6 columnas (tabla 9.3) presentan valores nulos o en blanco, por lo que sólo 3 297 columnas aportan información. Considerando lo anterior y que representan el 0.18 % del total de columnas, se eliminan las columnas vacías o con valores nulos (*NaN*). En cuanto al análisis de sus estadísticas, se nota un aumento significativo en el valor de desviación estándar hacia longitudes de onda mayores, lo que se observa en el extremo derecho del gráfico de la figura 8.1, coincidente con el drástico aumento en los valores máximos y valores absoluto de los valores mínimos de las longitudes de onda. En el caso de Espectros ASD-FTIR Santiago, presenta 57 562 muestras con valores de reflectancia en 3 300 longitudes de

onda, no tiene columnas vacías ni con valores nulos, por lo que no se modifican los datos. En relación con sus estadísticas, de manera análoga al caso de Perth, se observa un aumento significativo en los valores de desviación estándar hacia longitudes de onda más largas, congruente con el aumento disparado de valores máximos y valor absoluto de valores mínimos, observable en la distorsión de las curvas hacia el extremo derecho en la figura 8.2. Finalmente, tanto en Espectros ASD-FTIR Perth como Espectros ASD-FTIR Santiago, se observa una disminución de los cuantiles 25 %, 50 % y 75 % hacia longitudes de onda mayores, lo que es congruente con el comportamiento de las ondas electromagnéticas, donde a longitudes de onda más largas, se tiene menor energía y puede ocasionar distorsión en la medición de reflectancia por limitaciones del instrumento.

Tabla 9.1: Estadísticas básicas de Espectros ASD-FTIR Perth, donde en cada fila se tienen los valores del promedio (*mean*), la desviación estándar (*std*), mínimo (*min*), máximo (*max*), cuantiles 25 %, 50 % y 75 % para cada longitud de onda de la columna.

	Longitud de onda [nm]						
Estadísticas	350	351	352	...	43211.54	43943.93	44701.59
mean	0.35	0.35	0.35		0.09	0.08	0.11
std	0.08	0.08	0.08		0.40	0.89	2.35
min	0.10	0.09	0.10		-7.99	-93.49	-62.08
max	0.65	0.64	0.64		9.79	9.24	126.48
25 %	0.29	0.30	0.29		0.016	0.008	-0.001
50 %	0.33	0.34	0.33		0.07	0.07	0.07
75 %	0.40	0.40	0.40		0.13	0.14	0.15

Tabla 9.2: Estadísticas básicas de Espectros ASD-FTIR Santiago, donde en cada fila se tienen los valores del promedio (*mean*), la desviación estándar (*std*), mínimo (*min*), máximo (*max*), percentil 25%, 50% y 75% para cada longitud de onda de la columna.

Estadísticas	Longitud de onda [nm]						
	350	351	352	...	43211.54	43943.93	44701.59
mean	0.33	0.33	0.33		0.003	0.06	0.98
std	0.04	0.04	0.04		15.05	8.46	41.78
min	0.01	0.01	0.01		-1772.47	-543.51	-907.33
max	0.57	0.56	0.57		1812.05	465.55	3049.03
25 %	0.30	0.31	0.31		0.03	0.02	0.014
50 %	0.33	0.34	0.34		0.08	0.08	0.173
75 %	0.36	0.36	0.37		0.13	0.15	0.15

Tabla 9.3: Columnas eliminadas de Espectros ASD-FTIR Perth.

Columnas eliminadas de Espectros ASD-FTIR Perth
['4209.813', '9393.813', '18787.63', '4209.812', '9398.812', '18787.62']

9.1.2.2. Tratamiento de negativos

Como se mencionó en la metodología, al desconocer las condiciones de laboratorio así también como los instrumentos de medición utilizados, sumado a no contar con espectros representativos de señal, ruido y materiales, no es posible realizar un preprocesamiento de señales típico. Por lo tanto, en este paso de la metodología propuesta se analizan los valores negativos presentes en Espectros ASD-FTIR Perth y Espectros ASD-FTIR Santiago.

En cuanto a Espectros ASD-FTIR Perth, se observa la distribución de valores negativos de reflectancia en la figura 9.4, donde es notable un aumento en la frecuencia de valores negativos hacia valores extremos de longitud de onda. Mismo fenómeno se aprecia en la

figura 9.5 con la distribución de valores negativos de Espectros ASD-FTIR Santiago.

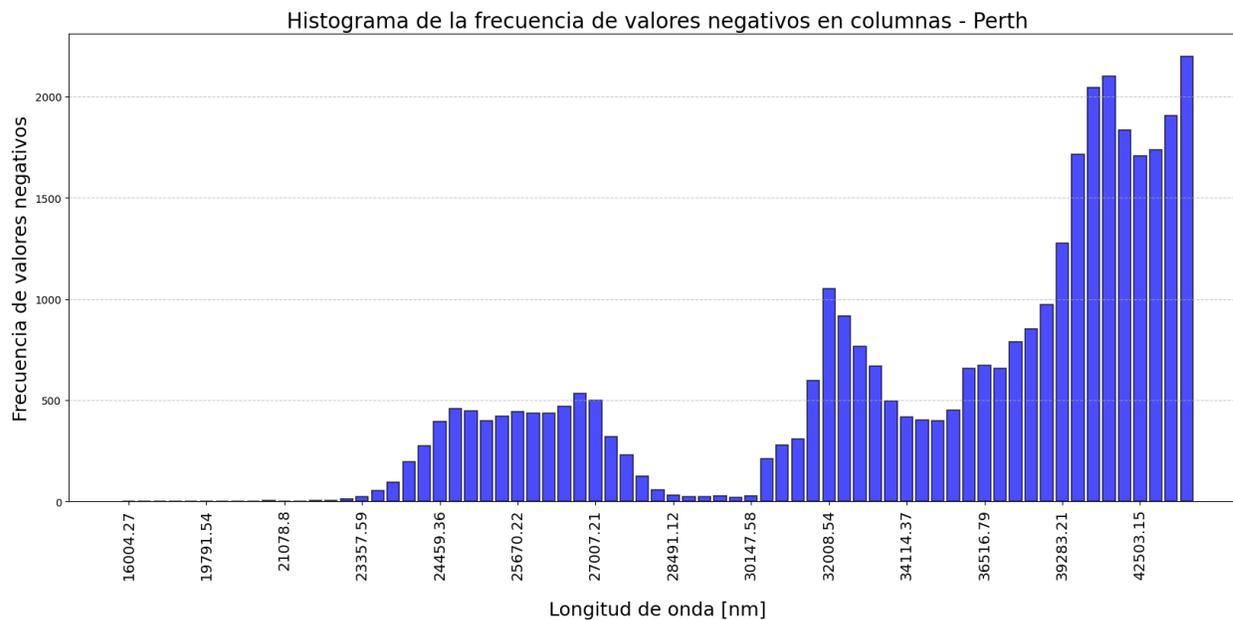


Figura 9.4: Distribución de los valores negativos según longitud de onda en Espectros ASD-FTIR Perth (muestras analizadas en laboratorios de Perth).

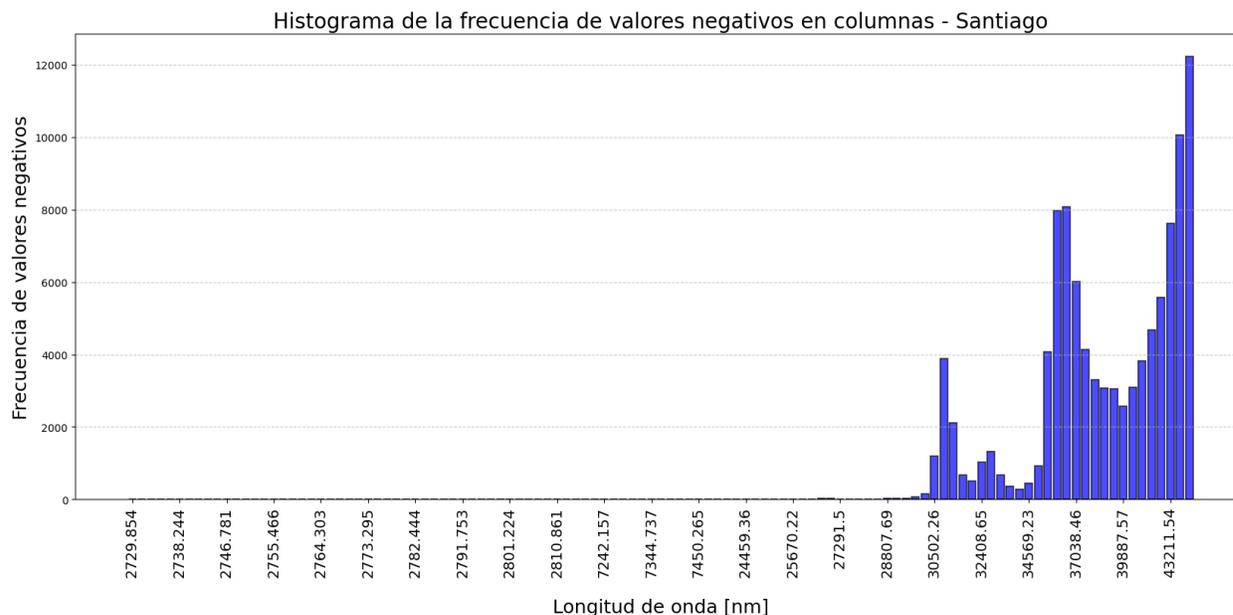


Figura 9.5: Distribución de los valores negativos según longitud de onda en Espectros ASD-FTIR Santiago (muestras analizadas en laboratorios de Santiago).

Ante esta situación, se presentan tres escenarios para tratar estos datos negativos:

1. **Eliminar filas con valores negativos:** este escenario consiste en a) eliminar 9 396

filas de Espectros ASD-FTIR Perth, lo que representa eliminar el 65 % de las muestras analizadas en Perth; y b) eliminar 26 959 filas de Espectros ASD-FTIR Santiago, lo que representa eliminar el 47 % del universo de muestras analizadas en Santiago. De este escenario, se crean dos nuevos conjunto de datos, Perth_1 y Santiago_1, donde el gráfico de las curvas espectrales de las muestras se observan en las figuras 9.6 y 9.7, respectivamente. No obstante, aún se aprecia distorsión de las curvas hacia los extremos, derivado de valores de reflectancia sobre 1.

2. **Eliminar columnas con valores negativos:** consiste en a) eliminar 70 columnas de Espectros ASD-FTIR Perth, lo que representa eliminar un 2 % del total de columnas; y b) eliminar 113 columnas de Espectros ASD-FTIR Santiago, lo que representa eliminar un 3.4 % del total de columnas. De este escenario, se crean dos nuevos conjuntos de datos, Perth_2 y Santiago_2, donde el gráfico de las curvas espectrales de las muestras se observan en las figuras 9.8 y 9.9. En contraste con el escenario anterior, se eliminan los valores extremos que distorsionan los gráficos de las curvas, lo que se puede asociar a que al eliminar columnas, también se reduce el rango de longitudes de onda medido para cada muestra, donde se pasan de máximos de 44 701.59 a 22159.76 nm.
3. **No eliminar filas ni columnas con valores negativos:** consiste en conservar tanto filas como columnas con valores negativos. De este escenario, se crean dos nuevos conjuntos de datos, Espectros_Perth y Espectros_Santiago. En cuanto a los gráficos de las curvas espectrales de cada conjunto, se mantienen los de la sección de metodología (figuras 8.1 y 8.2).

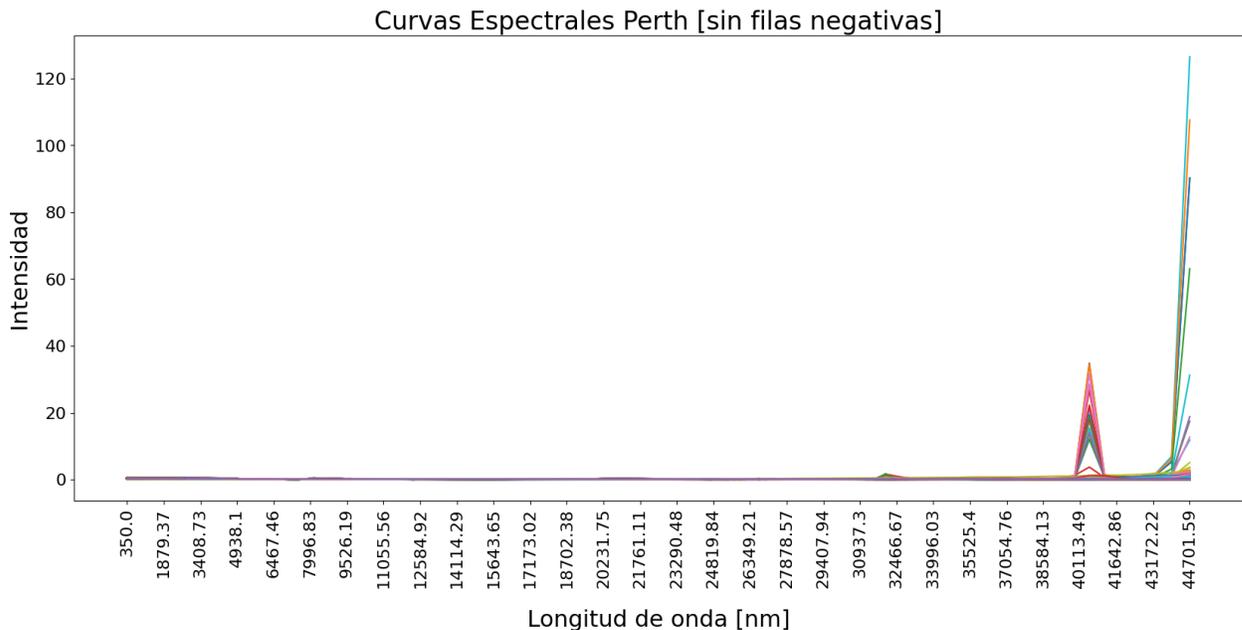


Figura 9.6: Curvas espectrales de las muestras de Perth_1, resultante de eliminar filas con valores negativos de Espectros ASD-FTIR Perth.

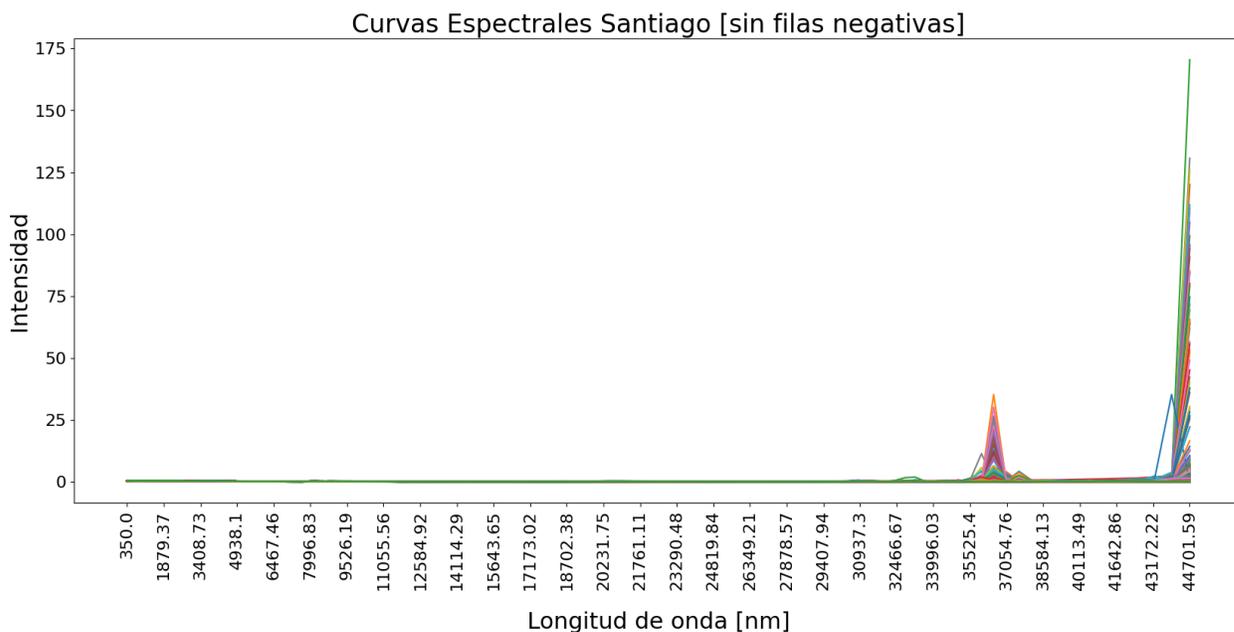


Figura 9.7: Curvas espectrales de las muestras de Santiago_1, resultante de eliminar filas con valores negativos de Espectros ASD-FTIR Santiago.

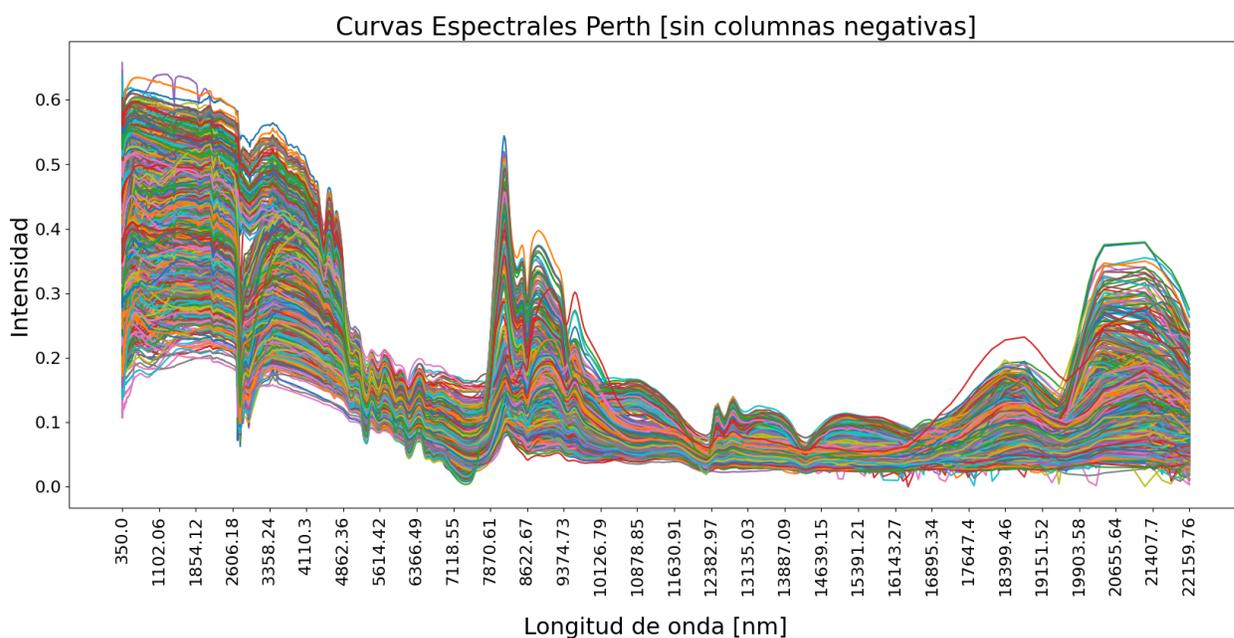


Figura 9.8: Curvas espectrales de las muestras de Perth_2, resultante de eliminar columnas con valores negativos de Espectros ASD-FTIR Perth.

9.1.2.3. Tratamiento de reflectancia sobre 1

Se tiene que por definición los valores de reflectancia se mueven entre 0 y 1, donde 0 indica absorción total y 1 indica sin absorción, por lo que valores fuera de estos límites se consideran como erróneos. Sin embargo, en el escenario 1 expuesto anteriormente, se identifican 113 muestras y 11 longitudes de onda (columnas) de Perth_1 con valores

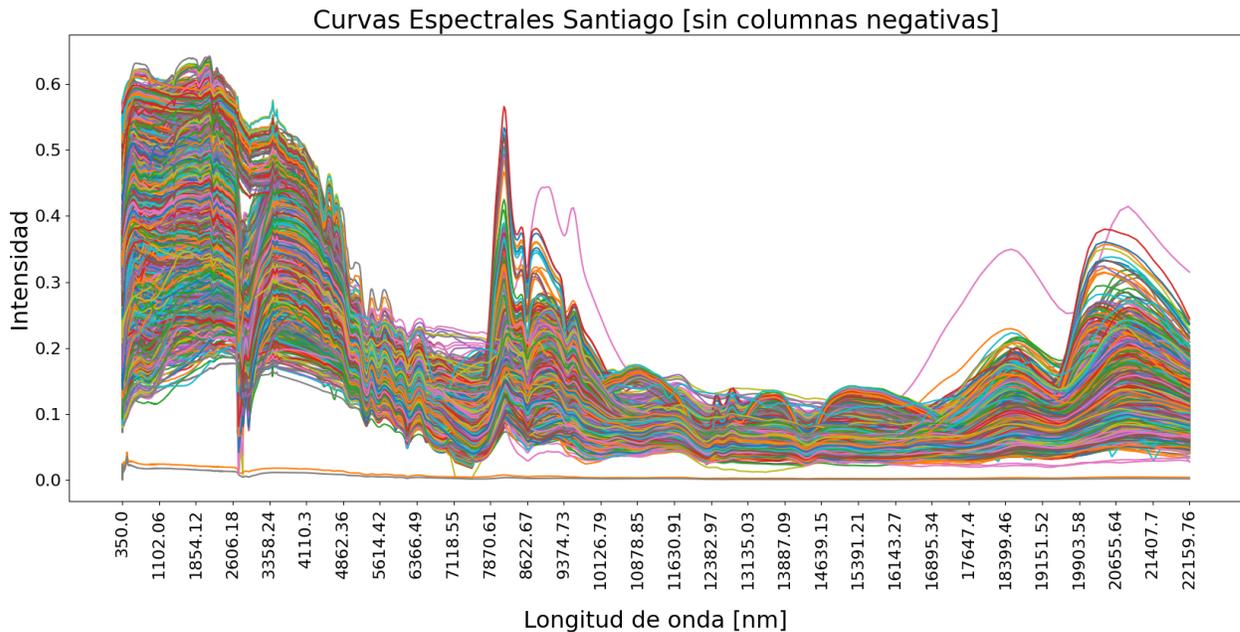


Figura 9.9: Curvas espectrales de las muestras de Santiago_2, resultante de eliminar columnas con valores negativos de Espectros ASD-FTIR Santiago.

de reflectancia sobre el valor límite. Del mismo modo, en Santiago_1 se reconocen 780 muestras y 14 longitudes de onda en la misma situación. En vista de lo anterior, se tienen dos opciones 1) eliminar las muestras (filas) con valores fuera de límites o 2) eliminar las longitudes de onda (columnas) con al menos un valor de reflectancia sobre 1. Finalmente, se decide eliminar las columnas, acotando el rango de longitudes de onda desde 350 hasta 38,696.9 nm para Perth_1 (figura 9.10) y 39,887.57 nm para Santiago_1 (figura 9.11). Cabe mencionar que la decisión anterior se basa en que 1) eliminar columnas representa la opción que conserva la mayor cantidad de información y 2) aún se notan anomalías en los gráficos de las curvas hacia longitudes de onda extremas.

9.1.2.4. Filtro 6: Separar muestras sin espectros

Este filtro consiste en separar las muestras de cada conjunto de datos que no presenten coincidencias en nombre (*Sample*) con Perth_Sample_Contacto, Perth_Sample_Sin_Contacto, Santiago_Sample_Contacto y Santiago_Sample_Sin_Contacto, generando cuatro nuevos conjuntos de datos para cada escenario, detallados a continuación:

- **Perth_Spectrum_Con_Contacto** : Conjunto de muestras analizadas en Perth que presentan contacto y espectro asociado. El nombre de este conjunto se mantiene para el escenario 3, sin embargo, para los dos escenarios restantes se encapsulan los datos resultantes en Perth_1_Spectrum_Con_Contacto y Perth_2_Spectrum_Con_Contacto.
- **Perth_Spectrum_Sin_Contacto** : Conjunto de muestras analizadas en Perth que **no** presentan contacto ni espectro asociado. De la misma manera, se mantiene el

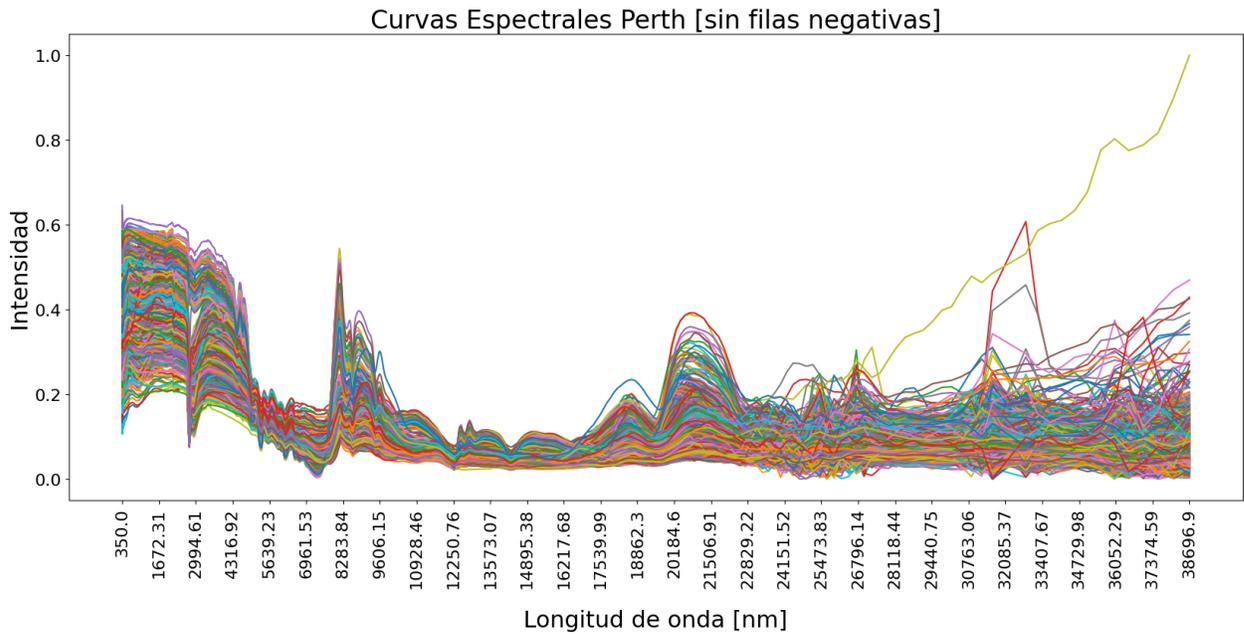


Figura 9.10: Curvas espectrales de las muestras de Perth_1, resultante de eliminar columnas con valores de reflectancia sobre 1.

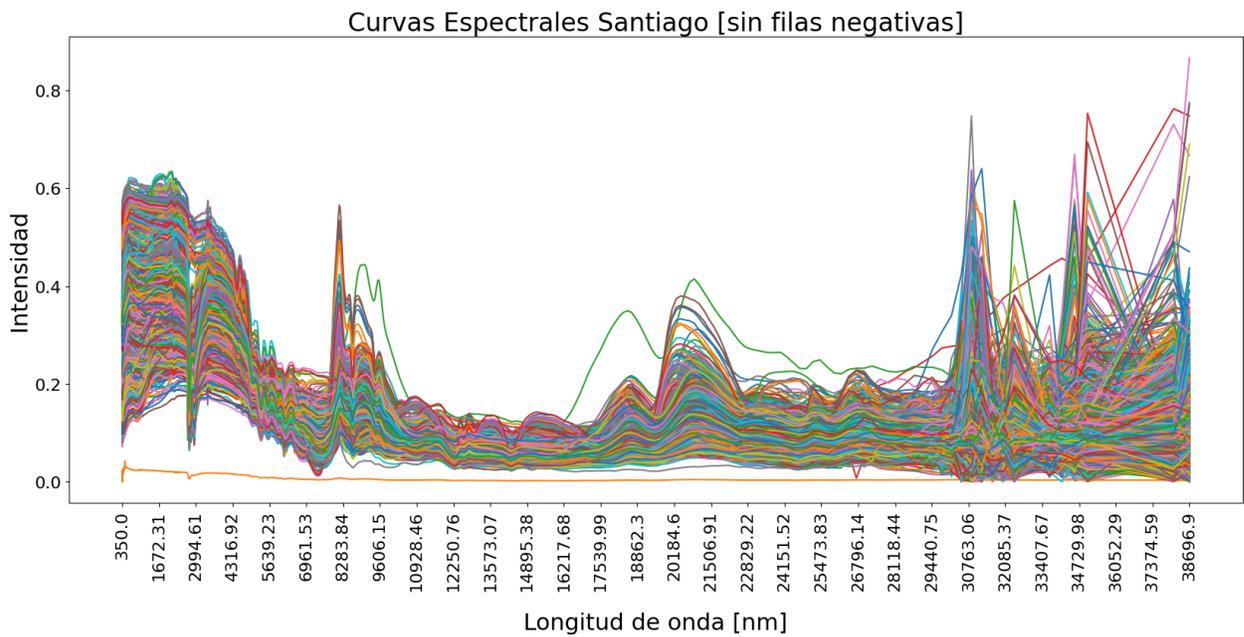


Figura 9.11: Curvas espectrales de las muestras de Santiago_1, resultante de eliminar columnas con valores de reflectancia sobre 1.

nombre para el escenario 3, mientras que para los dos escenarios restantes se nombran como Perth_1_Spectrum_Sin_Contacto y Perth_2_Spectrum_Sin_Contacto.

- **Santiago_Spectrum_Contacto** : Conjunto de muestras analizadas en Santiago que presentan contacto y espectro asociado. Del mismo que los conjuntos anteriores, el nombre se mantiene para el escenario 3, sin embargo, para los dos escenarios res-

tantes se encapsulan los datos resultantes en `Santiago_1_Spectrum_Con_Contacto` y `Santiago_2_Spectrum_Con_Contacto`.

- **Santiago_Spectrum_Con_Contacto** : Conjunto de muestras analizadas en Santiago que **no** presentan contacto ni espectro asociado. El nombre de este conjunto se mantiene para el escenario 3, sin embargo, para los dos escenarios restantes se encapsulan los datos resultantes en `Santiago_1_Spectrum_Sin_Contacto` y `Santiago_2_Spectrum_Sin_Contacto`.

9.1.2.5. Integrar columna del contacto

Inmediatamente posterior al filtro 6, se integra la columna del contacto primario-secundario presente en `Perth_Sample_Contacto` y `Santiago_Sample_Contacto` a cada archivo generado con el filtro anterior, dando como resultado las siguientes variables:

- Conjunto de espectros de muestras analizadas en laboratorios de Perth que cuentan con contacto reportado:
`Perth_Spectrum_Con_Contacto`, `Perth_1_Spectrum_Con_Contacto` y `Perth_2_Spectrum_Con_Contacto`.
- Conjunto de espectros de muestras analizadas en laboratorios de Santiago que cuentan con contacto reportado:
`Santiago_Spectrum_Con_Contacto`, `Santiago_1_Spectrum_Con_Contacto` y `Santiago_2_Spectrum_Con_Contacto`.

Finalmente, se adjunta una tabla resumen (tabla 9.4) con las variables finales a ser utilizadas en pasos posteriores.

Tabla 9.4: Conjuntos de datos derivados del preprocesamiento de los datos de Espectros ASD-FTIR Perth y Santiago, corresponden a los *inputs* de los modelos a crear.

Conjunto de datos	Nº muestras	Rango λ 's
	(filas)	(columnas) [nm]
Perth_Spectrum_Con_Contacto	6176	350 - 44701.59
Perth_1_Spectrum_Con_Contacto	1518	350 - 38696.90
Perth_2_Spectrum_Con_Contacto	6176	350 - 38696.90
Santiago_Spectrum_Con_Contacto	5040	350 - 44701.59
Santiago_1_Spectrum_Con_Contacto	2736	350 - 39887.57
Santiago_2_Spectrum_Con_Contacto	5040	350 - 39887.57

9.2. Modelo supervisado

9.2.1. Preparación de los datos

Posterior al preprocesamiento, se deben preparar los datos para ser ingresados a los modelos, esto consta de tres pasos que se detallan a continuación

9.2.1.1. Normalización de los datos

Esta transformación se utiliza para ajustar los valores a una distribución normal, como se puede observar en el cambio en la distribución de Perth antes y después del ajuste (figura 9.12). En particular, para este proyecto se utiliza la clase *StandardScaler* de la librería *scikit-learn* para llevar a cabo la normalización mediante la rutina 9.2.

Código 9.2: Función para ajustar los datos utilizando la clase *StandardScaler*

```
1 from sklearn.preprocessing import StandardScaler
2
3 # Preparar los datos
4 #X_Perth: Columnas de las longitudes de onda
5 #y_Perth: Columna del contacto
6 X_Perth = Perth_Spectrum_Con_Contacto_new.iloc[:, 1:-1]
7 y_Perth = Perth_Spectrum_Con_Contacto_new['Primario_Secundario_2024']
8
9 # Normalizar los datos
10 scaler = StandardScaler()
11 X_scaled_Perth = scaler.fit_transform(X_Perth)
```

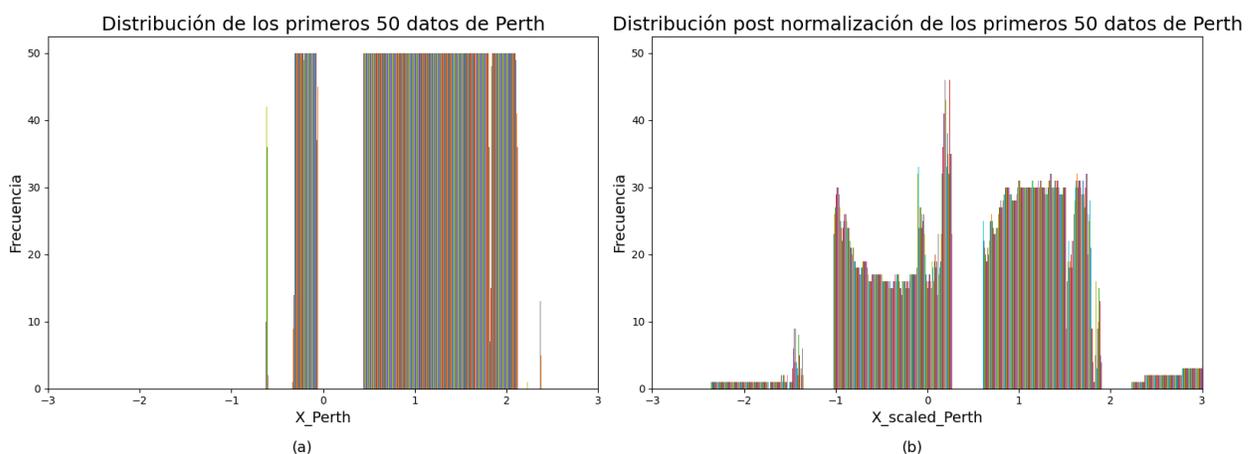


Figura 9.12: Distribución de los 50 primeros valores de a) *X_Perth* y b) *X_scaled_Perth*, donde *X_Perth* corresponde a los valores antes de ser normalizados, mientras que *X_scaled_Perth* presenta los datos transformados.

9.2.1.2. Codificar la variable categórica

Este paso consiste en codificar la columna del contacto primario secundario en 0 y 1, donde 0 se asigna para Secundario y 1 para Primario, mediante la rutina 9.3.

Código 9.3: Función para codificar la variable categórica del contacto primario secundario

```
1 #Rutina para codificar la variable categórica
2 # 'y_Perth' contiene la columna con la clasificación de dominio
3
4 y_Perth_encoded = y_Perth.map({'Primario': 1, 'Secundario': 0})
```

9.2.1.3. Dividir en conjunto de entrenamiento, prueba y validación

Como se mencionó anteriormente, la práctica común es utilizar el set de entrenamiento (*train set*) para entrenar el modelo y obtener los hiperparámetros, luego acudir al set de validación (*validation set*) para ajustar estos hiperparámetros, escoger el mejor modelo y finalmente, utilizar el set de testeo (*test set*) para evaluar el desempeño del modelo en datos no antes vistos. Sin embargo, la metodología propuesta en este proyecto utiliza el conjunto de validación como un segundo set de testeo para comparar el rendimiento entre modelos. Más aún, la partición de los datos es tal que el 70% de los datos corresponden al conjunto de prueba, 20% al set de prueba y el 10% restante conforman el set de validación, lo anterior da como resultado la distribución de datos expuesta en la tabla 9.5.

Tabla 9.5: Cantidad de datos de cada set para cada variable.

Conjunto de datos	Train set	Test set	Validation set
Perth_Spectrum_Con_Contacto	4323	1235	618
Perth_1_Spectrum_Con_Contacto	1063	304	151
Perth_2_Spectrum_Con_Contacto	4323	1235	618
Santiago_Spectrum_Con_Contacto	3528	1008	504
Santiago_1_Spectrum_Con_Contacto	1915	547	274
Santiago_2_Spectrum_Con_Contacto	3528	1008	504

9.2.2. Bosque aleatorio

9.2.2.1. Búsqueda mejores hiperparámetros

Para la selección de los mejores hiperparámetros del clasificador Bosque aleatorio o *Random Forest* (RF) se realiza una primera búsqueda con la función *RandomizedSearchCV*. Seguido de una segunda búsqueda con la clase *GridSearchCV*, tomando valores cercanos al óptimo identificado en el primer paso. Los resultados se resumen en la tabla 9.6 para Perth_1 y Perth_2 respectivamente. Del mismo modo, los mejores hiperparámetros para el conjunto de datos analizados en Santiago, se encuentran en C.1.1.

Tabla 9.6: Mejores hiperparámetros encontrados en la primera (*RandomizedSearchCV*) y segunda búsqueda (*GridSearchCV*) para el clasificador *Random Forest*, con los datos de Perth_1_Spectrum_Con_Contacto y Perth_2_Spectrum_Con_Contacto. En color amarillo se destacan los valores distintos entre una búsqueda y otra.

RANDOM FOREST				
	Perth_1		Perth_2	
Mejores hiperparámetros	R.SearchCV	G.SearchCV	R.SearchCV	G.SearchCV
criterion	entropy	entropy	entropy	entropy
max_depth	None	None	None	15
max_features	0.5	0.5	0.5	0.45
min_samples_leaf	2	2	1	2
min_samples_split	5	5	10	10
n_estimators	100	110	200	190
class_weight	None	None	balanced	balanced
bootstrap	True	True	True	True

Con respecto a las métricas de evaluación, las que derivan de evaluar el modelo con los mejores hiperparámetros encontrados en cada búsqueda, se resumen en la tabla 9.7 y C.1.2 para Perth_1 y Perth_2 respectivamente. Se puede observar en la tabla 9.7 que las métricas son similares, lo que estaría directamente relacionado con la ligera diferencia en el número óptimo de estimadores propuesto por cada búsqueda (tabla 9.6), siendo 100 para la primera búsqueda y 110 para la segunda. Del mismo modo, y considerando que tanto las métricas como los hiperparámetros encontrados por cada búsqueda son prácticamente idénticos, la elección del mejor método de búsqueda es indiferente, no obstante, se escoge el mejor modelo como el que tiene los hiperparámetros encontrados por *RandomizedSearchCV*, ya que presenta una pequeña ligera mejora de 0.005 en exactitud. Con respecto a las métricas de Perth_2, expuestas en la tabla C.1.2, se observa una pequeña mejora en exactitud en el modelo resultante de la segunda búsqueda, incluso presenta un equilibrio ligeramente mejor en términos de precisión y recall, especialmente para la clase Primario. Sumado a lo anterior, la configuración de los hiperparámetros es más restrictiva con *GridSearchCV*, lo que podría ayudar a prevenir sobreajuste. Considerando lo expuesto, se escoge el modelo resultante de la búsqueda con *GridSearchCV* como el mejor modelo de *Random Forest* para Perth_2.

Por otro lado, en la tabla C.1.4 se visualizan las métricas de los modelos resultantes para Santiago_1, al emplear los mejores hiperparámetros encontrados en cada búsqueda. Considerando que se observa un rendimiento bastante similar, se prefiere el modelo resultante de la segunda búsqueda, ya que presenta una ligera mejora en exactitud, precisión, *recall* y *f1-score*. Bajo la misma idea, se evalúan las métricas de los modelos resultantes para Santiago_2, expuestas en la tabla C.1.4, donde tanto los valores de precisión, *recall* como *f1-score* son prácticamente idénticos en ambos modelos, por lo que cada modelo es igualmente bueno para identificar positivos verdaderos sin incluir demasiados falsos positivos. En cuanto a la matriz de confusión, son similares. Dado que ambos modelos tienen un rendimiento similar, no se tiene una preferencia de uno sobre otro, no obstante, se nota que el modelo obtenido con los hiperparámetros encontrados con *RandomizedSearchCV* presenta una exactitud ligeramente superior, por lo que se escoge como el mejor modelo para Santiago_2. Finalmente, los modelos seleccionados como mejores para cada conjunto de datos se resumen en la tabla 9.8.

Tabla 9.7: Métricas de evaluación del modelo *Random Forest*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para *Perth_1*.

RANDOM FOREST									
Métricas de evaluación Perth_1									
1° Búsqueda - RandomizedSearchCV					2° Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.8980			Exactitud en el conjunto de prueba		0.8947		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.87	0.96	0.91	171	Secundario	0.86	0.96	0.91	171
Primario	0.95	0.81	0.87	133	Primario	0.95	0.80	0.87	133
accuracy			0.90	304	accuracy			0.89	304
macro avg	0.91	0.89	0.89	304	macro avg	0.91	0.88	0.89	304
weighted avg	0.90	0.90	0.90	304	weighted avg	0.91	0.89	0.90	304
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	165	6			Positivo	165	6		
Negativo	26	107			Negativo	26	107		

Tabla 9.8: Hiperparámetros seleccionados del modelo *Random Forest* para clasificar el contacto primario-secundario, con los datos de Perth_1, Perth_2, Santiago_1 y Santiago_2, según corresponda.

RANDOM FOREST				
Hiperparámetros de los modelos seleccionados				
	Perth_1	Perth_2	Santiago_1	Santiago_2
criterion	entropy	entropy	entropy	entropy
max_depth	None	15	10	None
max_features	0.5	0.45	0.45	0.5
min_samples_leaf	2	2	1	1
min_samples_split	5	10	4	10
n_estimators	100	190	100	200
class_weight	None	balanced	None	balanced
bootstrap	True	True	True	True

9.2.3. Máquina de vectores de soporte

9.2.3.1. Búsqueda mejores hiperparámetros

Para la construcción de los modelos del clasificador máquina de vectores de soporte o *Support Vector Machine* (SVM), al igual que RF, se realiza la búsqueda doble de hiperparámetros con *RandomizedSearchCV* y *GridSearchCV* para los conjuntos de datos de la sección anterior, de los que se obtiene lo siguiente:

- Para **Perth_Spectrum_Con_Contacto** se tienen los mejores hiperparámetros expuestos en las primeras dos columnas de la tabla C.2.1, mientras que las métricas al evaluar los modelos resultantes se plasman en la tabla C.2.3. Al comparar las

métricas de los modelos, se puede observar que los hiperparámetros obtenidos con *RandomizedSearchCV*, que sólo se diferencian en el valor C con la búsqueda de *GridSearchCV*, entregan una ligera mejora en la exactitud en el conjunto de prueba. Por lo anterior, se prefieren los hiperparámetros de la primera búsqueda como los mejores para el set de datos.

- Para **Perth_1_Spectrum_Con_Contacto** se tienen los mejores hiperparámetros expuestos en la tercera y cuarta columna de la tabla C.2.1, mientras que las métricas resultantes al evaluar los modelos en el conjunto de prueba se plasman en la tabla C.2.4. Ambos modelos muestran métricas y rendimiento bastante similar, por lo que no hay una preferencia de uno sobre otro. No obstante, se prefiere el modelo resultante de la búsqueda con *GridSearchCV*, ya que realiza una búsqueda exhaustiva, por lo que puede ser ligeramente más confiable.
- Para **Perth_2_Spectrum_Con_Contacto** se tienen los mejores hiperparámetros expuestos en las dos últimas columnas de la tabla C.2.1, mientras que las métricas de evaluación se plasman en la tabla C.2.5. Ambos modelos tienen una exactitud muy similar en el conjunto de prueba, sin embargo, el modelo resultante de la búsqueda con *RandomizedSearchCV* muestra una ligera mejora en precisión y *recall* para ambas clases, por lo que se elige como mejor modelo para el conjunto.
- Para **Santiago_Spectrum_Con_Contacto** se tienen los mejores hiperparámetros expuestos en las primeras dos columnas de la tabla C.2.2, mientras que las métricas al evaluar los modelos resultantes se plasman en la tabla C.2.6. Dada la similitud de las métricas y los hiperparámetros encontrados, que son prácticamente idénticos, la búsqueda con *RandomizedSearchCV* es suficiente y, por lo tanto, se escoge como mejor opción.
- Para **Santiago_1_Spectrum_Con_Contacto** se tienen los mejores hiperparámetros expuestos en la tercera y cuarta columna de la tabla C.2.2, mientras que las métricas resultantes al evaluar los modelos en el conjunto de prueba se plasman en la tabla C.2.7. Ambos modelos tienen un rendimiento similar, sin embargo, el modelo con los hiperparámetros encontrados con *GridSearchCV* tiene una ligera mejora en la precisión con un valor de C mayor, se escogen como mejores hiperparámetros para el set de datos.
- Para **Santiago_2_Spectrum_Con_Contacto** se tienen los mejores hiperparámetros expuestos en las dos últimas columnas de la tabla C.2.2, mientras que las métricas de evaluación se plasman en la tabla C.2.8. Los modelos tienen exactamente los mismos hiperparámetros y métricas de evaluación, no hay preferencia de uno sobre otro, por lo tanto, la búsqueda con *RandomizedSearchCV* demostró ser suficiente y se escoge como la mejor opción.

Finalmente, considerando lo expuesto, los modelos seleccionados para cada conjunto de datos se resumen en la tabla 9.9.

Tabla 9.9: Hiperparámetros seleccionados del modelo *Support Vector Machine* para clasificar el contacto primario-secundario, con los datos de Perth_1, Perth_2, Santiago_1 y Santiago_2, según corresponda.

SUPPORT VECTOR MACHINE

Hiperparámetros de modelos seleccionados

	Perth	Santiago	Perth_1	Perth_2	Santiago_1	Santiago_2
kernel	linear	linear	linear	linear	linear	linear
gamma	10	10	-	100	-	-
degree	4	4	-	3	-	-
class_weight	None	None	None	balanced	None	balanced
C	0.10	0.10	0.15	0.10	0.15	1.00

9.2.4. Regresión Logística

9.2.4.1. Búsqueda mejores hiperparámetros

Del mismo modo que los modelos anteriores, se realiza una doble búsqueda de los mejores hiperparámetros para la construcción del modelo de Regresión Logística o *Logistic Regression* (LR), obteniendo lo siguiente:

- Para **Perth_Spectrum_Con_Contacto** se tienen los mejores hiperparámetros expuestos en las primeras dos columnas de la tabla C.3.1, mientras que las métricas al evaluar los modelos resultantes se plasman en la tabla C.3.3. Dada la similitud de las métricas, la búsqueda con *RandomizedSearchCV* es suficiente y, por lo tanto, se escoge como mejor opción.
- Para **Perth_1_Spectrum_Con_Contacto** se tienen los mejores hiperparámetros expuestos en la tercera y cuarta columna de la tabla C.3.1, mientras que las métricas resultantes al evaluar los modelos en el conjunto de prueba se plasman en la tabla

C.3.4. Ambos modelos tienen rendimiento muy similar en términos de precisión, *recall* y *f1-score*. Considerando la búsqueda exhaustiva realizada por *GridSearchCV* y el ligero aumento en la exactitud en el conjunto de prueba, se escogen los hiperparámetros de la segunda búsqueda como la mejor opción.

- Para **Perth_2_Spectrum_Con_Contacto** se tienen los mejores hiperparámetros expuestos en las dos últimas columnas de la tabla C.3.1, mientras que las métricas de evaluación se plasman en la tabla C.3.5. Ambos modelos tienen métricas similares, sin embargo, el modelo obtenido con los parámetros encontrados por *RandomizedSearchCV* tiene una precisión ligeramente mayor para la clase secundaria (Secundario) y una tasa de verdaderos positivos ligeramente más baja para la misma clase, por lo que se elige como mejor modelo para el conjunto.
- Para **Santiago_Spectrum_Con_Contacto** se tienen los mejores hiperparámetros expuestos en las primeras dos columnas de la tabla C.3.2, mientras que las métricas al evaluar los modelos resultantes se plasman en la tabla C.3.6. Ambos modelos tienen un rendimiento similar, sin embargo, el modelo con los hiperparámetros encontrados con *GridSearchCV* tiene una ligera mejora en la exactitud, se escogen como mejores hiperparámetros para el set de datos.
- Para **Santiago_1_Spectrum_Con_Contacto** se tienen los mejores hiperparámetros expuestos en la tercera y cuarta columna de la tabla C.3.2, mientras que las métricas resultantes al evaluar los modelos en el conjunto de prueba se plasman en la tabla C.3.7. Los modelos, a excepción de 'max_iter', tienen los mismos hiperparámetros y métricas de evaluación, no hay preferencia de uno sobre otro, por lo tanto, la búsqueda con *RandomizedSearchCV* demostró ser suficiente y se escoge como la mejor opción.
- Para **Santiago_2_Spectrum_Con_Contacto** se tienen los mejores hiperparámetros expuestos en las dos últimas columnas de la tabla C.3.2, mientras que las métricas de evaluación se plasman en la tabla C.3.8. Las métricas de los modelos son bastante similares, sin embargo, el primer modelo resultante de la búsqueda con *RandomizedSearchCV* tiene una precisión ligeramente mayor para la clase primaria (Primario) y una tasa de verdaderos positivos ligeramente más baja para la misma clase, así que se escoge como mejor modelo.

Finalmente, considerando lo expuesto, los modelos seleccionados para cada conjunto de datos se resumen en la tabla 9.10.

Tabla 9.10: Hiperparámetros seleccionados del modelo *Logistic Regression* para clasificar el contacto primario-secundario, con los datos de Perth_1, Perth_2, Santiago_1 y Santiago_2, según corresponda.

LOGISTIC REGRESSION						
Hiperparámetros de modelos seleccionados						
	Perth	Santiago	Perth_1	Perth_2	Santiago_1	Santiago_2
penalty	l1	l1	l2	l2	elasticnet	l1
max_iter	3000	2500	1500	2000	1500	3000
learning_rate	adaptive	adaptive	adaptive	adaptive	adaptive	adaptive
eta0	0.001	0.0015	0.0005	0.001	0.1	0.001
class_weight	None	None	None	None	balanced	None
alpha	0.001	0.0015	1e-05	0.0001	0.01	0.001

9.2.5. Redes neuronales

9.2.5.1. Búsqueda mejores hiperparámetros y selección

En contraste con los modelos anteriores, los hiperparámetros en redes neuronales se establecen antes de comenzar el entrenamiento, de modo que, para la construcción del modelo, se crea un sintonizador *Hyperband* que busca los mejores hiperparámetros, mediante el método *search*, tal que maximizan la precisión en el conjunto de validación. Una vez encontrados estos mejores hiperparámetros, se construye y entrena un nuevo modelo. Para cada conjunto de datos se buscan dos hiperparámetros, el número de neuronas en la capa oculta y la tasa de aprendizaje, obteniendo los siguientes resultados:

- Para **Perth_Spectrum_Con_Contacto** se propone un modelo con 128 neuronas en la capa oculta y una tasa de aprendizaje de 0.00028, que al evaluar en el conjunto de prueba alcanza 0.922 de exactitud y métricas promedio de *precision*, *recall* y *f1-score* igual a 0.92. El detalle de sus métricas se puede revisar en la primera columna de la tabla C.4.1.

- Para **Perth_1_Spectrum_Con_Contacto** se tienen 160 neuronas en la capa oculta y una tasa de aprendizaje de 0.00022 como mejores hiperparámetros, alcanzando una exactitud en el conjunto de prueba de 0.901, con *precision*, *recall* y *f1-score* promedio igual a 0.90, tal como se muestra en la segunda columna de la tabla C.4.1.
- Para **Perth_2_Spectrum_Con_Contacto** se encuentran como mejores hiperparámetros 128 neuronas en la capa oculta y una tasa de aprendizaje de 0.00156, al crear y evaluar el modelo con estos hiperparámetros en el conjunto de prueba, se obtienen las métricas de la última columna de la tabla C.4.1. Con respecto a las métricas, se alcanza una exactitud de 0.93, además de valores promedio de *precision*, *recall* y *f1-score* sobre 0.93.
- Para **Santiago_Spectrum_Con_Contacto** se encuentra como óptimo un modelo con 96 neuronas en la capa oculta y una tasa de aprendizaje de 0.0070, que al evaluar en el conjunto de prueba, alcanza una exactitud de 0.933 y valores promedio de *precision*, *recall* y *f1-score* iguales a 0.93, tal como se muestra en la primera columna de la tabla C.4.2.
- Para **Santiago_1_Spectrum_Con_Contacto** se propone un modelo con 320 neuronas en la capa oculta y una tasa de aprendizaje de 0.0010. Este modelo, al ser evaluado en el conjunto de prueba, alcanza una exactitud de 0.943 y valores promedio de *precision*, *recall* y *f1-score* iguales a 0.94, tal como se muestra en la segunda columna de la tabla C.4.2.
- Para **Santiago_2_Spectrum_Con_Contacto** se tiene 288 neuronas en la capa oculta y una tasa de aprendizaje de 0.00518 como mejores hiperparámetros para el modelo, alcanzando una exactitud en el conjunto de prueba de 0.943 y valores promedio de *precision*, *recall* y *f1-score* iguales a 0.94, tal como se muestra en la última columna de la tabla C.4.2.

9.3. Modelo no supervisado

9.3.1. K-means

A pesar de ser un clasificador que no necesita datos etiquetados, se utilizan los mismos conjuntos de los modelos anteriores, es decir, espectros que cuentan con contacto primario-secundario, ya que así se podrá evaluar el rendimiento del modelo para cada caso. Anterior a la construcción del modelo, se determina el coeficiente de silueta, homogeneidad y completitud para cada conjunto de datos, según el número de clústeres, obteniendo lo siguiente:

- Para **Perth_Spectrum_Con_Contacto** se obtienen mejores métricas para 2 clústeres, con un coeficiente promedio de silueta de 0.38, homogeneidad de 0.01 y completitud de 0.01. Recordar que se busca que el coeficiente de silueta sea lo más

cercano a 1, así, al seleccionar 2 clústeres para el modelo *K-means* se obtienen las métricas de la primera columna de la tabla C.5.1. De la tabla anterior, se tiene que los resultados de homogeneidad, completitud y *V-measure* son bastante bajos, por lo que los *clusters* formados con el algoritmo no se alinean con las categorías reales (Primario/Secundario). En cuanto a la matriz de confusión, se tienen 1482 datos clasificados correctamente para Secundario y 1155 para Primario. Mientras que 1126 y 2413 fueron incorrectamente clasificadas para Secundario y Primario, respectivamente. En general, más de la mitad de los datos fueron clasificados de manera incorrecta. Por lo tanto, el modelo de *Kmeans* no es capaz de capturar la complejidad de los datos y recae en clasificaciones erróneas.

- Para **Perth_1_Spectrum_Con_Contacto** se obtienen mejores métricas para 2 clústeres, con un coeficiente promedio de silueta de 0.38, homogeneidad de 0.01 y completitud de 0.01. Así, al seleccionar 2 clústeres para el modelo *K-means* se obtienen las métricas de la segunda columna de la tabla C.5.1. De la tabla anterior, se tiene que los resultados de homogeneidad, completitud y *V-measure* son bastante bajos, por lo que los *clusters* formados con el algoritmo no se alinean con las categorías reales (Primario/Secundario). En cuanto a la matriz de confusión, se tienen 559 datos clasificados correctamente para Secundario y 298 para Primario. Mientras que 373 y 288 fueron incorrectamente clasificadas para Secundario y Primario, respectivamente. En general, más de la mitad de los datos fueron clasificados de manera incorrecta. Por lo tanto, el modelo de *Kmeans* no es capaz de capturar la complejidad de los datos y recae en clasificaciones erróneas.
- Para **Perth_2_Spectrum_Con_Contacto** se obtienen mejores métricas para 2 clústeres, con un coeficiente promedio de silueta de 0.39, homogeneidad de 0.01 y completitud de 0.01. Al seleccionar 2 clústeres para el modelo, se obtienen las métricas de la tercera columna en la tabla C.5.1, donde, al igual que los otros conjuntos de Perth, se obtienen valores muy bajos en las métricas de homogeneidad, completitud y *v-measure*. También, de la matriz de confusión se observa que un tercio de las muestras de Secundario y la mitad de las muestras de Primario fueron clasificadas erróneamente. Nuevamente, se confirma que *K-means* no es capaz de capturar la complejidad de los datos.
- Para **Santiago_Spectrum_Con_Contacto** se obtienen mejores métricas para 2 clústeres, con un coeficiente promedio de silueta de 0.36, homogeneidad y completitud cercanas a 0. Al igual que con los datos de Perth, se obtiene un mayor valor para el coeficiente de silueta con 2 *clusters*. Sin embargo, las métricas de homogeneidad y completitud son extremadamente bajas para todos los casos (iguales o cercanas a 0). Por lo anterior, se puede estimar que los datos no presentan una separación clara en grupos discretos. Esto podría deberse a que las muestras tienen características similares, en este caso, las muestras presentan intensidades similares, por lo que el algoritmo de clustering tiene dificultades para distinguir entre clases. De todas

formas, se evalúa el modelo con 2 clústeres, obteniendo las métricas de la primera columna de la tabla C.5.2, donde se puede observar que las métricas son cercanas a cero e inclusive se tienen 1093 y 1390 datos incorrectamente clasificados para Secundario y Primario, respectivamente.

- Para **Santiago_1_Spectrum_Con_Contacto** se obtienen mejores métricas para 2 clústeres, con un coeficiente promedio de silueta de 0.36, homogeneidad y completitud cercanas a 0. Al igual que con los datos de Perth, se obtiene un mayor valor para el coeficiente de silueta con 2 *clusters*. Así, al evaluar el modelo con 2 clústeres, se obtienen las métricas de la segunda columna de la tabla C.5.2, no obstante, los valores de homogeneidad, completitud y *v-measure* son bastante bajos, lo que indica que hay poca o ninguna correspondencia entre los *clusters* y las clasificaciones reales. Con respecto a la matriz de confusión, se obtienen 425 muestras de Secundario y 722 muestras de Primario en el cluster 0, mientras que en el cluster 1 se tienen 546 muestras de Secundario y 1043 de Primario. Nuevamente, se tiene solapamiento de las clases, por lo que el algoritmo de Kmeans no está capturando bien la estructura de los datos.
- Para **Santiago_2_Spectrum_Con_Contacto** se obtienen mejores métricas para 2 clústeres, con un coeficiente promedio de silueta de 0.38, homogeneidad y completitud cercanas a 0. Al seleccionar 2 clústeres para el modelo, se obtienen las métricas de la tercera columna en la tabla C.5.2, donde, al igual que los otros conjuntos de Santiago, se obtienen valores muy bajos en las métricas de homogeneidad, completitud y *v-measure*. Nuevamente, se confirma que *K-means* no es capaz de capturar la complejidad de los datos.

9.4. Comparación de modelos

Al comparar los modelos para cada conjunto, se obtiene lo siguiente:

- Para **Perth_1_Spectrum_Con_Contacto** se comparan 4 modelos (RF, SVM, LR y NN), dejando fuera el modelo de *clustering* con *K-means*, ya que las métricas de este último muestran que llega a clasificar más de la mitad de los datos erróneamente, demostrando que no es capaz de capturar la complejidad de los datos. Por lo tanto, se evalúan tres modelos (RF, SVM, LR), para luego comparar las métricas del mejor modelo con las métricas de la red neuronal, ambos evaluados en el conjunto de prueba. Los resultados de la primera comparación se encuentran en la tabla C.6.1, donde se puede observar notablemente mejores métricas (precisión, *recall* y *f1-score*) para el modelo LR. Luego, al comparar con el modelo de redes neuronales (tabla C.6.2), sigue presentando mejores métricas el modelo de LR. Además, tiene un mejor equilibrio en la detección de ambas clases, lo que se puede ver en su matriz de confusión, con 7 falsos positivos y 14 falsos negativos en comparación de los 10

falsos positivos y 20 falsos negativos de la red neuronal. Más aún, presenta valores más altos de precisión, *recall* y *f1-score*.

- Para **Perth_2_Spectrum_Con_Contacto**, al igual que el conjunto anterior, se comparan 4 modelos (RF, SVM, LR y NN), dejando fuera el modelo de *clustering* con *K-means* por su bajo rendimiento. Los resultados de la primera comparación se encuentran en la tabla C.6.3, donde se puede observar notablemente mejores métricas (exactitud, precisión, *recall* y *f1-score*) para el modelo SVM, si bien presenta resultados similares con LR, SVM muestra menores falsos negativos. Con respecto a la segunda comparación (tabla C.6.4), el modelo SVM presenta mayor exactitud (94.17%), una mejora en las métricas (precisión, *recall* y *f1-score*) y un menor número de falsos negativos.
- Para **Santiago_1_Spectrum_Con_Contacto** se comparan 4 modelos (RF, SVM, LR y NN). Los resultados de la primera comparación se encuentran en la tabla C.6.5, se pueden observar mejores métricas (exactitud, *recall* y *f1-score*) para el modelo LR, ligeramente mejor que SVM. Luego, al comparar con el modelo de redes neuronales (tabla C.6.6), si bien el modelo de red neuronal sigue siendo valioso por su capacidad de capturar relaciones complejas, el modelo de LR presenta una ligera mejora en las métricas.
- Para **Santiago_2_Spectrum_Con_Contacto**, al igual que el conjunto anterior, se comparan 4 modelos (RF, SVM, LR y NN), dejando fuera el modelo de *clustering* con *K-means* por su bajo rendimiento. Los resultados de la primera comparación se encuentran en la tabla C.6.7, donde se puede observar notablemente mejores métricas (exactitud, precisión, *recall* y *f1-score*) para el modelo LR. Con respecto a la segunda comparación (tabla C.6.8), el modelo LR presenta mayor exactitud (94.84%) y una mejora en las métricas (precisión, *recall* y *f1-score*).
- Para **Perth_Spectrum_Con_Contacto** se comparan 3 modelos (SVM, LR y NN), se deja fuera el modelo RF dado su bajo rendimiento en comparación con SVM y LR evidenciado en el análisis de los conjuntos anteriores, del mismo modo, se deja fuera el modelo de *clustering* con *K-means* por su bajo rendimiento. Los resultados de la primera comparación se encuentran en la tabla C.6.9, donde se puede observar notablemente mejores métricas (exactitud, precisión, *recall* y *f1-score*) para el modelo SVM. Al comparar con el modelo NN (tabla C.6.10), SVM sigue presentando un rendimiento superior.
- Para **Santiago_Spectrum_Con_Contacto**, al igual que el conjunto anterior, se comparan 3 modelos (SVM, LR y NN), dejando fuera los modelos RF y *K-means*. Los resultados de la primera comparación se encuentran en la tabla C.6.11, se puede observar una ligera mejora en las métricas (exactitud, precisión, *recall* y *f1-score*) para el modelo SVM. Cabe mencionar que se utilizan tres cifras significativas para

mostrar la cercanía de las métricas. En cuanto a la segunda comparación, se puede observar una notable mejora en las métricas para el modelo SVM (tabla C.6.12).

Finalmente, considerando lo expuesto, los modelos seleccionados para cada conjunto de datos se resumen en la tabla 9.11. De la tabla anterior, se puede notar que el pre-procesamiento de los datos afecta directamente el rendimiento de los modelos, donde los datos sin eliminar valores negativos ni sobre 1, muestran SVM como el mejor modelo para datos analizados en Perth (Perth_Spectrum_Con_Contacto) y Santiago (Santiago_Spectrum_Con_Contacto). En contraste con los datos generados al eliminar filas negativas (Perth_1_Spectrum_Con_Contacto y Santiago_1_Spectrum_Con_Contacto), que muestran el modelo LR como el mejor en rendimiento. Mientras que los datos generados al eliminar columnas negativas, muestran diferentes modelos según donde fueron analizados los datos, teniendo SVM como mejor modelo para los datos analizados en Perth y LR para datos analizados en Santiago.

Tabla 9.11: Modelos seleccionados para cada conjunto de datos, luego de analizar sus métricas de rendimiento.

SELECCIÓN DE MEJORES MODELOS						
	Perth	Santiago	Perth_1	Santiago_1	Perth_2	Santiago_2
Modelo seleccionado	SVM	SVM	LR	LR	SVM	LR

9.5. Implementación

Para visualizar los datos en 3D, se cargan al software *Leapfrog Geo*, seleccionando dos modelos para los datos analizados en Perth y dos para Santiago. Los primeros dos modelos que se consideran son los seleccionados para el escenario 3 (datos sin eliminar negativos ni valores de reflectancia sobre 1), ambos corresponden a modelos SVM. Los últimos dos modelos, corresponden al modelo LR y SVM para el escenario 2 (datos sin columnas negativas ni valores de reflectancia sobre 1), ya que estos presentan una ligera mejora en las métricas de rendimiento en comparación con los modelos del escenario 1. Teniendo esto, se clasifican todos los espectros que coinciden en *SampleID* con los sondajes de *Merged_table*. Luego, se cargan a *Leapfrog Geo*, obteniendo la visualización 3D de los sondajes de las figuras 9.13 y 9.14 para el escenario 2 y 3, respectivamente.

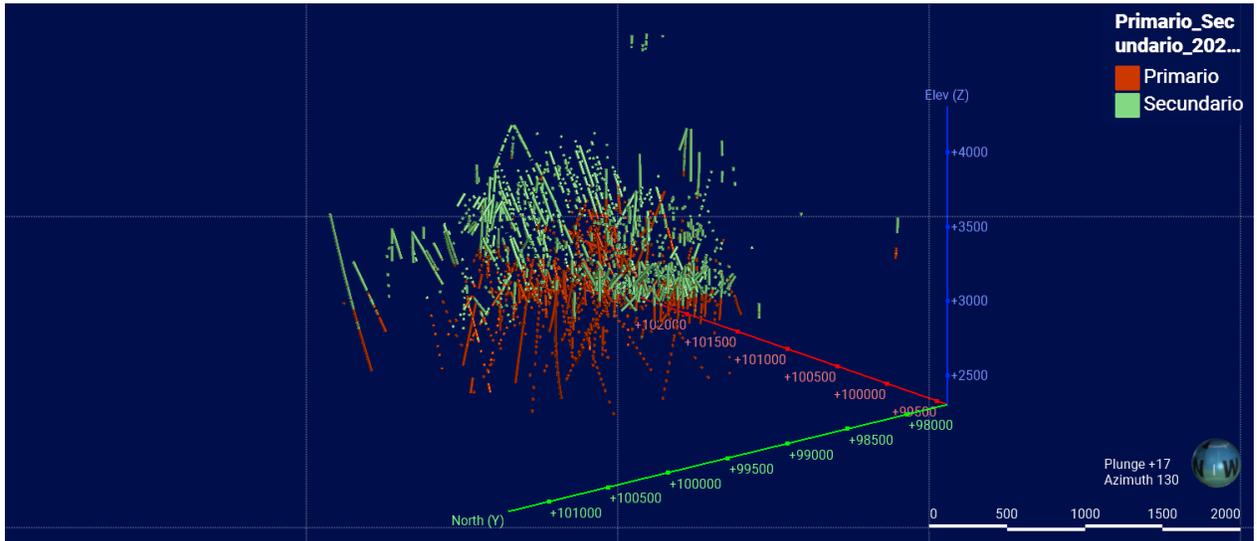


Figura 9.13: Visualización 3D de sondajes en *Leapfrog Geo* para el escenario 2, coloreados según contacto estimado por los modelos SVM para datos analizados en laboratorios de Perth y LR para datos analizados en laboratorios de Santiago.

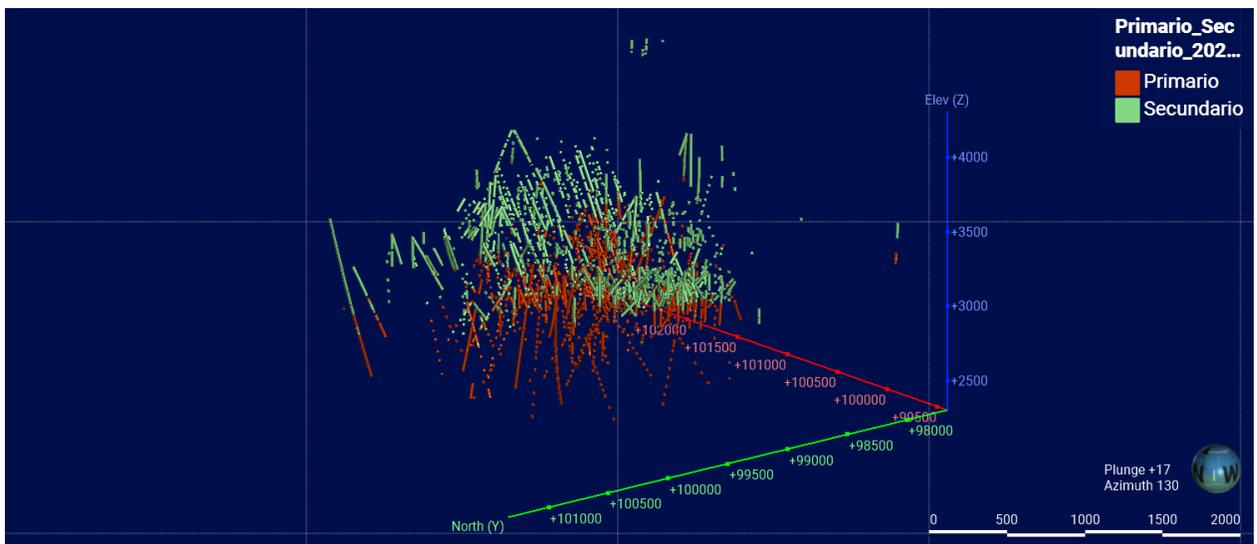


Figura 9.14: Visualización 3D de sondajes en *Leapfrog Geo* para el escenario 3, coloreados según contacto estimado por los modelos SVM para datos analizados en Santiago y Perth.

9.5.1. Empaquetamiento de la solución

Para simplificar el flujo de trabajo de los pasos del proyecto, se empaqueta la solución en cuatro rutinas que utilizan la interfaz *tkinter* de Python, sin embargo, la búsqueda doble de hiperparámetros para los modelos supervisados no se incluye en las rutinas, ya que es un paso exigente en tiempo computacional y supervisión, por lo que se estima adecuado realizarlo de manera manual. Bajo la misma idea, considerando la complejidad de los

modelos y el tiempo computacional, no se añade el modelo RF ni NN en la interfaz de entrenamiento. A continuación, se muestra el resultado de cada rutina:

- **Primer paso** (Figura 9.15): Filtrar los archivos que contengan espectros ('Archivo a filtrar') según las coincidencias en la columna *SampleID* del archivo de sondajes ('Archivo guía'). El usuario ingresa el archivo con espectros, el archivo guía y la carpeta de salida. Luego, si lo estima conveniente, modifica los nombres de los archivos que serán generados. Seguido de esto, debe ingresar el nombre de las columnas que tienen los nombres de las muestras, si no, se toman los valores por defecto. Finalmente, decide si ajustar los nombres de la columna de muestras a la nomenclatura.

Filtrar espectros según SampleID

Archivo a filtrar: Seleccionar archivo

Archivo guía: Seleccionar archivo

Carpeta de salida: Seleccionar carpeta

Nombre del archivo con coincidencias (no incluir comillas ni '.csv'):

Nombre del archivo sin coincidencias (no incluir comillas ni '.csv'):

Nombre de la columna de muestras en archivo a filtrar (no incluir comillas):

Nombre de la columna de muestras en archivo guía (no incluir comillas):

¿Ajustar nomenclatura de muestras en archivo a filtrar? Formato 'CODIGO_5 digitos'

¿Ajustar nomenclatura de muestras en archivo guía? Ejemplo 'MET_04319'

Ejecutar

Figura 9.15: Interfaz para filtrar espectros según *SampleID*.

- **Segundo Paso** (Figura 9.16): Filtrar los espectros ('Arhivo a filtrar') según la columna del contacto primario secundario desde el archivo de sondajes ('Archivo guía'). El usuario ingresa el archivo de espectros, el archivo guía y selecciona la carpeta de salida. Luego, modifica los nombres de los archivos que serán generados. Posterior a esto, debe modificar la columna con los nombres de las muestras y la del contacto. Finalmente, decide si ajustar los nombres de la columna de muestras a la nomenclatura.

Filtrar Contacto Primario Secundario

Archivo a filtrar:

Archivo guía:

Carpeta de salida:

Nombre del archivo con contacto (no incluir comillas ni '.csv'):

Nombre del archivo sin contacto (no incluir comillas ni '.csv'):

Nombre de la columna de muestras en archivo a filtrar (no incluir comillas):

Nombre de la columna del contacto en archivo guía (no incluir comillas):

¿Ajustar nomenclatura de muestras en archivo a filtrar? Formato 'CODIGO_5 digitos'

¿Ajustar nomenclatura de muestras en archivo guía? Ejemplo 'MET_04319'

Figura 9.16: Interfaz para filtrar contacto primario-secundario.

- **Tercer Paso** (Figura 9.17): Entrenamiento de los modelos SVM y LR. El usuario ingresa el archivo con espectros que cuenten con etiqueta del contacto primario secundario y selecciona la carpeta de salida. Luego, debe ingresar el nombre para guardar el escalador, la lista de características y los modelos SVM y LR. También, se le da la opción de modificar los hiperparámetros, por defecto están los mejores hiperparámetros para el escenario 3. Finalmente, el usuario selecciona si quiere entrenar un modelo u otro o ambos.

Entrenamiento de Modelos SVM y Logistic Regression

Archivo de Espectros con Contacto Primario Secundario:

Carpeta de Salida:

Nombre para guardar escalador :

Nombre para guardar archivo de características :

Nombre para guardar modelo SVM :

Valor de C para SVM: [0.0, inf)

Kernel para SVM: {linear, poly, rbf, sigmoid, precomputed}

Valor de gamma para SVM: {scale, auto} o [0.0, inf)

Valor de degree para SVM: número entero

Class weight para SVM: dict, {balanced, None}

Nombre para guardar modelo Logistic Regression :

Valor de alpha para LR: [0.0, inf)

Valor de penalty para LR: {l2, l1, elasticnet, None}

Valor de max_iter para LR: número entero

Valor de learning_rate para LR: {constant, optimal, adaptive}

Valor de eta0 para LR: [0.0, inf)

Valor de class_weight para LR: dict, {balanced, None}

Entrenar SVM Entrenar Logistic Regression

Figura 9.17: Entrenamiento de modelos SVM y LR.

- **Cuarto Paso** (Figura 9.18): Clasificación del contacto en espectros. El usuario ingresa el archivo con los espectros a clasificar, selecciona la carpeta de salida, así también como el archivo con las características del modelo. Opcionalmente, puede seleccionar el archivo con el escalador utilizado para normalizar los datos en el entrenamiento del modelo, y puedo cargar un archivo donde añadir la columna del contacto clasificado con el modelo, por rutina, se añade como columna final del archivo seleccionado. Luego, ingresa el nombre de la columna de muestras, seguido de ingresar el nombre de la nueva columna del contacto a generar así también como el archivo para guardar el archivo de espectros clasificados. Opcionalmente, puede ingresar el nombre del archivo al que se le añade la nueva columna del contacto. Finalmente, selecciona el modelo a utilizar para la clasificación.

Clasificación del contacto en espectros

Archivo de Espectros a clasificar:

Carpeta de Salida:

Archivo de modelo para clasificación:

Archivo con características del modelo:

Archivo con escalador utilizado en el modelo (Opcional):

Archivo donde añadir la columna del contacto clasificada:

Nombre de la columna con nombre de las muestras en archivo de espectros:

Nombre de la columna con nombre de las muestras en archivo de sondajes:

Nombrar la nueva columna del contacto a generar:

Nombre para guardar archivo de espectros clasificados:

Nombre para guardar archivo de sondajes con nueva columna del contacto:

Tipo de modelo para clasificación:

Modelo Support Vector Machine Modelo Logistic Regression

Modelo Support Random Forest Modelo Neural Network

Figura 9.18: Clasificación del contacto a partir de espectros.

Estas rutinas están particularmente diseñadas para los archivos del proyecto, por lo que ante cualquier modificación de las columnas, debe modificarse la interfaz para asegurar el correcto funcionamiento de estas.

CAPÍTULO 10

DISCUSIÓN

Como primer punto a tocar en la discusión, se tiene la distribución de litologías presentes en cada dominio, obtenida posterior a clasificar las muestras con los modelos. Si los modelos logran capturar de buena manera la complejidad de los datos, no debiesen variar de manera significativa con respecto a los datos originales. Para el escenario 2, que utiliza un modelo SVM para clasificar los datos analizados en laboratorios de Perth y LR para las muestras analizadas en Santiago, se tienen las distribuciones de la figura D.0.1 para el dominio primario y de la figura D.0.2 para el dominio secundario. Al comparar con la distribución de los datos originales (figura 9.2 para primario y 9.3 para secundario), se observa que 1) se mantiene la predominancia de QM en ambos dominios; 2) mantiene la alta frecuencia de BXPR, BXA y BXBM en el dominio primario sobre el secundario; 3) mantiene la baja presencia de BXA en el dominio secundario; 4) se captura la baja frecuencia de las muestras de RIO y AND en el dominio primario, siendo más frecuentes en el secundario; y 5) concordancia en la baja presencia de muestras de BXI clasificadas como primario. Con lo anterior, se puede concluir que los modelos generados para el escenario 2 mantienen la distribución de las litologías clasificadas en el dominio primario y secundario. Con respecto al escenario 3, que utiliza modelos SVM para muestras analizadas tanto en laboratorios de Perth como Santiago, se tienen las siguientes distribuciones de la figura D.0.3 para el dominio primario y la figura D.0.4 para el dominio secundario. Del mismo modo que el escenario 2, se compara con los datos originales, obteniendo que mantiene las relaciones expuestas (1,2,3,4,y 5), por lo que se puede concluir que logra capturar la complejidad de los datos y mantener la distribución litológica de cada dominio.

Ahora, si bien se desconoce una descripción litológica de cada dominio, a partir de la distribución de los datos originales se puede estimar que:

- La litología QM no es diagnóstica de los dominios, ya que esta altamente presente en ambos.
- BXA puede ser una litología indicativa de encontrarse en el dominio primario, dada su alta presencia, ya que si recordamos el capítulo 7, se tiene que la anhidrita se asocia a la mineralización hipógena más profunda en zonas de alteración potásica

y, por lo tanto, estaría ausente en zonas superficiales. Lo anterior, se apoya dada la alta presencia de BXBM en el dominio primario, brecha que se caracteriza por presentarse siempre en ambiente potásico. En cuanto a BXBM, puede ser indicativa de encontrarse en el dominio primario, ya que su presencia en el dominio secundario es notablemente baja.

- La litología BXPR puede asociarse a la presencia del dominio primario, dada su alta frecuencia en este en comparación con el dominio secundario. Inclusive, la baja presencia de BXI en el dominio primario, puede reforzar la hipótesis de que esta brecha correspondería a la continuidad en profundidad de BXI.

Por otro lado, con respecto a la metodología propuesta de preprocesamiento de los datos de espectros ASD-FTIR, si bien no cuenta con un respaldo teórico, los modelos obtenidos a partir de los escenarios generados presentaron métricas sobre el 80 % en la mayoría de los casos. En particular, para el escenario 1, en el cual se eliminan 65 % de las muestras analizadas en Perth y 47 % de las muestras analizadas en Santiago, a pesar de tener menos datos para la construcción de los modelos, los datos analizados en Perth obtuvieron como exactitud más baja un valor de 89 % para el modelo RF, mientras que en el modelo NN alcanzó un 90.1 %, seguido de un 93 % de exactitud con el modelo LR y finalmente el modelo SVM con una exactitud del 95.3 %. Por lo tanto, la metodología entrega resultados robustos al eliminar los valores negativos y de reflectancia sobre 1, que se presumen corresponderían a datos anómalos e incluso ruido producto de perturbaciones durante el proceso de medición. Finalmente, tanto para Perth como Santiago, los modelos de *K-means* muestran un bajo rendimiento.

En cuanto al escenario 2, en este se comprime el rango electromagnético estudiado al eliminar un 2 % y 3.4 % de las columnas con valores negativos de muestras analizadas en Perth y Santiago, respectivamente. Por lo que cuenta con la misma cantidad de muestras del escenario 3, en el cual no se eliminan valores negativos ni reflectancias sobre 1, pero abarca hasta la longitud 22,159.76 nm, no hasta 44,701.59 nm como en el escenario 3. Con respecto a las métricas de sus modelos, se obtienen los valores más bajos para Perth en el modelo RF con un 91 % de exactitud, mientras que para Santiago la exactitud más baja se obtiene en el modelo RF con un valor de 93 %. En particular, los datos de Perth alcanzan una exactitud del 93.7 % para el modelo NN, 94.1 % para LR y 94.9 % para SVM. Del mismo modo, los datos de Santiago alcanzan una exactitud del 94.3 % para el modelo NN, 94.8 % para LR y 95.4 % para SVM. No obstante, al evaluar el modelo SVM y LR en muestras no antes vistas, se obtiene una exactitud de 93.6 % y 93.8 %, respectivamente, por lo que se selecciona el modelo LR como el mejor para Santiago en el escenario 2.

Por último, se tiene el escenario 3, para este conjunto no se construyen modelos RF, ya que analizando los escenarios anteriores, no presentaban mejores métricas que LR y SVM. Por lo tanto, para los datos de Perth se alcanzan valores de exactitud en el conjunto de prueba de 93.2 % para NN, 92.6 % para LR y 93.6 % para SVM. Al evaluar los datos

de Santiago, se obtienen valores de exactitud en el conjunto de prueba de 93.3% para NN, 94.7% para SVM y 95% para LR. No obstante, al evaluar en datos no antes vistos, el modelo SVM supera las métricas de LR, por lo que se selecciona como el mejor modelo para Santiago en el escenario 3. Más aún, al comparar los modelos de los tres escenarios, se evidencia que el escenario 2 presenta mejores métricas que el escenario 3, que a su vez presenta mejores métricas que el escenario 1.

Dado que la metodología de preprocesamiento de los datos difiere del estándar, se recomienda utilizar los datos del escenario 3, es decir, no eliminar los valores que se presumen como anómalos, ya que las métricas de los modelos asociados a este conjunto de datos no presentan una diferencia significativa con los otros escenarios, en particular con el escenario 2. Así también, se recomienda aplicar métodos estadísticos como las pruebas de *t-Student* y *Fisher* para analizar el impacto en las estadísticas al modificar los datos originales. Cabe mencionar que la metodología de preprocesamiento se propone para superar la limitación de no contar con información detallada sobre el proceso de medición en laboratorio de los espectros, en el caso que si se cuenta con mayor detalle, se recomienda preferir metodologías estándar de procesamiento de señales, como la expuesta en 8.2.2.

Ahora bien, con respecto a la búsqueda doble de hiperparámetros para los modelos supervisados (RF, SVM y LR), en la mayoría de los casos, la primera búsqueda con *RandomizedSearchCV* demostró ser suficiente, por lo tanto, con miras a optimizar el proceso, se recomienda realizar sólo una búsqueda simple con *RandomizedSearchCV* para búsquedas futuras.

Por otro lado, se evidencia que el algoritmo *K-means* presenta bajo rendimiento en todos los escenarios, por lo que no es capaz de capturar la complejidad de los datos. Se propone que, dado que *K-means* asume que los clústeres son isotrópicos y equidistantes respecto a su centroide, puede no ser el mejor algoritmo para este tipo de datos. Se recomienda para trabajos futuros probar otros algoritmos de clustering como DBSCAN, HDBSCAN o Modelos de mezcla gaussiana.

Con respecto al empaquetamiento de la solución, la interfaz generada facilita el flujo de trabajo para entrenar y clasificar muestras que cuenten con un archivo de espectros y coordenadas geográficas asociadas, ya que el archivo generado no sólo entrega la clasificación del dominio de la muestra, si no que la clasificación puede añadirse como columna a archivos de sondeos y cargar a *Leapfrog Geo*. No obstante, la interfaz fue diseñada para recibir archivos con el mismo formato que los archivos de esta memoria, por lo tanto, ante cualquier cambio en los archivos, debe modificarse la rutina de la interfaz para asegurar su correcto funcionamiento.

CAPÍTULO 11

CONCLUSIÓN

Los modelos supervisados fueron lo bastante robustos para clasificar las muestras, tanto las analizadas en laboratorios de Perth como las de Santiago, presentando métricas sobre el 80 % en la mayoría de los casos e incluso alcanzando valores de exactitud del 95 %. Más aún, a pesar de implementar una metodología de preprocesamiento no estándar, los modelos fueron capaces de caracterizar el contacto entre la roca primaria y secundaria del yacimiento Los Bronces. Por lo tanto, se recomienda utilizar modelos supervisados como *Logistic Regression*, *Support Vector Machine* y *Neural Network* para el entrenamiento y clasificación de nuevos conjuntos de datos, ya que estos modelos demostraron mejor rendimiento. Del mismo modo, se descarta utilizar modelos no supervisados como *K-means*, dado que demostró no ser capaz de capturar la complejidad de los datos. Finalmente, en cuanto al modelo *Random Forest*, sumado a que sus métricas son ligeramente más bajas que los demás modelos supervisados, exige una alta capacidad computacional. Considerando la inmediatez de los datos en minería, se recomienda no utilizar modelos costosos computacionalmente, a menos que sus métricas sean notablemente superiores.

Por otro lado, analizando la distribución de las litologías, se propone considerar la alta o baja ocurrencia de ciertas litologías en cada dominio para la descripción de estos. Más aún, se puede recurrir a las descripciones de las litologías y su ambiente de ocurrencia para caracterizar de manera preliminar los dominios según litología. De lo anterior, se rescata que la alta presencia de BXA, BXBM o BXPR puede indicar que la muestra se encuentra en el dominio primario.

CAPÍTULO 12

BIBLIOGRAFÍA

Abarzúa, A. (2014). Zonación de los tipos de alteración hidrotermal en el sector Los Bronces, como parte del sistema porfídico Cu-Mo Río Blanco-Los Bronces, Región Metropolitana, Chile. Memoria para optar al título de Geólogo, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago.

Aguirre, L. (1960). Geología de Los Andes de Chile Central, Provincia de Aconcagua. Instituto de Investigaciones Geológicas. Vol. 9. Pp. 70. Santiago, Chile.

Agus, Abigail June L. (2011) Mapping white mica in milled porphyry copper pebbles using hyperspectral imagery : an exploratory study.

AngloAmerican. (2012). Informe a comunidades Los Bronces - SEAT III. Santiago, Chile, pp. 10-11. <https://chile.angloamerican.com/~media/Files/A/Anglo-American-Group-v5/Chile/sustentabilidad/comunidades/seat-los-bronces-2012.pdf>

AngloAmerican. (2021) Superintendencia Geología Mina - General Geology. [Presentación]. Santiago, Chile.//

Anglo American (2023). Geología de Los Bronces. Informe Interno.

Anglo American Chile. (2 de febrero de 2023a). *Informe de producción correspondiente al cuarto trimestre finalizado el 31 de diciembre de 2022*. <https://chile.angloamerican.com/medios/press-releases/pr-2023/2023-02-02.aspx>

Anglo American Chile. (8 de febrero de 2024). *Informe de producción del cuarto trimestre finalizado el 31 de diciembre de 2023*. <https://chile.angloamerican.com/medios/press-releases/pr-2024/2024-02-08.aspx>

Ávila, M. (2019). Desarrollo de guía de mejores prácticas en estimación de recursos mineros para yacimientos tipo pórfido cuprífero en Chile. [Memoria para optar al título de Ingeniero Civil de Minas, Universidad de Chile]. Disponible en <https://repositorio.uchile.cl/handle/2250/174371>

Amazon Web Services, Inc. (AWS). (n.d.). *¿Qué es una red neuronal? - Explicación de las redes neuronales artificiales*. <https://aws.amazon.com/es/what-is/neural-network/#:~:text=Una%20red%20neuronal%20es%20un,lo%20hace%20el%20cerebro%20humano.>

Azhari, F., Sennersten, C., Lindley, C. y Sellers, E. (2023). Deep learning implemen-

tations in mining applications: a compact critical review. *Artificial Intelligence Review*. 56. 1-36. 10.1007/s10462-023-10500-9.

Barros, C. (2007). Distribución y Modelamiento preliminar del Arsénico, Bismuto, Plomo, Zinc, Plata y Oro en la mineralización profunda de la Mina Los Bronces, Región Metropolitana, Chile. Memoria de título, Universidad Católica del Norte, Facultad de Ingeniería y Ciencias Geológicas, Departamento de Ciencias Geológicas, Antofagasta, Chile.

Betancourt, Gustavo. (2005). LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs). *Scientia Et Technica*.

Bruker. (s.f.). *Guía de espectroscopía infrarroja*. <https://www.bruker.com/es/products-and-solutions/infrared-and-raman/ft-ir-routine-spectrometer/what-is-ft-ir-spectroscopy.html>

Bruker. (2019). FT-IR Basics – Principles of Infrared Spectroscopy [Video] Youtube. [https://www.youtube.com/watch?v=KR0WMB3AR3s&&\\$ab\\$_\\$channel=Bruker](https://www.youtube.com/watch?v=KR0WMB3AR3s&&ab_$channel=Bruker)

Bureau Veritas Chile. (s.f). *Machine Learning*. <https://www.bureauveritas.cl/es/machine-learning>

Carrizo, D; Silva, W.; Vela, I. y Comte, D. 2012. Búsqueda del control estructural en la evolución de la permeabilidad del Yacimiento Río Blanco - Los Bronces. T1 13º Congreso Geológico Chileno.

Carrizo, D. y Comte, D. (2014). Estudio estructural 3D del proyecto los Sulfatos: geología estructural y tomografía sísmica, distrito Los Bronces, Anglo American Chile. Tecno explora SpA Applied Geoscience Solutions, reporte inédito interno para Anglo American Chile.

Cepeda, A. (1974). Geología del Área del Yacimiento Río Blanco, Provincia de Aconcagua, Instituto de Investigaciones Geológicas [Informe Inédito], Santiago, Chile.

Charrier, R., Wyss, A.R., Flynn, J.J., Swisher, C.C. III, Norell, M.A., Zapatta, F., McKenna, M. C., Novaceck, M.J. (1996). New evidence for late Mesozoic - early Cenozoic evolution of the Chilean Andes in the upper Tinguiririca valley (35o S), Central Chile. *Journal of South American Earth Sciences*, Vol. 9, No 2, pp. 1-30.

Charrier, R., Elgueta, S., Flynn, J., Gans, P., Kay, S., Muñoz, N., Wyss, AR., Zurita, E. (2002). Evidence for Cenozoic extensional basin development and tectonic inversion south of the flat-slab segment, southern Central Andes, Chile (33-36° S.L.). *J S Am Earth Sci* 15, 117-139.

Charrier, R., Bustamante, M., Comte, D., Elgueta, S., Flynn, J. J., Lturra, N., Munoz, N., Pardo, M., Thiele, R., Wyss, A. R. (2005). The Abanico extensional basin: Regional extension, chronology of tectonic inversion and relation to shallow seismic activity and Andean uplift. *Neues Jahrbuch für Geologie und Paläontologie - Abhandlungenvol.* 236. pp: 43-77.

Charrier, R., Farías, M. y Makshev, V. (2009). Evolución tectónica, paleogeográfica y metalogénica durante el Cenozoico en los Andes de Chile norte y central e implicaciones para las regiones adyacentes de Bolivia y Argentina. *Revista de la Asociación Geológica*

Argentina 65, pp. 05-35.

Clark, R. N. (1999). Spectroscopy of rocks and minerals, and principles of spectroscopy. Remote Sensing for the Earth Sciences: Manual of Remote Sensing, 3, 3–58. <https://doi.org/10.1111/j.1945-5100.2004.tb00079.x>

Contreras, A. (2003). Geología Sondajes Profundos en la Mina Los Bronces. Informe Interno de la Compañía Minera Anglo American, División Los Bronces. Superintendencia de Geología. Pp. 38. Santiago, Chile.

Contreras, Y. (2018). Análisis geológico mineralógico y químico de rocas reactivas yacimiento Río Blanco, Región de Valparaíso, Chile. [Memoria para optar al Título de Geóloga, Universidad de Concepción]. Disponible en <http://repositorio.udec.cl/xmlui/handle/11594/3112>

Dalm, M. , Buxton, M. , Ruitenbeek, F.J.A. y Voncken, J. (2014). Application of near-infrared spectroscopy to sensor based sorting of a porphyry copper ore. Minerals Engineering. 58. 7–16. 10.1016/j.mineng.2013.12.016.

Dalm, M., Buxton, M. y Ruitenbeek, F.J.A.. (2017). Discriminating ore and waste in a porphyry copper deposit using short-wavelength infrared (SWIR) hyperspectral imagery. Minerals Engineering. 105. 10-18. 10.1016/j.mineng.2016.12.013.

Davidson, P., Kamenetsky, V., Cooke, D., Frikken, P., Hollings, P., Ryan, C., Achterbergh, E., Mernagh, T., Skarmeta, J., Serrano, L. y Vargas, R. (2005). Magmatic Precursors of Hydrothermal Fluids at the Río Blanco Cu-Mo Deposit, Chile: Links to Silicate Magmas and metal Transport. Economic Geology, v. 100, pp. 963-978.

Deckart, K., Silva, W.; Spröhnle, C. y Vela, I. (2014). Timing and duration of hydrothermal activity at the Los Bronces porphyry cluster: an update. Miner Deposita. DOI: [10.1007/s00126-014-0512-9](https://doi.org/10.1007/s00126-014-0512-9).

Deckart, K., Clark, A. H., Cuadra, P. y Fanning, M. (2013). Refinement of the time-space evolution of the giant Mio-Pliocene Río Blanco-Los Bronces porphyry Cu-Mo cluster, Central Chile: new U-Pb (SHRIMP II) and Re-Os geochronology and $^{40}\text{Ar}/^{39}\text{Ar}$ thermochronology data. Miner Deposita (2013) 28: 57-79.

Deckart, K., Clark, A. H., Aguilar, C. y Vargas, R. (2005). Magmatic y hydrothermal chronology of the supergiant Río Blanco Porphyry Copper Deposit, Central Chile: Implications of an integrated UPb y $^{40}\text{Ar}-^{39}\text{Ar}$ database. Economic Geology 100, p.905-934.

Demierres, K. (2021). Modelación numérica hidrogeológica del rajo Los Bronces. [Memoria para optar al Título de Geóloga, Universidad de Chile]. Disponible en <https://repositorio.uchile.cl/handle/2250/182777>

Fock, A., (2005). Cronología y tectónica de la exhumación en el neógeno de los Andes de Chile central entre los 33° y los 34° S. Tesis para optar al grado de magíster en Ciencias, mención Geología, Departamento de Geología, Universidad de Chile, Santiago, 235p.

Fock, A., Charrier, R., Farías M., Muñoz, M. (2006). Fallas de vergencia oeste en la Cordillera Principal de Chile Central: Inversión de la cuenca de Abanico (33°-34°S). Revista de la Asociación Geológica Argentina, Publicación Especial. 6. 48-55.

Gautam, R., Vanga, S., Ariese, F., Umapathy, S. (2015). Review of multidimensional data processing approaches for Raman and infrared spectroscopy. EPJ Tech. Instrum. 2, 1–38.

Gustafson, L. y Hunt, J. (1975). The porphyry copper deposit at El Salvador, Chile. Economic Geology. 70. 857-912. 10.2113/gsecongeo.70.5.857.

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor-Flow. O'Reilly Media.

Grez, E. y Spröhnle, C. (2003). Los Bronces Porphyry Copper Target, Anglo American Chile, División Los Bronces, Gerencia de Exploraciones, Exploraciones Categoría III, [Informe interno].

Hadid, A., Chakraborty, T., Busby, D. (2024). When Geoscience Meets Generative AI and Large Language Models: Foundations, Trends, and Future Challenges. ArXiv, abs/2402.03349.

Hémera. (2019). Documento técnico: Información Geología, Geomorfología y Riesgos Naturales Sectores Los Bronces, Los Bronces Subterráneo y STP. Anexo AD-151 Adenda 1 Proyecto “Los Bronces Integrado”. Elaborado por Hémera Centro de Observación de la Tierra, Universidad Mayor para AngloAmerican.

Huet, P. (13 de abril de 2023). *Qué son las redes neuronales y sus aplicaciones*. Open-Webinars. <https://openwebinars.net/blog/que-son-las-redes-neuronales-y-sus-aplicaciones/>

Hunt, G. R. (1977). Spectral signatures of particulate minerals in the visible and near infrared. Geophysics, 42(3). <https://doi.org/10.1190/1.1440721>

Irrarrazaval, V., Sillitoe, R., Wilson, A., Toro, J. (2010). Discovery History of a Giant, High-Grade, Hypogene Porphyry Copper-Molybdenum. Deposit at Los Sulfatos, Los Bronces-Río Blanco District, Central Chile. 2010 Society of Economic Geologists, Inc. Special Publication 15, pp. 253–269.

Jara, S. (2005). Distribución y Composición de la Bornita, en relación con la Génesis de la Mina Los Bronces, Yacimiento Río Blanco-Los Bronces, Región Metropolitana, Chile. Memoria de título, Universidad de Concepción, Facultad de Ciencias Químicas, Departamento de Ciencias de la Tierra, Concepción, Chile.

Jaime Llanes & Asociados. (2019). Estudio de Impacto Ambiental "Proyecto Los Bronces Integrado". Cap. 3.10. Hidrogeología.

Jhean, G. (21 de mayo de 2024). *AI vs Generative AI: What's the Difference?*. MyCase. <https://www.mycase.com/blog/ai/ai-vs-generative-ai/>

Kraal, K. y Ayling, B. (2019). Hyperspectral Characterization of Fallon FORGE Well 21-31: New Data and Technology Applications.

Kay, S.M. and Kurtz, A. (1995). Magmatic and tectonic characterization of the El Teñiente region. Informe Inédito para CODELCO Chile, 180 p.

Keboola. (24 de agosto de 2020). *The Ultimate Guide to logistic Regression for Machine*

Learning. <https://www.keboola.com/blog/logistic-regression-machine-learning>

Kerr, A., Rafuse, H., Sparkes, G., Hinchey, J. y Sandeman, H. (2011). Visible/infrared spectroscopy (VIRS) as a research tool in economic geology; background and pilot studies from New Foundland and Labrador. 145-166.

Klohn, C. (1960). Geología de la Cordillera de los Andes de Chile Central, Provincias de Santiago, O'Higgins, Colchagua y Curicó. Instituto de Investigaciones Geológicas. Boletín N° 8. Pp. 95. Santiago, Chile.

Levi, B. (1970). Burial metamorphic episodes in the Andean geosyncline, Central Chile. Geol. Rundsch., Vol. 59, N°3, p. 994-1013.

Lois Morales, P. (2015). Interacción agua/roca durante molienda e impactos fisicoquímicos y geoquímicos en la flotación: Caso pórfido Escondida. Disponible en <https://repositorio.uchile.cl/handle/2250/134060>

Luengo González, E. A. (2018). Análisis operacional y estudio experimental base para el desarrollo de herramienta estimativa de consumo de cal-División Andina-CODELCO [Memoria de Título Para optar al Título de Ingeniero Civil Metalúrgico, Universidad de Concepción]. <http://repositorio.udec.cl/handle/11594/3413>

Maksaev, V., Munizaga, F., Zentilli, M. y Charrier, R., (2009). Fission track thermochronology of Neogene plutons in the Principal Andean Cordillera of central Chile (33°-35°S): Implications for tectonics evolution and porphyry Cu-Mo mineralization. *Andean Geology*, 36 (2), p. 153-171.

Malvern Panalytical. (n.d.). Espectrómetro de minerales ASD TerraSpec 4 | Hi-Res.

Matich, D. (2001). Redes Neuronales: Conceptos Básicos y Aplicaciones.

https://www.frro.utn.edu.ar/repositorio/catedras/quimica/5_anio/orientadora1/monografias/matich-redesneuronales.pdf

Mery, N. y Morales, N. (2023). Aprendizaje no supervisado. En Mery, N. (Ed.), *Análisis de datos en sistemas y procesos mineros*. Universidad de Chile

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill Education.

Mokari, A., Guo, S. y Bocklitz, T. (2023). Exploring the Steps of Infrared (IR) Spectral Analysis: Pre-Processing, (Classical) Data Modelling, and Deep Learning. *Molecules*. 28. 10.3390/molecules28196886.

Muñoz, M., Fuentes, F., Vergara, M., Aguirre, L., Nyström, J.; Féraud, G. y Demant, A. (2006). Abanico East Formation: petrology and geochemistry of volcanic rocks behind Cenozoic arc front in the Andean Cordillera, Central Chile (33°50'S), *Revista Geológica de Chile*, v. 33, n. 1, pp. 109- 140.

Ni, L., Xu, H. y Zhou, X. (2020). Mineral Identification and Mapping by Synthesis of Hyperspectral VNIR/SWIR and Multispectral TIR Remotely Sensed Data With Different Classifiers. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. PP. 1-1. 10.1109/JSTARS.2020.2999057.

Ortiz, D. (2020). Modelo Mineralógico de corto plazo en base a mineralogía automati-

zada TIMA-X en la mina Los Bronces, Región Metropolitana, Chile. Memoria para optar al título de Geólogo, Facultad de Ingeniería, Universidad del Desarrollo, Santiago.

Petrelli, M. (2021). *Introduction to Python in Earth Science Data Analysis: From Descriptive Statistics to Machine Learning*. Springer Nature.

Piquer, J., Skarmeta, J., Cooke, D. (2012). Structural Evolution of the Río Blanco – Los Bronces District, Andes of Central Chile: Controls on Stratigraphy, Magmatism, and Mineralization. *Society of Economic Geologists. Economic Geology*, v. 110, pp. 1995 – 2023.

Rajbanshi, S. (30 de marzo de 2021). *Everything you need to know about machine learning*. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/03/everything-you-need-to-know-about-machine-learning/>

Ramírez, I. (2018). Diseño y desarrollo de un nuevo reporte compacto y automático de planificación mina División Los Bronces, Anglo American S.A.. [Memoria para optar al título de Ingeniero Civil de Minas, Universidad de Chile]. Disponible en <https://repositorio.uchile.cl/handle/2250/169973>

Rimmelín, R. (2019). Respuesta del macizo rocoso a la minería de cielo abierto de gran profundidad. [Tesis para optar al grado de Magíster en Ciencias de la Ingeniería, Mención Ingeniería Estructural, Sísmica y Geotécnica, Universidad de Chile]. Disponible en <https://repositorio.uchile.cl/handle/2250/173916>

Rivano, S., Godoy, E., Vergara, M. y Villaroel, M. (1990). Redefinición de la Formación Farellones de los Andes de Chile Central (32°-34°S). *Revista Geológica de Chile*. N°15. Pp. 59-70. Santiago, Chile.

Rojas Verasay, Felipe. (2022). Análisis estructural de vetillas mediante mediciones de televiewer para su uso en la geología de exploración en la mina Los Bronces.

Quigley, M. y Yildirim, B. (2015). Mineral Identification and Domain Characterisation Using Two Automated Hyperspectral Core Logging Systems, Los Bronces Cu-Mo Porphyry Deposit.

Sánchez Villar, S. (2017). Logueo de alteración potásica y sericitica en el pórfido cuprífero de los sulfatos, Chile Central, usando el sistema HYLOGGER?. Disponible en <https://repositorio.uchile.cl/handle/2250/144522>

Serrano, L., Vargas, R., Stambuck, V., Aguilar, C., Galeb, M., Holmgren, C., Contreras, A., Godoy, S., Vela, I., Skewes, M. y Stern, C. (1996). The Late Miocene Early Pliocene Río Blanco - Los Bronces Copper Deposit, Central Chilean Andes. In *Andean Copper Deposits: New Discoveries, Mineralization, Styles and Metallogeny*. Society of Economic Geologists. Special Publication N°5. Pp. 119-130.

Skewes, A., Holmgren, C. y Stern, CH. (2003). The Donoso Copper-Rich, Tourmaline-Bearing Breccia Pipe in Central Chile: Petrology, Fluid Inclusion and Stable Isotope Evidence for an Origin from Magmatic Fluids. *Mineralium Deposita*. N°38. Pp. 2-21.

Stambuk, V., Serrano, L. y Blondel, J. (1982). Geología del Yacimiento Río Blanco. III Congreso Geológico Chileno, Tomo II, Concepción, Chile.

Sillitoe, R. y Perelló, J. (2005). Andean copper province: toctoomagmatic setting deposit types, metallogeny, exploration, and discovery: Economic Geology 100th Anniversary Volume.

Skewes, M. y Stern, C. (1994). Tectonic trigger for the formation of late Miocene Cu-rich breccia pipes in the Andes of central Chile. *Geology* 22, 551-554.

Sorrentino, A., Chirico, R., Corrado, F., Laukamp, C., Di Martire, D. y Mondillo, N. (2024). The application of PRISMA hyperspectral satellite imagery in the delineation of distinct hydrothermal alteration zones in the Chilean Andes: The Marimaca IOCG and the Río Blanco-Los Bronces Cu-Mo porphyry districts. *Ore Geology Reviews*, Volume 167, 105998, ISSN 0169-1368, <https://doi.org/10.1016/j.oregeorev.2024.105998>.

Stern, C., (2004). Active Andean Volcanism: its geologic and tectonic setting. *Revista Geológica de Chile*, Vol. 31 (2), p. 161-206.

Thiele, R. (1980). Hoja de Santiago. Carta Geológica. Instituto de Investigaciones Geológicas. N° 39. Pp. 51. Santiago, Chile.

Toro, J., Ortúzar, J., Zamorano, J., Cuadra, P., Hermosilla, J. Y Spronle, C. (2012). Protracted Magmatic-Hydrothermal History of the Río Blanco-Los Bronces District, Central Chile: Development of World's Greatest Known Concentration of Copper. Society of Economic Geologists, Inc. Special Publication 16, pp. 000-000.

Toro, J.C., Ortúzar, J., Makshev, V. y Barra, F. (2009). Nuevos antecedentes geocronológicos franja de pórfidos Cu-Mo del Mioceno-Plioceno, Chile Central: Implicancias metalogénicas. S11_047. 12° Congreso Geológico Chileno, Santiago.

Universidad Politécnica de Madrid. (2021). *02.3 Ratios e indicadores que califican una clasificación*. Introducción al Aprendizaje Automático. https://dcain.etsin.upm.es/~carlos/bookAA/02.3_MetodosdeClasificacion-Ratioseindicadores.html

Varela Miranda, J. I. (2019). Análisis geológico-geomecánico de pilares de roca mediante perforaciones de auscultación, en los sectores productivos Mina Esmeralda y Mina Pilar Norte, Pertenecientes a la División El Teniente, CODELCO, Región del Libertador General Bernardo O'Higgins, Chile. [Memoria para optar al título de geólogo, Universidad de Concepción]. <http://repositorio.udec.cl/handle/11594/3488>

Vargas, R. (1985). Estudio Estructural del Yacimiento Los Bronces, Región Metropolitana, Chile, Síntesis Geológica, Memoria para optar al Título de Geólogo, Universidad Católica del Norte, Departamento de Ciencias Geológicas (Inédito): 133 pp. Antofagasta.

Vargas, F. (2015). Estudio de mineralización y distribución de Ag en el sector Los Bronces del complejo porfídico Cu-Mo Río Blanco - Los Bronces, Región Metropolitana, Chile. Memoria para optar al título de Geólogo, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago.

Vargas, F., Gustafson, L., Vukasovic, M. y Skewes, A. (1999). Ore Breccias in the Río Blanco-Los Bronces Porphyry Copper Deposit, Chile. Society of Economic Geologists. Special Publication N° 7. Pp. 281-297.

Vela, I. (2014). El Rol del "techo de sulfatos" en el comportamiento Hidrogeológico y

Geotécnico de las zonas Primaria y Secundaria del Socavón Los Sulfatos. Informe Interno. Santiago: LBD Project, División Los Bronces, Anglo American

Velásquez, G. , Estay, H. , Vela, I. , Salvi, S. y Pablo, M. (2020). Metal-Selective Processing from the Los Sulfatos Porphyry-Type Deposit in Chile: Co, Au, and Re Recovery Workflows Based on Advanced Geochemical Characterization. *Minerals*. 10. 531. 10.3390/min10060531.

Vergara, M., Charrier, M., Munizaga, F., Rivano, S., Sepulveda, P., Thiele, R. y DRAKE, R. (1988). Miocene volcanism in the central Chilean Andes (31°30'S-34°35'S). *Journal of South American Earth Sciences*. Vol 1, No. 2, Pp. 199-209.

Warnaars, F., Holmgren, C. y Barassi, S. (1985). Porphyry Copper and Tourmaline Breccias at Los Bronces, Chile. *Society of Economic Geologists*. Vol. 80. Pp. 1544-1565.

Wagh, A. (25 de marzo de 2023). *What's Generative AI? Explore Underlying Layers of Machine Learning and Deep Learning*. Medium. <https://medium.com/@amol-wagh/whats-generative-ai-explore-underlying-layers-of-machine-learning-and-deep-learning-8f99272e0b0d>

Yáñez, G., Cembrano, J., Pardo, M., Ranero, C. y Selles, D., (2002). The Challenger-Juan Fernández-Maipo major tectonic transition of the Nazca-Andean subduction system at 33–34°S: Geodynamic evidence and implications. *Journal of South American Earth Sciences*, v. 15, p. 23–38.

ANEXO A

INTRODUCCIÓN

PLAN DE TRABAJO

TÍTULO DEL PROYECTO: Modelamiento espectral del contacto primario-secundario en Mno Los Bronces
 RESPONSABLE DEL PROYECTO: Thore Sotzner
 FECHA DE INICIO: 09-18
 12-15



NOMBRE DE LA TAREA	SEGUNDO SEMESTRE 2023																				
	AGOSTO			SEPTIEMBRE			OCTUBRE			NOVIEMBRE			DICIEMBRE								
	SEMANA	SEMANA	SEMANA	SEMANA	SEMANA	SEMANA	SEMANA	SEMANA	SEMANA	SEMANA	SEMANA	SEMANA	SEMANA								
	18/08	28/08	07/09	08/09	15/09	22/09	29/09	06/10	13/10	20/10	27/10	03/11	10/11	17/11	24/11	01/12	08/12	15/12	22/12	29/12	
1 DEFINICIONES INICIALES																					
1.1 Problema de investigación, objetivos, Plan de trabajo																					
1.2 Reconocimiento de la base de datos disponible																					
2 REVISIÓN BIBLIOGRÁFICA																					
2.1 Trabajos anteriores																					
2.2 Marco geológico																					
2.3 Marco tectónico																					
3 METODOLOGÍA																					
3.1.1 Obtención de la base de datos de espectros ASD-FIR (1ª parte FeMn)																					
3.1.2 Obtención de la base de datos de espectros ASD-FIR (2ª parte FeMn)																					
3.1.3 Obtención de la base de datos de espectros ASD-FIR (Sulfuro)																					
3.2 Obtención de la base de datos de mapeo geológico																					
3.3 Limpieza y hotamiento de la BBDD (1ª parte)																					
3.4 Modelo no supervisado																					
3.4.1 Caracterización de clusters																					
3.5 Limpieza y hotamiento de la BBDD (2ª parte)																					
3.5.1 Solicitud de licencia para software de modelamiento (Leapfrog)																					
3.5.2 Composición de datos																					
3.6 Modelo supervisado																					
3.7 Comparación de modelos																					
3.8 Implementación con nueva data																					
HITO				Presentación de las definiciones iniciales de la memoria y diseño general de la metodología.																	
4 RESULTADOS																					
4.1 Resultados del análisis de datos																					
4.2 Modelo no supervisado																					
4.3 Caracterización de clusters; Incluye el código para la identificación de peaks																					
4.4 Modelo supervisado																					
4.5 Modelo contacto primario-secundario																					
4.6 Comparación de modelos																					
4.7 Implementación con nueva data																					
HITO				Presentación de los resultados preliminares de la caracterización de la zona primaria y secundaria; Incluyendo modelo 3D en Leapfrog.																	
5.3 Análisis de modelos																					
6 CONCLUSIONES																					
6.1 Recomendaciones y oportunidades futuras																					
7 MEJORA CONTINUA																					
7.1 Redacción de slides finales y validación de resultados																					
7.2 Redacción final del informe de memoria																					
7.3 Entrega de la versión final del informe de memoria																					
HITO				Entrega de informe final de memoria, exposición final de los resultados y conclusiones.																	

Figura A.0.1: Plan de trabajo de la memoria de título.

ANEXO B

MARCO TEÓRICO

Mineral Group	Mineral Species (Examples)	Assignment of Absorption	Lower Limit (nm)	Upper Limit (nm)
sheet silicate	kaolinite	2νM _n OH	1392 to 1415	
sheet silicate	muscovite		1406 to 1415	
sheet silicate	lepidolite		1387 to 1425	
sheet silicate	talc		1392 to 1412	
sheet silicate	smectites		1394 to 1445	
sheet silicate	palygorskite		1386 to 1443	
sheet silicate	chlorite		1391 to 1415	
double chain silicate	calcic amphibole		1392 to 1413	
Mineral Group	Mineral Species (Examples)	Assignment of Absorption	Lower Limit (nm)	Upper Limit (nm)
sulphate	jarosite	2νM _n OH, ν + 2δH ₂ O	1471	
sulphate	alunite	2νM _n OH, ν + 2δH ₂ O	1473 (K-rich)	1491 (Na-rich)
sheet silicate	prehnite	2νM ₂ OH	1476 (Fe-rich)	1478 (Mg-rich)
sorosilicate	epidote, clinozoisite, pistasite	2νM ₂ OH	1545 (Fe-rich)	1563 (Al-rich)
NH ₄	alunite, montmorillonite, buddingtonite, white mica	ν ₁ + ν ₃ NH ₄	1520 to 1560	
sulphate	gypsum	ν + δOH/H ₂ O and rotational fundamentals or δS-O overtones	1751	
sulphate	alunite	ν + 2δM _n OH	1762 (Na-rich)	1764 (K-rich)
sulphate	jarosite	ν + 2δM _n OH	1849	
Mineral Group	Mineral Species (Examples)	Assignment of Absorption	Lower Limit (nm)	Upper Limit (nm)
talc	talc	ν + δM ₃ OH	2080	
NH ₄	alunite, montmorillonite, buddingtonite, white mica	ν ₁ + ν ₄ NH ₄	2020 to 2120	
di-oct. sheet silicate	kaolin group	ν + δM ₂ OH _o	2159	
di-oct. sheet silicate	pyrophyllite	ν + δM _n OH	2166	
sulphate	alunite	ν + δM _n OH	2172	
cyclosilicate	tourmaline	ν + δM _n OH	2174	
di-oct. sheet silicate	kaolin group	ν + δM ₂ OH _i	2209	
di-oct. sheet silicate	muscovite, phengite, paragonite	ν + δ(M) ₂ OH	2185 ([VI]Al-rich)	2215 ([VI]Al-poor)
cyclosilicate	tourmaline	ν + δM _n OH	2204	
sulphate	jarosite	ν + δM _n OH	2212	
di-oct. sheet silicate	prehnite	ν + δM _n OH	2235 ³⁾ , 2280, 2333, 2358, 2384	

Figura B.0.1: Modos vibracionales comúnmente observados en rocas y minerales de alteración, incluyendo sus posiciones en longitud de onda. ν - *stretching*/estiramiento, δ - *bending*/flexión, E - emisividad, R - reflectancia, * - Transmisión. Extraído de Laukamp et al., 2021.

Mineral Group	Mineral Species (Examples)	Assignment of Absorption	Lower Limit (nm)	Upper Limit (nm)
tri-oct. sheet silicate	clinochlore, chamosite, ripidolite	v + δ M2OH	2248 (Mg-rich)	2261 (Fe ²⁺ -rich)
tri-oct. sheet silicate	biotite, annite, phlogopite (only when Al ^{VI} is present)	v + δ M2OH	2248 (Mg-rich)	2259 (Fe ²⁺ -rich)
sorosilicate	epidote group	v + δ M _n OH		2250
cyclosilicate	tourmaline	v + δ M _n OH		2256
sulphate	jarosite	v + δ M _n OH		2263
hydroxide	gibbsite	v + δ M _n OH		2268
talc	talc	v + δ M3OH		2279
tri-oct. sheet silicate	Fe-smectite	v + δ M2OH	2288	2317
talc	talc	v + δ M3OH		2300
cyclosilicate	tourmaline	v + δ M _n OH		2302
Ca-amphibole	tremolite, actinolite	v + δ Mg3OH	2296 (Mg-rich)	2318 (Fe ²⁺ -rich)
Ca-amphibole	tremolite, actinolite	v + δ Mg3OH	2318 (Mg-rich)	2338 (Fe ²⁺ -rich)
carbonate	ankerite, calcite, dolomite, magnesite, siderite	3n3CO ₃	2300 (Mg-rich)	2340 (Ca only)
chlorite	clinochlore, chamosite, ripidolite	v + δ Mg3OH	2331 (Mg-rich)	2358 (Fe ²⁺ -rich)
di-oct. sheet silicate	muscovite, phengite, paragonite	v + δ Mg3OH	2348 ([VI]Al-rich)	2366 ([VI]Al-poor)
tri-oct. sheet silicate	biotite, annite, phlogopite	v + δ M2OH	2320 (Mg-rich)	2350 (Fe ²⁺ -rich)
sorosilicate	epidote group	v + δ M2OH		2340
cyclosilicate	tourmaline	v + δ M _n OH	2359 (ca. 5% to 11% MgO)	2375 (0% MgO)
tri-oct. sheet silicate	biotite, annite, phlogopite	v + δ M2OH	2377 (Mg-rich)	2390 (Fe ²⁺ -rich)
talc	talc	v + δ Mg3OH		2380
Ca-amphibole	tremolite, actinolite	v + δ Mg3OH	2382 (Mg-rich)	2406 (Fe ²⁺ -rich)
carbonate	ankerite, calcite, dolomite, magnesite, siderite	2v3 + v1	2505 (Mg-rich)	2541 (Ca only)

Mineral Group	Mineral Species (Examples)	Assignment of Absorption	Lower Limit (nm)	Upper Limit (nm)
all hydroxylated minerals				
		vOH	2667	5000
di-oct. sheet silicate	kaolin group	vOH	2705, 2725, 2738, 2761	
di-oct. sheet silicate	white mica	vOH	2677, 2762, 2775, 2799, 2807	
tri-oct. sheet silicate	dark mica	vOH	2668, 2677, 2680	
di-/tri-oct. sheet silicate	white and dark micas	vOH	2810 (Mg-rich)	2853 (Fe ²⁺ -rich)
double chain silicate	amphibole	vOH	2723 (Mg-rich)	2759 (Fe ²⁺ -rich)
cyclosilicate	tourmaline	vOH	2675 to 2890	
iron oxides	goethite	vH-O-H	2899	2903
di-oct. sheet silicate	Al-smectites	vH-O-H	2946 (Mg-rich)	2987 (K-rich)
all H2O-bearing minerals				
		vH-O-H		3000
iron oxides				
		vOH	3113	3131
cyclosilicate	tourmaline	2vBO ₃		3704
cyclosilicate	tourmaline	2vBO ₃		3811
carbonate	carbonate	2vCO ₃ , v + vCO ₃	3930	4150
tectosilicate	quartz	2vSiO		4470
tectosilicate	quartz	2vSiO		4670
phosphate				
		2v ₃ (PO ₄)		4805
tectosilicate	quartz	2vSiO		4920
sulphate	jarosite	2v ₃ (SO ₄) ²⁻ , 2d(OH)		4960
sulphate	jarosite	2v ₃ (SO ₄) ²⁻ , 2d(OH)		5099
tectosilicate	quartz	2vSiO		5330
di-oct. sheet silicate	kaolin group	2vSiO		5480

Figura B.0.2: Cont.

Mineral Group	Mineral Species (Examples)	Assignment of Absorption	Lower Limit (nm)	Upper Limit (nm)
carbonate	ankerite, calcite, dolomite, magnesite, siderite	ν_3 asymmetric stretch CO_3	6405R to 6598R	
NH_4 (multiple minerals)		NH bending	6993* to 7400*	
cyclosilicate	tourmaline (schorl)	$\nu_{\text{as}}(\text{BO}_3)$	7391*	
cyclosilicate	tourmaline (schorl)	$\nu_{\text{as}}(\text{BO}_3)$	7968* (7880 to 7995)	
sulphate	alunite, jarosite	$\nu_3\text{SO}_4$	7899E, 8197E	
sulphate	jarosite, barite, alunite, anhydrite	$\nu_3\text{SO}_4$	all*: 8439, 8480 to 8540, 8576, 8650	
sulphate	anhydrite, barite, alunite, jarosite	$\nu_3\text{SO}_4$	all*: 8920, 8930 to 8980, 8969, 8993	
tectosilicate	quartz	νSiO	8496R (8598*)	
tectosilicate	quartz	νSiO	9074R (9328*)	
tectosilicate	quartz	"quartz RSC feature"	8625R	
tectosilicate	K-feldspar (orthoclase)	νSiO	8739R	
single chain silicate	pyroxene (diopside)	νSiO	8960 to 9100	
phosphate	apatite	$\nu_{\text{as}}(\text{PO}_4)$	9050R (9160*)	

Mineral Group	Mineral Species (Examples)	Assignment of Absorption	Lower Limit (nm)	Upper Limit (nm)
phosphate	apatite	"apatite RSC feature"	9210R	
phosphate	apatite	$\nu_{\text{as}}(\text{PO}_4)$	9550R (9620*)	
tri-oct. sheet silicate	talc	νSiO	9280R	
tri-oct. sheet silicate	talc	"talc RSC feature"	9470R	
tri-oct. sheet silicate	talc	νSiO	9680R	
di-oct. sheet silicate	muscovite	νSiO	9130R	
tri-oct. sheet silicate	biotite, phlogopite	νSiO	9200R	
tri-oct. sheet silicate	ripidolite (clinochlore)	νSiO	9217* to 9225*	
single chain silicate	pyroxene (diopside)	νSiO	9290 to 9390	
nesosilicates	olivine (Fo#91 to Fo#0)	νSiO	9337 to 9852	
sulphate	jarosite, barite	$\nu_3\text{SO}_4$	all*: 9217, 9230 to 9260	
sulphate	alunite, jarosite, barite	$\nu_1\text{SO}_4$	all*: 9737, 9940, 10150 to 10180	
di-oct. sheet silicate	muscovite	νSiO	9600R	
di-oct. sheet silicate	montmorillonite (+- mica)	νSiO	9600* to 9710*	
tri-oct. sheet silicate	biotite, phlogopite	νSiO	9780R to 9800R	
di-oct. sheet silicate	kaolinite	νSiO	9074*, 9681*, 9891*	
tectosilicate	K-feldspar (orthoclase)	νSiO	9558R	

Figura B.0.3: Cont.

tectosilicate	plagioclase feldspars	νSiO	9660R
tectosilicate	plagioclase feldspars	νSiO	9920R
nesosilicates	vesuvianite	νSiO in Si_2O_7 group	9900R
nesosilicates	olivine (Fo#91 to Fo#0)	νSiO	10,121R to 10,537R
nesosilicates	olivine (Fo#91 to Fo#0)	$n\text{SiO}$	10,515R to 11,099R
di-oct. sheet silicate	kaolin group	δOH	10,661*
single chain silicate	pyroxene (diopside)	νSiO	10,850R
di-oct. sheet silicate	kaolin group	δOH	10,929*
nesosilicates	grandite series	$\nu\text{Si-O}$ in SiO_4 group ("B")	10,953R to 11,368R
carbonate	ankerite, calcite, dolomite, magnesite, siderite	ν_2 out-of-plane bend CO_3	11,058R to 11,372R
nesosilicates	grandite series	grandite feature "T"	11,250R to 11,600R
single chain silicate	pyroxene (enstatite)	νSiO	11,470R to 11,570R
carbonate	ankerite, calcite, dolomite, magnesite, siderite	ν_4 in-plane bend CO_3 (trough!)	13,200R to 14,000R
sulphate	anhydrite	$n_4\text{SO}_4$	14,780*

Figura B.0.4: Cont.

ANEXO C

RESULTADOS

C.1. Bosque aleatorio

Tabla C.1.1: Mejores hiperparámetros encontrados en la primera (*RandomizedSearchCV*) y segunda búsqueda (*GridSearchCV*) para el clasificador *Random Forest*, con los datos de escenario 1 (*Santiago_1_Spectrum_Con_Contacto*) y escenario 2 (*Santiago_2_Spectrum_Con_Contacto*). En color amarillo se destacan los valores distintos entre una búsqueda y otra.

RANDOM FOREST				
	Santiago_1		Santiago_2	
Mejores hiperparámetros	R.SearchCV	G.SearchCV	R.SearchCV	G.SearchCV
criterion	entropy	entropy	entropy	entropy
max_depth	None	10	None	None
max_features	0.5	0.45	0.5	0.5
min_samples_leaf	2	1	1	1
min_samples_split	5	4	2	3
n_estimators	100	100	100	110
class_weight	None	None	balanced	None
bootstrap	True	True	True	True

Tabla C.1.2: Métricas de evaluación del modelo *Random Forest*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para el escenario 2 de Perth (Perth_2).

RANDOM FOREST									
Métricas de evaluación Perth_2									
1º Búsqueda - RandomizedSearchCV					2º Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9101			Exactitud en el conjunto de prueba		0.9150		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.91	0.95	0.93	781	Secundario	0.92	0.95	0.93	781
Primario	0.91	0.84	0.87	455	Primario	0.91	0.85	0.88	455
accuracy			0.91	1236	accuracy			0.92	1236
macro avg	0.91	0.90	0.90	1236	macro avg	0.91	0.90	0.91	1236
weighted avg	0.91	0.91	0.91	1236	weighted avg	0.91	0.92	0.91	1236
Matriz de confusión					Matriz de confusión				
	Positivo		Negativo			Positivo		Negativo	
Positivo	743		38		Positivo	743		38	
Negativo	73		382		Negativo	67		388	

Tabla C.1.3: Métricas de evaluación del modelo *Random Forest*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para *Santiago_1*.

RANDOM FOREST									
Métricas de evaluación Santiago_1									
1° Búsqueda - RandomizedSearchCV					2° Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9288			Exactitud en el conjunto de prueba		0.9343		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.92	0.87	0.89	188	Secundario	0.93	0.88	0.90	188
Primario	0.93	0.96	0.95	360	Primario	0.94	0.96	0.95	360
accuracy			0.93	548	accuracy			0.93	548
macro avg	0.93	0.91	0.92	548	macro avg	0.93	0.92	0.93	548
weighted avg	0.93	0.93	0.93	548	weighted avg	0.93	0.93	0.93	548
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	163	25			Positivo	165	23		
Negativo	14	346			Negativo	13	347		

Tabla C.1.4: Métricas de evaluación del modelo *Random Forest*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para *Santiago_2*.

RANDOM FOREST									
Métricas de evaluación Santiago_2									
1° Búsqueda - RandomizedSearchCV					2° Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9325			Exactitud en el conjunto de prueba		0.9315		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.91	0.91	0.91	373	Secundario	0.90	0.91	0.91	373
Primario	0.95	0.95	0.95	635	Primario	0.95	0.94	0.95	635
accuracy			0.93	1008	accuracy			0.93	1008
macro avg	0.93	0.93	0.93	1008	macro avg	0.93	0.93	0.93	1008
weighted avg	0.93	0.93	0.93	1008	weighted avg	0.93	0.93	0.93	1008
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	338	35			Positivo	340	33		
Negativo	33	602			Negativo	36	599		

C.2. Máquina de vectores de soporte

Tabla C.2.1: Mejores hiperparámetros encontrados en la primera (*RandomizedSearchCV*) y segunda búsqueda (*GridSearchCV*) para el clasificador *Support Vector Machine*, con los datos de Perth en escenario 3 (*Perth_Spectrum_Con_Contacto*), escenario 1 (*Perth_1_Spectrum_Con_Contacto*) y escenario 2 (*Perth_2_Spectrum_Con_Contacto*). En color amarillo se destacan los valores distintos entre una búsqueda y otra.

SUPPORT VECTOR MACHINE						
	Perth		Perth_1		Perth_2	
Mejores hiperparámetros	R.SearchCV	G.SearchCV	R.SearchCV	G.SearchCV	R.SearchCV	G.SearchCV
kernel	linear	linear	linear	linear	linear	linear
gamma	10	-	10	-	100	-
degree	4	-	4	-	3	-
class_weight	None	None	None	None	balanced	balanced
C	0.1	0.05	0.1	0.15	0.1	0.05

Tabla C.2.2: Mejores hiperparámetros encontrados en la primera (*RandomizedSearchCV*) y segunda búsqueda (*GridSearchCV*) para el clasificador *Support Vector Machine*, con los datos de Santiago en escenario 3 (*Santiago_Spectrum_Con_Contacto*), escenario 1 (*Santiago_1_Spectrum_Con_Contacto*) y escenario 2 (*Santiago_2_Spectrum_Con_Contacto*). En color amarillo se destacan los valores distintos entre una búsqueda y otra.

SUPPORT VECTOR MACHINE						
	Santiago		Santiago_1		Santiago_2	
Mejores hiperparámetros	R.SearchCV	G.SearchCV	R.SearchCV	G.SearchCV	R.SearchCV	G.SearchCV
kernel	linear	linear	linear	linear	linear	linear
gamma	10	-	10	-	0.1	-
degree	4	-	4	-	2	-
class_weight	None	None	None	None	balanced	balanced
C	0.1	0.1	0.1	0.15	1.0	1.0

Tabla C.2.3: Métricas de evaluación del modelo *Support Vector Machine*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para Perth.

SUPPORT VECTOR MACHINE									
Métricas de evaluación Perth									
1° Búsqueda - RandomizedSearchCV					2° Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9360			Exactitud en el conjunto de prueba		0.9344		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.94	0.96	0.95	782	Secundario	0.95	0.95	0.95	782
Primario	0.93	0.90	0.91	454	Primario	0.92	0.91	0.91	454
accuracy			0.94	1236	accuracy			0.93	1236
macro avg	0.93	0.93	0.93	1236	macro avg	0.93	0.93	0.93	1236
weighted avg	0.94	0.94	0.94	1236	weighted avg	0.93	0.93	0.93	1236
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	749	33			Positivo	744	38		
Negativo	46	408			Negativo	43	411		

Tabla C.2.4: Métricas de evaluación del modelo *Support Vector Machine*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para Perth_1.

SUPPORT VECTOR MACHINE									
Métricas de evaluación Perth_1									
1º Búsqueda - RandomizedSearchCV					2º Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9539			Exactitud en el conjunto de prueba		0.9539		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.93	0.99	0.96	171	Secundario	0.93	0.99	0.96	171
Primario	0.99	0.90	0.94	133	Primario	0.99	0.90	0.94	133
accuracy			0.95	304	accuracy			0.95	304
macro avg	0.96	0.95	0.95	304	macro avg	0.96	0.95	0.95	304
weighted avg	0.96	0.95	0.95	304	weighted avg	0.96	0.95	0.95	304
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	170	1			Positivo	170	1		
Negativo	13	120			Negativo	13	120		

Tabla C.2.5: Métricas de evaluación del modelo *Support Vector Machine*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para *Perth_2*.

SUPPORT VECTOR MACHINE									
Métricas de evaluación Perth_2									
1º Búsqueda - RandomizedSearchCV					2º Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9498			Exactitud en el conjunto de prueba		0.9498		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.96	0.96	0.96	781	Secundario	0.96	0.97	0.96	781
Primario	0.94	0.94	0.94	455	Primario	0.94	0.92	0.93	455
accuracy			0.95	1236	accuracy			0.95	1236
macro avg	0.95	0.95	0.95	1236	macro avg	0.95	0.94	0.95	1236
weighted avg	0.95	0.95	0.95	1236	weighted avg	0.96	0.95	0.95	1236
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	753	28			Positivo	754	27		
Negativo	34	421			Negativo	35	420		

Tabla C.2.6: Métricas de evaluación del modelo *Support Vector Machine*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para Santiago.

SUPPORT VECTOR MACHINE									
Métricas de evaluación Santiago									
1º Búsqueda - RandomizedSearchCV					2º Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9474			Exactitud en el conjunto de prueba		0.9474		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.92	0.94	0.93	367	Secundario	0.92	0.94	0.93	367
Primario	0.96	0.95	0.96	641	Primario	0.96	0.95	0.96	641
accuracy			0.95	1008	accuracy			0.95	1008
macro avg	0.94	0.95	0.94	1008	macro avg	0.94	0.95	0.94	1008
weighted avg	0.95	0.95	0.95	1008	weighted avg	0.95	0.95	0.95	1008
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	344	23			Positivo	344	23		
Negativo	30	611			Negativo	30	611		

Tabla C.2.7: Métricas de evaluación del modelo *Support Vector Machine*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para Santiago_1.

SUPPORT VECTOR MACHINE									
Métricas de evaluación Santiago_1									
1º Búsqueda - RandomizedSearchCV					2º Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9580			Exactitud en el conjunto de prueba		0.9598		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.94	0.94	0.94	188	Secundario	0.94	0.94	0.94	188
Primario	0.97	0.97	0.97	360	Primario	0.97	0.97	0.97	360
accuracy			0.96	548	accuracy			0.96	548
macro avg	0.95	0.94	0.94	548	macro avg	0.96	0.96	0.96	548
weighted avg	0.96	0.96	0.96	548	weighted avg	0.96	0.96	0.96	548
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	177	11			Positivo	177	11		
Negativo	12	348			Negativo	11	349		

Tabla C.2.8: Métricas de evaluación del modelo *Support Vector Machine*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearch* para Santiago_2.

SUPPORT VECTOR MACHINE										
Métricas de evaluación Santiago_2										
1º Búsqueda - RandomizedSearchCV					2º Búsqueda - GridSearchCV					
Exactitud en el conjunto de prueba					0.9543	Exactitud en el conjunto de prueba				
Reporte de clasificación						Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support	
Secundario	0.92	0.96	0.94	373	Secundario	0.92	0.96	0.94	373	
Primario	0.97	0.95	0.96	360	Primario	0.97	0.95	0.96	360	
accuracy			0.95	548	accuracy			0.95	548	
macro avg	0.95	0.95	0.95	548	macro avg	0.95	0.95	0.95	548	
weighted avg	0.95	0.95	0.95	548	weighted avg	0.96	0.95	0.95	548	
Matriz de confusión						Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo			
Positivo	357	16			Positivo	357	16			
Negativo	30	605			Negativo	30	605			

C.3. Regresión Logística

Tabla C.3.1: Mejores hiperparámetros encontrados en la primera (*RandomizedSearchCV*) y segunda búsqueda (*Grid-SearchCV*) para el clasificador *Logistic Regression*, con los datos de Perth en escenario 3 (*Perth_Spectrum_Con_Contacto*), escenario 1 (*Perth_1_Spectrum_Con_Contacto*) y escenario 2 (*Perth_2_Spectrum_Con_Contacto*). En color amarillo se destacan los valores distintos entre una búsqueda y otra.

LOGISTIC REGRESSION						
	Perth		Perth_1		Perth_2	
Mejores hiperparámetros	R.SearchCV	G.SearchCV	R.SearchCV	G.SearchCV	R.SearchCV	G.SearchCV
penalty	l1	l1	l2	l2	l2	l2
max_iter	3000	2500	2000	1500	2000	1500
learning_rate	adaptive	adaptive	adaptive	adaptive	adaptive	adaptive
eta0	0.001	0.0005	0.001	0.0005	0.001	0.002
class_weight	None	None	None	None	None	balanced
alpha	0.001	0.0005	0.0001	1e-0.5	0.0001	1e-0.5

Tabla C.3.2: Mejores hiperparámetros encontrados en la primera (*RandomizedSearchCV*) y segunda búsqueda (*GridSearchCV*) para el clasificador *Logistic Regression*, con los datos de Santiago en escenario 3 (Santiago_Spectrum_Con_Contacto), escenario 1 (Santiago_1_Spectrum_Con_Contacto) y escenario 2 (Santiago_2_Spectrum_Con_Contacto). En color amarillo se destacan los valores distintos entre una búsqueda y otra.

LOGISTIC REGRESSION						
	Santiago		Santiago_1		Santiago_2	
Mejores hiperparámetros	R.SearchCV	G.SearchCV	R.SearchCV	G.SearchCV	R.SearchCV	G.SearchCV
penalty	l1	l1	elasticnet	elasticnet	l1	l1
max_iter	3000	2500	2000	1500	3000	2500
learning_rate	adaptive	adaptive	adaptive	adaptive	adaptive	adaptive
eta0	0.001	0.0015	0.1	0.1	0.001	0.0005
class_weight	None	None	balanced	balanced	None	None
alpha	0.001	0.0015	0.01	0.01	0.001	0.0005

Tabla C.3.3: Métricas de evaluación del modelo *Logistic Regression*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para Perth.

LOGISTIC REGRESSION									
Métricas de evaluación Perth									
1° Búsqueda - RandomizedSearchCV					2° Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9263			Exactitud en el conjunto de prueba		0.9263		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.93	0.95	0.94	782	Secundario	0.93	0.95	0.94	782
Primario	0.92	0.88	0.90	454	Primario	0.91	0.88	0.90	454
accuracy			0.93	1236	accuracy			0.93	1236
macro avg	0.92	0.92	0.92	1236	macro avg	0.92	0.92	0.92	1236
weighted avg	0.93	0.93	0.93	1236	weighted avg	0.93	0.93	0.93	1236
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	754	37			Positivo	744	38		
Negativo	54	400			Negativo	53	401		

Tabla C.3.4: Métricas de evaluación del modelo *Logistic Regression*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para *Perth_1*.

LOGISTIC REGRESSION									
Métricas de evaluación Perth_1									
1º Búsqueda - RandomizedSearchCV					2º Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9276			Exactitud en el conjunto de prueba		0.9309		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.92	0.95	0.94	171	Secundario	0.92	0.96	0.94	171
Primario	0.94	0.89	0.92	133	Primario	0.94	0.89	0.92	133
accuracy			0.93	304	accuracy			0.93	304
macro avg	0.93	0.92	0.93	304	macro avg	0.93	0.93	0.93	304
weighted avg	0.93	0.93	0.93	304	weighted avg	0.93	0.93	0.93	304
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	163	8			Positivo	164	7		
Negativo	14	119			Negativo	14	119		

Tabla C.3.5: Métricas de evaluación del modelo *Logistic Regression*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para *Perth_2*.

LOGISTIC REGRESSION									
Métricas de evaluación Perth_2									
1º Búsqueda - RandomizedSearchCV					2º Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9417			Exactitud en el conjunto de prueba		0.9377		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.95	0.96	0.95	781	Secundario	0.96	0.94	0.95	781
Primario	0.93	0.91	0.92	455	Primario	0.91	0.93	0.92	455
accuracy			0.94	1236	accuracy			0.94	1236
macro avg	0.94	0.93	0.94	1236	macro avg	0.93	0.94	0.93	1236
weighted avg	0.94	0.94	0.94	1236	weighted avg	0.94	0.94	0.94	1236
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	752	29			Positivo	738	43		
Negativo	43	412			Negativo	34	421		

Tabla C.3.6: Métricas de evaluación del modelo *Logistic Regression*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para Santiago.

LOGISTIC REGRESSION									
Métricas de evaluación Santiago									
1° Búsqueda - RandomizedSearchCV					2° Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9494			Exactitud en el conjunto de prueba		0.9503		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.93	0.93	0.93	367	Secundario	0.93	0.94	0.93	367
Primario	0.96	0.96	0.96	641	Primario	0.96	0.96	0.96	641
accuracy			0.95	1008	accuracy			0.95	1008
macro avg	0.94	0.95	0.95	1008	macro avg	0.95	0.95	0.95	1008
weighted avg	0.95	0.95	0.95	1008	weighted avg	0.95	0.95	0.95	1008
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	343	24			Positivo	344	23		
Negativo	27	614			Negativo	27	614		

Tabla C.3.7: Métricas de evaluación del modelo *Logistic Regression*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para Santiago_1.

LOGISTIC REGRESSION											
Métricas de evaluación Santiago_1											
1º Búsqueda - RandomizedSearchCV					2º Búsqueda - GridSearchCV						
Exactitud en el conjunto de prueba					0.9489	Exactitud en el conjunto de prueba					0.9489
Reporte de clasificación					Reporte de clasificación						
	precision	recall	f1-score	support		precision	recall	f1-score	support		
Secundario	0.91	0.94	0.93	188	Secundario	0.91	0.94	0.93	188		
Primario	0.97	0.95	0.96	360	Primario	0.97	0.95	0.96	360		
accuracy			0.95	548	accuracy			0.95	548		
macro avg	0.94	0.95	0.94	548	macro avg	0.94	0.95	0.95	548		
weighted avg	0.95	0.95	0.95	548	weighted avg	0.95	0.95	0.95	548		
Matriz de confusión					Matriz de confusión						
	Positivo	Negativo				Positivo	Negativo				
Positivo	177	11			Positivo	177	11				
Negativo	17	343			Negativo	17	343				

Tabla C.3.8: Métricas de evaluación del modelo *Logistic Regression*, derivadas de evaluar en el conjunto de prueba los mejores hiperparámetros encontrados con *RandomizedSearchCV* y *GridSearchCV* para Santiago_2.

LOGISTIC REGRESSION									
Métricas de evaluación Santiago_2									
1º Búsqueda - RandomizedSearchCV					2º Búsqueda - GridSearchCV				
Exactitud en el conjunto de prueba		0.9484			Exactitud en el conjunto de prueba		0.9464		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.93	0.94	0.93	373	Secundario	0.92	0.94	0.93	373
Primario	0.96	0.96	0.96	635	Primario	0.96	0.95	0.96	635
accuracy			0.95	1008	accuracy			0.95	1008
macro avg	0.94	0.95	0.94	1008	macro avg	0.94	0.94	0.94	1008
weighted avg	0.95	0.95	0.95	1008	weighted avg	0.95	0.95	0.95	1008
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	349	24			Positivo	349	24		
Negativo	28	607			Negativo	30	605		

C.4. Redes neuronales

Tabla C.4.1: Métricas de evaluación del modelo *Neural Network* al evaluar en el conjunto de prueba los mejores hiperparámetros para Perth, Perth_1 y Perth_2.

NEURAL NETWORK														
Métricas de evaluación														
Perth				Perth_1				Perth_2						
Exactitud en el conjunto de prueba				0.922	Exactitud en el conjunto de prueba				0.901	Exactitud en el conjunto de prueba				0.937
Reporte de clasificación				Reporte de clasificación				Reporte de clasificación						
	precision	recall	f1-score	support		precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.94	0.94	0.94	782	Secundario	0.89	0.94	0.91	171	Secundario	0.94	0.96	0.95	781
Primario	0.89	0.89	0.89	454	Primario	0.92	0.85	0.88	133	Primario	0.94	0.89	0.91	455
accuracy			0.92	1236	accuracy			0.90	304	accuracy			0.94	1236
macro avg	0.92	0.92	0.92	1236	macro avg	0.90	0.90	0.90	304	macro avg	0.94	0.93	0.93	1236
weighted avg	0.92	0.92	0.92	1236	weighted avg	0.90	0.90	0.90	304	weighted avg	0.94	0.94	0.94	1236
Matriz de confusión				Matriz de confusión				Matriz de confusión						
	Positivo	Negativo			Positivo	Negativo			Positivo	Negativo				
Positivo	734	48		Positivo	161	10		Positivo	753	28				
Negativo	48	406		Negativo	20	113		Negativo	49	406				

Tabla C.4.2: Métricas de evaluación del modelo *Neural Network* al evaluar en el conjunto de prueba los mejores hiperparámetros para Santiago, Santiago_1 y Santiago_2.

NEURAL NETWORK														
Métricas de evaluación														
Santiago				Santiago_1				Santiago_2						
Exactitud en el conjunto de prueba		0.933		Exactitud en el conjunto de prueba		0.943		Exactitud en el conjunto de prueba		0.943				
Reporte de clasificación				Reporte de clasificación				Reporte de clasificación						
	precision	recall	f1-score	support		precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.91	0.90	0.91	367	Secundario	0.92	0.92	0.92	188	Secundario	0.92	0.92	0.92	373
Primario	0.94	0.95	0.95	641	Primario	0.96	0.96	0.96	133	Primario	0.96	0.95	0.96	635
accuracy			0.93	1008	accuracy			0.94	548	accuracy			0.94	1008
macro avg	0.93	0.93	0.93	1008	macro avg	0.94	0.94	0.94	548	macro avg	0.94	0.94	0.94	1008
weighted avg	0.93	0.93	0.93	1008	weighted avg	0.94	0.94	0.94	548	weighted avg	0.94	0.94	0.94	1008
Matriz de confusión				Matriz de confusión				Matriz de confusión						
	Positivo	Negativo			Positivo	Negativo			Positivo	Negativo				
Positivo	331	36		Positivo	173	15		Positivo	345	28				
Negativo	31	610		Negativo	16	344		Negativo	29	606				

C.5. K-means

Tabla C.5.1: Métricas de evaluación del modelo *K-means* para Perth, Perth_1 y Perth_2.

K-MEANS								
Métricas de evaluación								
Perth			Perth_1			Perth_2		
Homogeneidad	0.0092		Homogeneidad	0.0083		Homogeneidad	0.0092	
Compleitud	0.0089		Compleitud	0.0082		Compleitud	0.0089	
Puntuación V-measure	0.0090		Puntuación V-measure	0.0083		Puntuación V-measure	0.0090	
Matriz de confusión			Matriz de confusión			Matriz de confusión		
	Positivo	Negativo		Positivo	Negativo		Positivo	Negativo
Positivo	1482	2413	Positivo	559	373	Positivo	2412	1483
Negativo	1126	1155	Negativo	288	298	Negativo	1155	1126

Tabla C.5.2: Métricas de evaluación de los modelos *K-means* para Santiago, Santiago_1 y Santiago_2.

K-MEANS								
Métricas de evaluación								
Santiago			Santiago_1			Santiago_2		
Homogeneidad	0.0005		Homogeneidad	0.0005		Homogeneidad	0.0004	
Compleitud	0.0005		Compleitud	0.0005		Compleitud	0.0004	
Puntuación V-measure	0.0005		Puntuación V-measure	0.0005		Puntuación V-measure	0.0004	
Matriz de confusión			Matriz de confusión			Matriz de confusión		
	Positivo	Negativo		Positivo	Negativo		Positivo	Negativo
Positivo	753	1093	Positivo	425	546	Positivo	1095	751
Negativo	1390	1804	Negativo	722	1043	Negativo	1813	1381

C.6. Comparación de modelos

Tabla C.6.1: Métricas de evaluación en los modelos de Perth_1 al evaluar en el conjunto de validación.

COMPARACIÓN DE MODELOS								
Perth_1								
RANDOM FOREST			LOGISTIC REGRESSION			SUPPORT VECTOR MACHINE		
Exactitud en el set de validación	0.87		Exactitud en el set de validación	0.93		Exactitud en el set de validación	0.91	
Reporte de clasificación			Reporte de clasificación			Reporte de clasificación		
Precisión	0.86		Precisión	0.92		Precisión	0.91	
Recall	0.83		Recall	0.93		Recall	0.89	
F1-score	0.85		F1-score	0.92		F1-score	0.90	
Matriz de confusión			Matriz de confusión			Matriz de confusión		
	Positivo	Negativo		Positivo	Negativo		Positivo	Negativo
Positivo	120	8	Positivo	121	7	Positivo	122	6
Negativo	16	46	Negativo	5	57	Negativo	10	52

Tabla C.6.2: Comparación final de métricas de evaluación en los modelos *Logistic Regression* y *Neural Network* de Perth_1 al evaluar en el conjunto de prueba.

COMPARACIÓN DE MODELOS II											
Perth_1											
LOGISTIC REGRESSION					NEURAL NETWORK						
Exactitud en el conjunto de prueba					0.93	Exactitud en el conjunto de prueba					0.90
Reporte de clasificación					Reporte de clasificación						
	precision	recall	f1-score	support		precision	recall	f1-score	support		
Secundario	0.92	0.96	0.94	171	Secundario	0.89	0.94	0.91	171		
Primario	0.94	0.89	0.92	133	Primario	0.92	0.85	0.88	133		
accuracy			0.93	304	accuracy			0.90	304		
macro avg	0.93	0.93	0.93	304	macro avg	0.90	0.90	0.90	304		
weighted avg	0.93	0.93	0.93	304	weighted avg	0.90	0.90	0.90	304		
Matriz de confusión					Matriz de confusión						
	Positivo	Negativo				Positivo	Negativo				
Positivo	164	7			Positivo	161	10				
Negativo	14	119			Negativo	20	113				

Tabla C.6.3: Métricas de evaluación en los modelos de Perth_2 al evaluar en el conjunto de validación.

COMPARACIÓN DE MODELOS								
Perth_2								
RANDOM FOREST			LOGISTIC REGRESSION			SUPPORT VECTOR MACHINE		
Exactitud en el set de validación		0.901	Exactitud en el set de validación		0.931	Exactitud en el set de validación		0.939
Reporte de clasificación			Reporte de clasificación			Reporte de clasificación		
Precisión		0.90	Precisión		0.93	Precisión		0.93
Recall		0.88	Recall		0.91	Recall		0.93
F1-score		0.89	F1-score		0.92	F1-score		0.93
Matriz de confusión			Matriz de confusión			Matriz de confusión		
	Positivo	Negativo		Positivo	Negativo		Positivo	Negativo
Positivo	462	18	Positivo	464	16	Positivo	462	18
Negativo	58	234	Negativo	37	255	Negativo	29	263

Tabla C.6.4: Comparación final de métricas de evaluación en los modelos *Support Vector Machine* y *Neural Network* de Perth_2 al evaluar en el conjunto de prueba.

COMPARACIÓN DE MODELOS II									
Perth_2									
SUPPORT VECTOR MACHINE					NEURAL NETWORK				
Exactitud en el conjunto de prueba		0.941			Exactitud en el conjunto de prueba		0.937		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.95	0.96	0.95	781	Secundario	0.94	0.96	0.95	781
Primario	0.93	0.91	0.92	455	Primario	0.94	0.89	0.91	455
accuracy			0.94	1236	accuracy			0.94	1236
macro avg	0.94	0.93	0.94	1236	macro avg	0.94	0.93	0.93	1236
weighted avg	0.94	0.94	0.94	1236	weighted avg	0.94	0.94	0.94	1236
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	752	29			Positivo	753	28		
Negativo	43	412			Negativo	49	406		

Tabla C.6.5: Métricas de evaluación en los modelos de Santiago_1 al evaluar en el conjunto de validación.

COMPARACIÓN DE MODELOS								
Santiago_1								
RANDOM FOREST			LOGISTIC REGRESSION			SUPPORT VECTOR MACHINE		
Exactitud en el set de validación		0.932	Exactitud en el set de validación		0.941	Exactitud en el set de validación		0.941
Reporte de clasificación			Reporte de clasificación			Reporte de clasificación		
Precisión		0.92	Precisión		0.93	Precisión		0.93
Recall		0.92	Recall		0.93	Recall		0.93
F1-score		0.92	F1-score		0.93	F1-score		0.93
Matriz de confusión			Matriz de confusión			Matriz de confusión		
	Positivo	Negativo		Positivo	Negativo		Positivo	Negativo
Positivo	114	12	Positivo	118	8	Positivo	117	9
Negativo	11	205	Negativo	12	204	Negativo	11	205

Tabla C.6.6: Comparación final de métricas de evaluación en los modelos *Logistic Regression* y *Neural Network* de Santiago_1 al evaluar en el conjunto de prueba.

COMPARACIÓN DE MODELOS II									
Santiago_1									
LOGISTIC REGRESSION					NEURAL NETWORK				
Exactitud en el conjunto de prueba		0.948			Exactitud en el conjunto de prueba		0.943		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.91	0.94	0.93	188	Secundario	0.92	0.92	0.92	188
Primario	0.97	0.95	0.96	360	Primario	0.96	0.96	0.96	360
accuracy			0.95	548	accuracy			0.94	548
macro avg	0.94	0.95	0.94	548	macro avg	0.94	0.94	0.94	548
weighted avg	0.95	0.95	0.95	548	weighted avg	0.94	0.94	0.94	548
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	177	11			Positivo	173	15		
Negativo	17	343			Negativo	16	344		

Tabla C.6.7: Métricas de evaluación en los modelos de Santiago_2 al evaluar en el conjunto de validación.

COMPARACIÓN DE MODELOS								
Santiago_2								
RANDOM FOREST			LOGISTIC REGRESSION			SUPPORT VECTOR MACHINE		
Exactitud en el set de validación		0.915	Exactitud en el set de validación		0.938	Exactitud en el set de validación		0.936
Reporte de clasificación			Reporte de clasificación			Reporte de clasificación		
Precisión		0.91	Precisión		0.93	Precisión		0.92
Recall		0.90	Recall		0.93	Recall		0.93
F1-score		0.90	F1-score		0.93	F1-score		0.93
Matriz de confusión			Matriz de confusión			Matriz de confusión		
	Positivo	Negativo		Positivo	Negativo		Positivo	Negativo
Positivo	193	30	Positivo	203	20	Positivo	207	16
Negativo	23	384	Negativo	19	388	Negativo	24	383

Tabla C.6.8: Comparación final de métricas de evaluación en los modelos *Logistic Regression* y *Neural Network* de Santiago_2 al evaluar en el conjunto de prueba.

COMPARACIÓN DE MODELOS II									
Santiago_2									
LOGISTIC REGRESSION					NEURAL NETWORK				
Exactitud en el conjunto de prueba		0.948			Exactitud en el conjunto de prueba		0.943		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.93	0.94	0.93	373	Secundario	0.92	0.92	0.92	373
Primario	0.96	0.96	0.96	635	Primario	0.96	0.95	0.96	635
accuracy			0.95	1008	accuracy			0.94	1008
macro avg	0.94	0.95	0.94	1008	macro avg	0.94	0.94	0.94	1008
weighted avg	0.95	0.95	0.95	1008	weighted avg	0.94	0.94	0.94	1008
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	349	24			Positivo	345	28		
Negativo	28	607			Negativo	29	606		

Tabla C.6.9: Métricas de evaluación en los modelos de Perth al evaluar en el conjunto de validación.

COMPARACIÓN DE MODELOS					
Perth					
LOGISTIC REGRESSION			SUPPORT VECTOR MACHINE		
Exactitud en el set de validación	0.92		Exactitud en el set de validación	0.94	
Reporte de clasificación			Reporte de clasificación		
Precisión	0.92		Precisión	0.94	
Recall	0.91		Recall	0.94	
F1-score	0.92		F1-score	0.94	
Matriz de confusión			Matriz de confusión		
	Positivo	Negativo		Positivo	Negativo
Positivo	470	24	Positivo	475	19
Negativo	32	246	Negativo	21	257

Tabla C.6.10: Comparación final de métricas de evaluación en los modelos *Support Vector Machine* y *Neural Network* de Perth al evaluar en el conjunto de prueba.

COMPARACIÓN DE MODELOS II									
Perth									
SUPPORT VECTOR MACHINE					NEURAL NETWORK				
Exactitud en el conjunto de prueba		0.936			Exactitud en el conjunto de prueba		0.922		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.94	0.96	0.95	782	Secundario	0.94	0.94	0.94	782
Primario	0.93	0.90	0.91	454	Primario	0.89	0.89	0.89	454
accuracy			0.94	1236	accuracy			0.92	1236
macro avg	0.93	0.93	0.93	1236	macro avg	0.92	0.92	0.92	1236
weighted avg	0.94	0.94	0.94	1236	weighted avg	0.92	0.92	0.92	1236
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	749	33			Positivo	734	48		
Negativo	46	408			Negativo	48	406		

Tabla C.6.11: Métricas de evaluación en los modelos de Santiago al evaluar en el conjunto de validación.

COMPARACIÓN DE MODELOS					
Santiago					
LOGISTIC REGRESSION			SUPPORT VECTOR MACHINE		
Exactitud en el set de validación	0.936		Exactitud en el set de validación	0.938	
Reporte de clasificación			Reporte de clasificación		
Precisión	0.931		Precisión	0.932	
Recall	0.929		Recall	0.933	
F1-score	0.930		F1-score	0.932	
Matriz de confusión			Matriz de confusión		
	Positivo	Negativo		Positivo	Negativo
Positivo	204	21	Positivo	206	19
Negativo	19	386	Negativo	20	385

Tabla C.6.12: Comparación final de métricas de evaluación en los modelos *Support Vector Machine* y *Neural Network* de Santiago al evaluar en el conjunto de prueba.

COMPARACIÓN DE MODELOS II									
Santiago									
SUPPORT VECTOR MACHINE					NEURAL NETWORK				
Exactitud en el conjunto de prueba		0.947			Exactitud en el conjunto de prueba		0.933		
Reporte de clasificación					Reporte de clasificación				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Secundario	0.92	0.94	0.93	367	Secundario	0.91	0.90	0.91	367
Primario	0.96	0.95	0.96	641	Primario	0.94	0.95	0.95	641
accuracy			0.95	1008	accuracy			0.93	1008
macro avg	0.94	0.95	0.94	1008	macro avg	0.93	0.93	0.93	1008
weighted avg	0.95	0.95	0.95	1008	weighted avg	0.93	0.93	0.93	1008
Matriz de confusión					Matriz de confusión				
	Positivo	Negativo				Positivo	Negativo		
Positivo	344	23			Positivo	331	36		
Negativo	30	611			Negativo	31	610		

ANEXO D

RESULTADOS

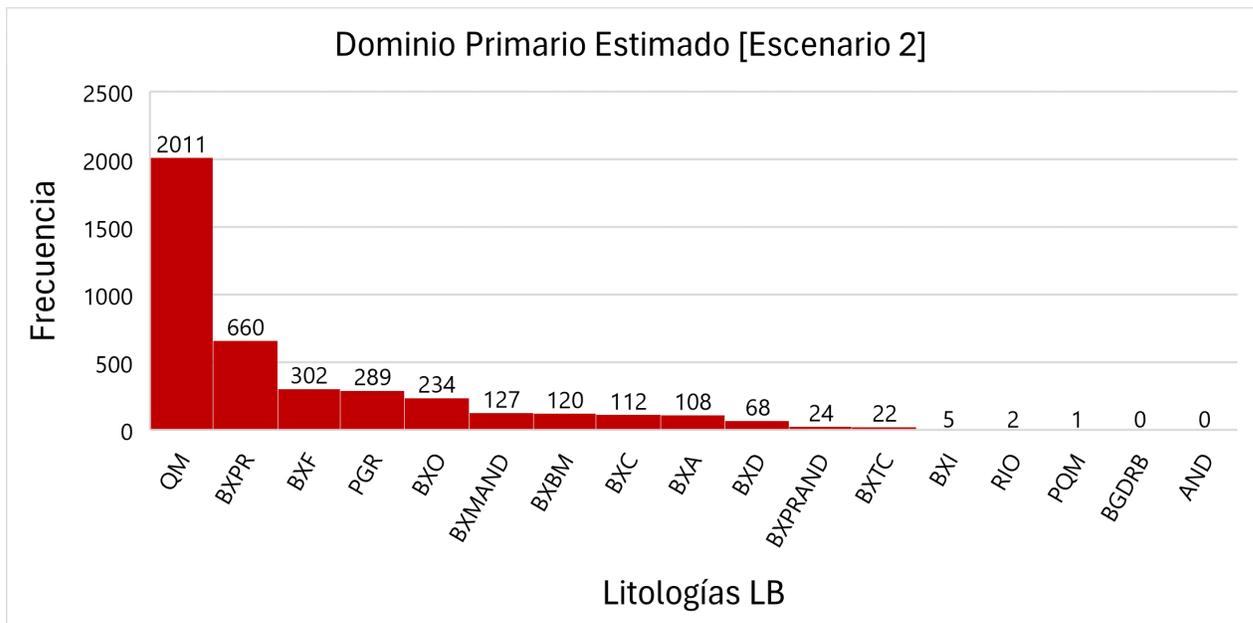


Figura D.0.1: Distribución de las litologías presentes en el dominio primario, basado en la clasificación del dominio a partir de los modelos del escenario 2.

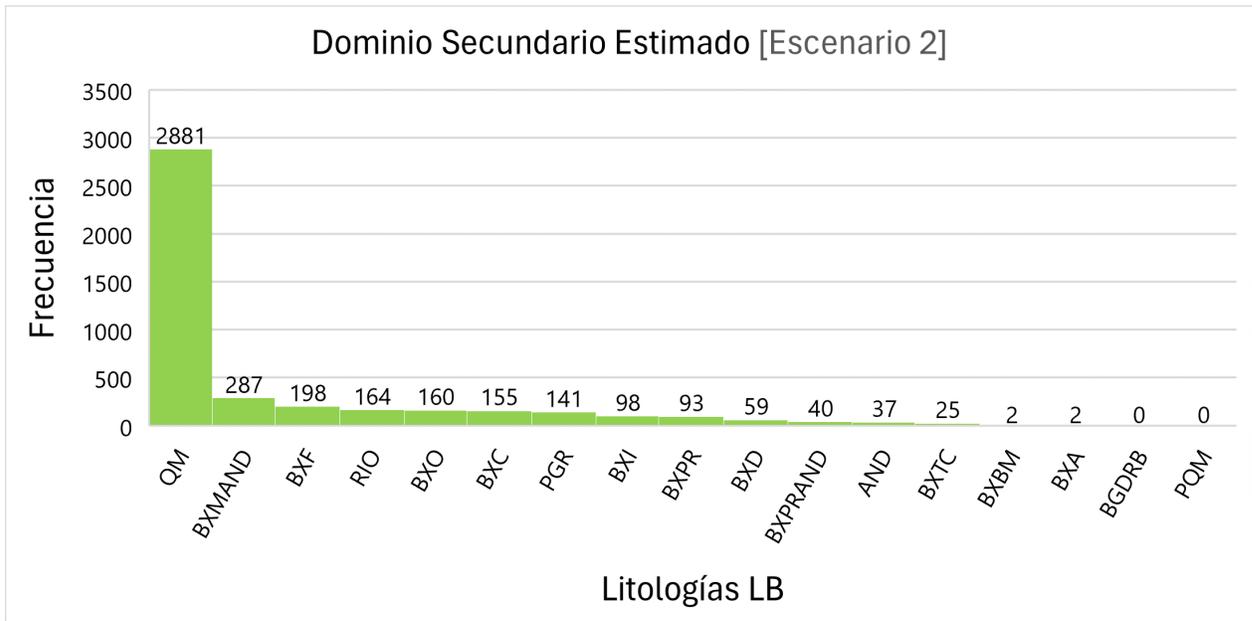


Figura D.0.2: Distribución de las litologías presentes en el dominio secundario, basado en la clasificación del dominio a partir de los modelos del escenario 2.

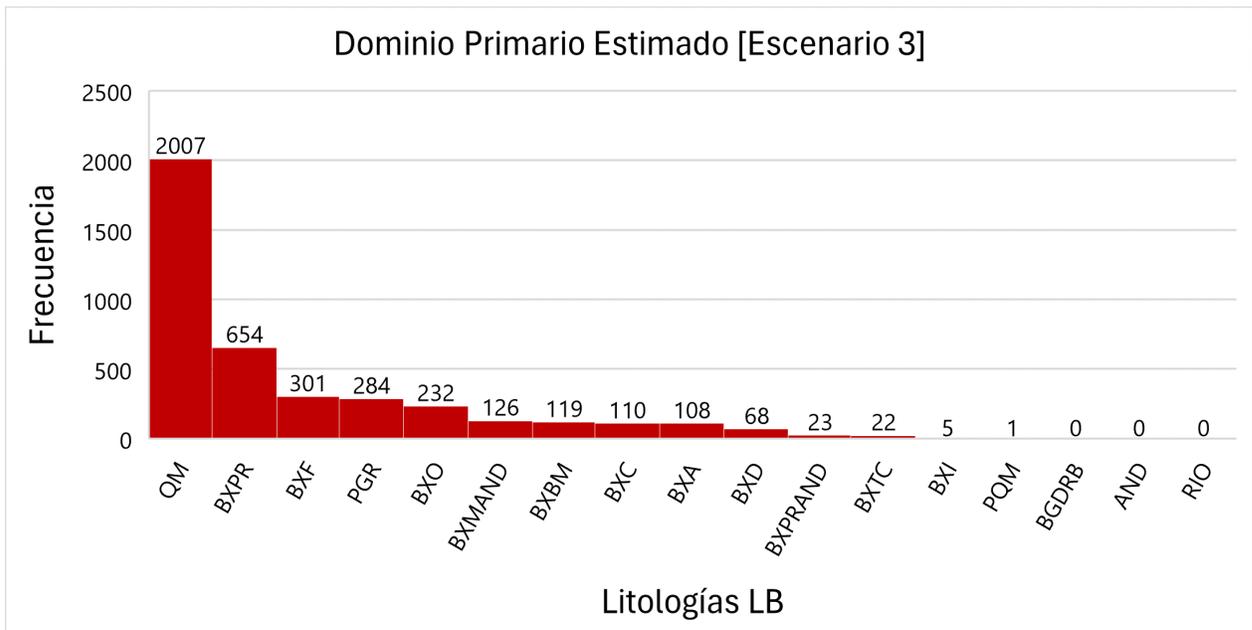


Figura D.0.3: Distribución de las litologías presentes en el dominio primario, basado en la clasificación del dominio a partir de los modelos del escenario 3.

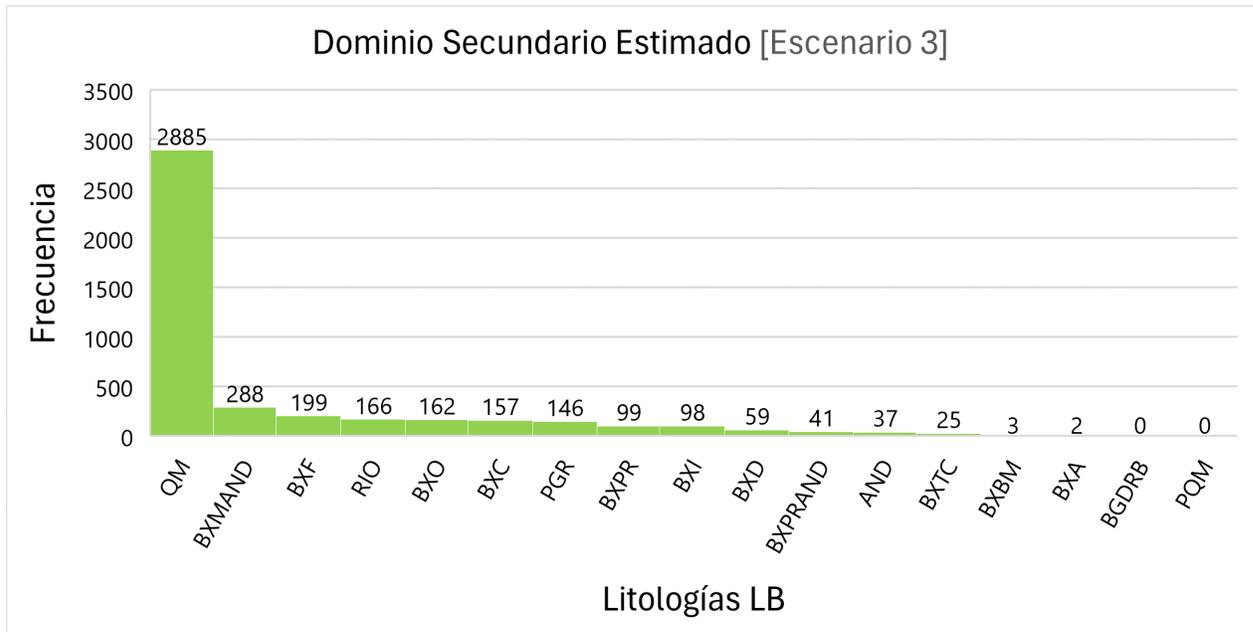


Figura D.0.4: Distribución de las litologías presentes en el dominio secundario, basado en la clasificación del dominio a partir de los modelos del escenario 3.