



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**Rediseñar el proceso de Estudios y Calidad de Servicio, integrando un
modelo de predicción para las desconexiones de curso forzoso en los
Transformadores de Poder**

*PROYECTO DE GRADO PARA OPTAR AL GRADO DE MAGÍSTER EN INGENIERÍA
DE NEGOCIOS CON TECNOLOGÍAS DE INFORMACIÓN*

GIOVANNI ITALO MARCELLO NOCETI GUZMÁN

PROFESOR GUÍA:
EZEQUIEL MUÑOZ KRSULOVIC

MIEMBROS DE LA COMISIÓN:
CINTHYA VERGARA SILVA
EDUARDO CONTRERAS VILLABLANCA

SANTIAGO DE CHILE
2023

RESUMEN EJECUTIVO

La industria eléctrica en Chile se divide en 3 grandes líneas de negocios. Primero la Generación eléctrica, en donde se encuentran instalaciones hidroeléctricas, geotérmicas, solares, eólicas y termoeléctricas. La segunda línea es la Transmisión eléctrica (Tx), que cuenta con Subestaciones de poder (S/E) y torres de alta tensión, para transportar y transformar la energía en grandes distancias. Finalmente, tenemos la Distribución eléctrica, conformada de alimentadores de cabecera para entregar energía a los hogares, sectores públicos y privados.

Este trabajo se centra en el Sistema de Transmisión del Sur S.A (STS), que pertenece a la segunda línea de negocios. Desde sus inicios en el 2001, hasta el 2017, se encontraba geográficamente en el sur de Chile, desde Concepción hasta Chiloé. Pero desde el año 2018, gracias a nuevas políticas públicas para fomentar la competitividad, la compañía se adjudicó una serie de licitaciones en el norte y centro del país, las cuales permitieron a STS, crecer y aumentar su cobertura en kilómetros de líneas de transmisión eléctrica, consiguiendo además nuevos clientes para operar en tiempo real las instalaciones eléctricas y prestar servicios de mantenimiento en las instalaciones eléctricas de diferentes industrias, como por ejemplo la industria minera.

Precisamente por lo anterior, la empresa tiene hoy en día el desafío estratégico de redefinir y satisfacer las necesidades de sus clientes; entregar calidad de servicio de excelencia; disminuir el índice que mide el tiempo de desconexión eléctrico (SAIDI); hacer más eficiente su operación bajando sus Costos Operacionales y de Mantenimiento (COMA); apuntar a una transformación digital para modernizar la industria eléctrica; asegurar la cultura y los talentos de sus profesionales. Todo lo anterior con un intransable, que es el consolidar el liderazgo en seguridad laboral de sus trabajadores.

Enmarcado en sus desafíos estratégicos, el presente proyecto de tesis se enfoca en rediseñar el proceso de estudios y calidad de servicio, integrando un modelo de predicción para las desconexiones de curso forzoso en los transformadores de poder, que permite disminuir los COMA con mantenimientos predictivos y obtener continuidad operacional disminuyendo los índices SAIDI. Esta tesis utiliza dos metodologías. Por un lado, la de Ingeniería de Negocios para los procesos. Por otro lado, como apoyo tecnológico, se utiliza el ciclo estandarizado CRISP-DM para proyectos de Machine Learning.

TABLA DE CONTENIDO

| | |
|--|-----------|
| CAPITULO 1 INTRODUCCIÓN Y CONTEXTO..... | 1 |
| 1.1 Antecedentes de la Industria..... | 1 |
| 1.2 Descripción General de la Empresa | 3 |
| 1.2.1 Visión..... | 3 |
| 1.3 Acerca del problema y su justificación | 4 |
| 1.4 Objetivos y Resultados Esperados del Proyecto | 5 |
| 1.4.1 Objetivo General..... | 5 |
| 1.4.2 Objetivos Específicos | 5 |
| 1.4.3 Resultados esperados | 6 |
| 1.5 Alcance | 6 |
| 1.5.1 Exclusiones | 7 |
| 1.6 Riesgos Potenciales | 7 |
| CAPÍTULO 2. MARCO TEÓRICO | 11 |
| 2.1 Metodología de Ingeniería de Negocios..... | 11 |
| 2.2 Marco Teórico CRISP-DM | 14 |
| 2.3 Algoritmos de clasificación y aprendizaje supervisado | 16 |
| 2.3.1 K Nearest Neighbor (KNN)..... | 17 |
| 2.3.2 Naive Bayes (NB)..... | 18 |
| 2.3.3 Random Forest (RF) | 20 |
| CAPÍTULO 3. PLANTEAMIENTO ESTRATÉGICO Y ANÁLISIS DE LA SITUACIÓN ACTUAL | 22 |
| 3.1 Posicionamiento Estratégico | 22 |
| 3.2 Modelo de Negocios..... | 23 |
| 3.2.1 Segmentos de Clientes | 24 |
| 3.2.2 Propuesta de valor..... | 24 |
| 3.2.3 Canales..... | 24 |

| | |
|--|-----------|
| 3.2.4 Relaciones con el cliente..... | 24 |
| 3.2.5 Fuentes de ingreso | 25 |
| 3.2.6 Recursos clave | 25 |
| 3.2.7 Actividades clave..... | 25 |
| 3.2.8 Socios claves..... | 25 |
| 3.2.9 Estructura de Costos | 26 |
| 3.3 Diagnóstico de la Situación Actual | 26 |
| 3.3.1 Problema(s) Identificado(s) / Oportunidad(es) identificada(s)..... | 26 |
| 3.3.2 Arquitectura de Procesos AS IS (Patrones)..... | 39 |
| 3.3.3 Modelamiento Detallado de Procesos AS IS (BPMN)..... | 44 |
| 3.4 Cuantificación del Problema u Oportunidad | 46 |
| CAPÍTULO 4. PROPUESTA DE DISEÑO DE PROCESOS | 47 |
| 4.1 Direcciones de cambio y alcance | 47 |
| 4.1.1 Estructura de la Empresa y Mercado | 48 |
| 4.1.2 Anticipación..... | 48 |
| 4.1.3 Coordinación..... | 49 |
| 4.1.4 Prácticas de Trabajo..... | 49 |
| 4.1.5 Integración de Procesos Conexos. | 50 |
| 4.1.6 Mantenimiento Consolidado de Estados. | 51 |
| 4.2 Propuesta de Solución | 52 |
| 4.2.1 Arquitectura de Procesos To Be (Patrones)..... | 52 |
| 4.2.2 Modelamiento Detallado de Procesos To Be (BPMN)..... | 54 |
| 4.2.3 Diseño de Lógica de Negocios | 55 |
| 4.2.3.1 Entendimiento de los datos | 55 |
| 4.2.3.2 Preparación de los datos..... | 58 |
| 4.3 Resultados Obtenidos | 61 |
| 4.3.1 Modelamiento de los datos | 61 |
| 4.3.1.1 Resultados KNN, NB, RF | 61 |

| | |
|--|-----------|
| 4.4 Retroalimentación juicio experto | 65 |
| CAPÍTULO 5. PROPUESTA DE APOYO TECNOLÓGICO..... | 68 |
| 5.1 Arquitectura Tecnológica | 68 |
| 5.2 Visualización de los datos | 69 |
| CAPÍTULO 6. PROPUESTA DE IMPLEMENTACIÓN | 74 |
| 6.1 Con texto de la Empresa..... | 74 |
| 6.2 Análisis de los Principios de Diseño | 75 |
| 6.2.1 Sentido y estrategia del proceso de cambio | 75 |
| 6.2.2 Cambio y conservación..... | 76 |
| 6.2.3 Gestión de las comunicaciones | 76 |
| 6.2.4 Gestión del aprendizaje y las habilidades | 76 |
| CAPÍTULO 7. EVALUACIÓN DEL PROYECTO | 78 |
| 7.1 Producto Mínimo Viable (MVP)..... | 78 |
| 7.1.1 Definición del MVP | 78 |
| 7.1.2 Resultados Obtenidos | 79 |
| 7.2 Definición de Variables y Criterios de Evaluación | 80 |
| 7.2.1 Tipo de Evaluación | 80 |
| 7.2.2 Horizonte de Evaluación..... | 81 |
| 7.2.3 Tasa de Descuento | 81 |
| 7.2.4 Escenarios de Evaluación | 81 |
| 7.2.5 Costos..... | 82 |
| 7.2.7 Inversión | 83 |
| 7.2.8 Consideración Social de Beneficios | 83 |
| 7.3 Flujo de Caja | 83 |
| 7.4 Análisis de Sensibilidad | 86 |
| CAPÍTULO 8. CONCLUSIONES, LECCIONES APRENDIDAS, TRABAJOS FUTUROS..... | 88 |
| 8.1 Lecciones Aprendidas | 89 |

| | |
|---|-----------|
| 8.2 Trabajos Futuros..... | 90 |
| 9. BIBLIOGRAFÍA..... | 91 |
| 10. ANEXOS | 93 |
| Anexo A: Extracto Metodología vBPM..... | 93 |
| Triangulación para mejorar procesos existentes..... | 93 |
| Anexo B: Selección del Problema..... | 96 |
| Causa Raíz | 96 |
| Anexo C: Algoritmos de Entrenamiento Supervisado con Python | 100 |
| PRIMEROS RESULTADOS | 100 |
| PRIMERA ITERACIÓN..... | 103 |
| SEGUNDA ITERACIÓN | 107 |
| TERCERA ITERACIÓN | 110 |
| CUARTA ITERACIÓN | 113 |
| QUINTA ITERACIÓN | 116 |

INDICE DE FIGURAS

| | |
|--|----|
| <i>Figura 1: Sistemas eléctricos en Chile. SEN, Aysén y Magallanes</i> | 1 |
| <i>Figura 2: Torres y líneas de Transmisión eléctrica y sus niveles de tensión (voltaje)</i> | 2 |
| <i>Figura 3: Ejemplo de una Subestación de Transmisión eléctrica y sus componentes</i> | 2 |
| <i>Figura 4: Metodología Ingeniería de Negocios</i> | 11 |
| <i>Figura 5: Macroprocesos de Barros</i> | 13 |
| <i>Figura 6: Metodología CRISP-DM</i> | 14 |
| <i>Figura 7: Evolución metodologías DataScience</i> | 16 |
| <i>Figura 8: Descripción general de K-vecinos más cercanos (KNN)</i> | 17 |
| <i>Figura 9: Diagrama ejemplificando la Probabilidades de predecir de Naive Bayes</i> | 19 |
| <i>Figura 10: Técnica Random Forest</i> | 21 |
| <i>Figura 11: Posicionamiento estratégico de STS de acuerdo con el modelo Delta</i> | 22 |
| <i>Figura 12: CANVAS de STS</i> | 23 |
| <i>Figura 13: Proyecciones del aporte al EBITDA</i> | 27 |
| <i>Figura 14: Proyección SAIDI</i> | 27 |
| <i>Figura 15: Objetivos estratégicos</i> | 28 |
| <i>Figura 16: Primeros Value Drivers</i> | 29 |
| <i>Figura 17: Segundos Value Drivers</i> | 30 |
| <i>Figura 18: Árbol de Problemas</i> | 31 |
| <i>Figura 19: Framework APQC con enfoque en al Utilities</i> | 32 |
| <i>Figura 20: Primeros Procesos con nota y ponderación</i> | 33 |
| <i>Figura 21: Primeros Procesos seleccionados con nota y ponderación</i> | 33 |
| <i>Figura 22: Matriz de estandarización</i> | 34 |

| | |
|--|----|
| <i>Figura 23: Gráfico matriz de estandarización</i> | 35 |
| <i>Figura 24: Triangulación para evaluar primer proceso seleccionado</i> | 36 |
| <i>Figura 25: Tabla Matrix Impacto / Esfuerzo</i> | 36 |
| <i>Figura 26: Gráfico Matrix Impacto / Esfuerzo</i> | 37 |
| <i>Figura 27: Cambio de transformador en falla en S/E Melipulli de la empresa STS</i> | 38 |
| <i>Figura 28: Traslado de nuevo transformador hacia S/E Melipulli de la empresa STS</i> .. | 38 |
| <i>Figura 29: Análisis de Macroprocesos de STS con el Coordinador Eléctrico Nacional</i> | 40 |
| <i>Figura 30: Macro 1 STS</i> | 41 |
| <i>Figura 31: Macro 1 Transmisión de Energía Eléctrica de STS</i> | 42 |
| <i>Figura 32: Macro 1 Administración con el cliente y autoridad fiscalizadora CEN</i> | 43 |
| <i>Figura 33: Macro 1 Estudio y Calidad de Servicio</i> | 44 |
| <i>Figura 34: BPMN Área de Estudios y Calidad de Servicio</i> | 45 |
| <i>Figura 35: Proceso rediseñado integrando un modelo de predicción de mantenimientos programados y de curso forzoso</i> | 52 |
| <i>Figura 36: Desarrollar y Actualizar Modelos Predictivos, Rediseño</i> | 53 |
| <i>Figura 37: Proceso rediseñado integrando un modelo de predicción de mantenimientos programados y de curso forzoso</i> | 54 |
| <i>Figura 38: Variables del set de datos (21.009 filas y 37 columnas (variables))</i> | 55 |
| <i>Figura 39: Identificación del Target</i> | 56 |
| <i>Figura 40: Grafico de variables distribuidas por temporada</i> | 56 |
| <i>Figura 41: Grafico de variables distribuidas por temporada</i> | 57 |
| <i>Figura 42: Grafico de variables distribuidas por meses</i> | 57 |
| <i>Figura 43: Variables seleccionadas para el entrenamiento</i> | 58 |

| | |
|---|----|
| <i>Figura 44: Muestra gráficamente que se limpian y no hay espacios vacíos en los datos</i> | 59 |
| <i>Figura 45: Variables seleccionadas y transformadas a numéricas para entrenar los modelos de predicción</i> | 60 |
| <i>Figura 46: Matriz de confusión con los resultados de los algoritmos en la tercera iteración</i> | 63 |
| <i>Figura 47: Curva de ROC de los algoritmos en la tercera iteración</i> | 65 |
| <i>Figura 48: Arquitectura de Apoyo Tecnológico</i> | 69 |
| <i>Figura 49: Visualización grafica del Jupyter Notebook</i> | 70 |
| <i>Figura 50: Gráfico de barras</i> | 71 |
| <i>Figura 51: Gráfico de histograma</i> | 71 |
| <i>Figura 52: Gráfico de caja</i> | 72 |
| <i>Figura 53: Gráfico de dispersión</i> | 72 |
| <i>Figura 54: Matriz de correlación</i> | 73 |
| <i>Figura 55: Triangulación para evaluar primer proceso seleccionado</i> | 93 |
| <i>Figura 56: Triangulación para evaluar segundo proceso seleccionado</i> | 94 |
| <i>Figura 57: Triangulación para evaluar tercer proceso seleccionado</i> | 94 |
| <i>Figura 58: Triangulación para evaluar cuarto proceso seleccionado</i> | 95 |
| <i>Figura 59: Triangulación para evaluar quinto proceso seleccionado</i> | 95 |
| <i>Figura 60: Design Thinking</i> | 96 |
| <i>Figura 61: Causa Raíz de Ishikawa</i> | 97 |
| <i>Figura 62: Causa Raíz de los 5 ¿por qué?</i> | 98 |
| <i>Figura 63: Brainstorming</i> | 98 |
| <i>Figura 64: Matriz Impacto esfuerzo con potenciales soluciones</i> | 99 |

| | |
|---|------------|
| <i>Figura 65: Propuesta de proyecto piloto.....</i> | <i>99</i> |
| <i>Figura 66: Primeros resultados.....</i> | <i>100</i> |
| <i>Figura 67: Primeros resultados.....</i> | <i>103</i> |
| <i>Figura 68: Curva de ROC de los Primeros resultado</i> | <i>103</i> |
| <i>Figura 69: Grid Search de la primera iteración</i> | <i>104</i> |
| <i>Figura 70: Resultados de los algoritmos en la primera iteración.....</i> | <i>105</i> |
| <i>Figura 71: Matriz de confusión con los resultados de los algoritmos en la primera iteración</i> | <i>106</i> |
| <i>Figura 72: Curva de ROC de los algoritmos en la primera iteración.....</i> | <i>107</i> |
| <i>Figura 73: Resultados de los ajustes de hiperparametros con Random Search</i> | <i>107</i> |
| <i>Figura 74: Resultados de los algoritmos en la segunda iteración</i> | <i>108</i> |
| <i>Figura 75: Matriz de confusión con los resultados de los algoritmos en la segunda iteración</i> | <i>109</i> |
| <i>Figura 76: Curva de ROC de los algoritmos en la segunda iteración</i> | <i>110</i> |
| <i>Figura 77: Nueva selección de atributos en una tercera iteración</i> | <i>110</i> |
| <i>Figura 78: Resultados de los algoritmos en la tercera iteración</i> | <i>111</i> |
| <i>Figura 79: Matriz de confusión con los resultados de los algoritmos en la tercera iteración</i> | <i>112</i> |
| <i>Figura 80: Curva de ROC de los algoritmos en la tercera iteración.....</i> | <i>113</i> |
| <i>Figura 81: Resultados de los ajustes de hiperparametros con Grid Search en la cuarta iteración</i> | <i>114</i> |
| <i>Figura 82: Resultados de los algoritmos en la cuarta iteración</i> | <i>114</i> |
| <i>Figura 83: Matriz de confusión con los resultados de los algoritmos en la cuarta iteración</i> | <i>115</i> |
| <i>Figura 84: Curva de ROC de los algoritmos en la cuarta iteración.....</i> | <i>116</i> |

Figura 85: Resultados de los ajustes de hiperparametros con Random Search en la quinta iteración 117

Figura 86: Resultados de los algoritmos en la quinta iteración 117

Figura 87: Matriz de confusión con los resultados de los algoritmos en la quinta iteración 118

Figura 88: Curva de ROC de los algoritmos en la quinta iteración 119

INDICE DE TABLAS

| | |
|---|----|
| <i>Tabla 1: Probabilidad de ocurrencia de los riesgos.</i> | 8 |
| <i>Tabla 2: Impacto de los riesgos</i> | 8 |
| <i>Tabla 3: Matriz de Riesgos.</i> | 9 |
| <i>Tabla 4: Estructura Empresa y Mercado.</i> | 48 |
| <i>Tabla 5: Anticipación.</i> | 48 |
| <i>Tabla 6: Coordinación.</i> | 49 |
| <i>Tabla 7: Prácticas de Trabajo.</i> | 50 |
| <i>Tabla 8: Integración de Procesos Conexos.</i> | 51 |
| <i>Tabla 9: Mantenimiento Consolidado del Estado.</i> | 51 |
| <i>Tabla 10: Entradas y salidas obtenidas por los algoritmos</i> | 62 |
| <i>Tabla 11: Matriz de confusión 2x2. TP, TN, FP y FN</i> | 63 |
| <i>Tabla 12: Comparación matriz de confusión entre la segunda, tercera, cuarta y quinta iteración</i> | 64 |
| <i>Tabla 13: Retroalimentación Juicio Experto.</i> | 66 |
| <i>Tabla 14: Costos de mantenimientos TTFB y falla ocurrida el 2020</i> | 79 |
| <i>Tabla 15: Costos de STS</i> | 80 |
| <i>Tabla 16: Costos de mantenimientos en los TTFB</i> | 82 |
| <i>Tabla 17: Flujo de caja, escenario sin Proyecto</i> | 84 |
| <i>Tabla 18: Indicadores, escenario sin Proyecto</i> | 84 |
| <i>Tabla 19: Flujo de caja, escenario con Proyecto</i> | 85 |
| <i>Tabla 20: Indicadores, escenario con Proyecto</i> | 86 |
| <i>Tabla 21: Índice SAIDI por año</i> | 86 |

| | |
|---|------------|
| <i>Tabla 22: Escenario ficticio donde se consideran, el porcentaje de desconexiones programadas vs el porcentaje desconexiones de curso forzoso.</i> | <i>87</i> |
| <i>Tabla 23: Entradas y salidas obtenidas por los algoritmos</i> | <i>101</i> |
| <i>Tabla 24: Primeros resultados del Accuracy.....</i> | <i>101</i> |
| <i>Tabla 25: Primeros resultados de los algoritmos KNN, NB y RF, en la clasificación de desconexiones programadas</i> | <i>101</i> |
| <i>Tabla 26: Primeros resultados de los algoritmos KNN, NB y RF en la clasificación de desconexiones de curso forzoso</i> | <i>102</i> |
| <i>Tabla 27: Matriz de confusión 2x2. TP, TN, FP y FN.....</i> | <i>102</i> |
| <i>Tabla 28: Accuracy de la primera iteración.....</i> | <i>105</i> |
| <i>Tabla 29: Precision y Recall de la primera iteración para desconexión programada .</i> | <i>105</i> |
| <i>Tabla 30: Precision y Recall de la primera iteración para curso forzoso</i> | <i>106</i> |
| <i>Tabla 31: Accuracy segunda iteración</i> | <i>108</i> |
| <i>Tabla 32: Precision y Recall segunda iteración para desconexiones programadas.....</i> | <i>109</i> |
| <i>Tabla 33: Precision y Recall segunda iteración para desconexiones de curso forzoso</i> | <i>109</i> |
| <i>Tabla 34: Accuracy tercera iteración</i> | <i>111</i> |
| <i>Tabla 35: Precision y recall en la tercera iteración para desconexión programada ...</i> | <i>111</i> |
| <i>Tabla 36: Precision y recall en la tercera iteración para desconexión de curso forzoso</i> | <i>112</i> |
| <i>Tabla 37: Comparación matriz de confusión entre la segunda y tercera iteración</i> | <i>113</i> |
| <i>Tabla 38: Precision y recall en la cuarta iteración para desconexión programada.....</i> | <i>115</i> |
| <i>Tabla 39: Precision y recall en la cuarta iteración para desconexión de curso forzoso</i> | <i>115</i> |
| <i>Tabla 40: Comparación matriz de confusión entre la segunda, tercera y cuarta iteración</i> | <i>116</i> |
| <i>Tabla 41: Precision y recall en la quinta iteración para desconexión programada.....</i> | <i>118</i> |

Tabla 42: Precision y recall en la quinta iteración para desconexión de curso forzoso
..... 118

Tabla 43: Comparación matriz de confusión entre la segunda, tercera, cuarta y quinta iteración
..... 119

CAPÍTULO 1. INTRODUCCIÓN Y CONTEXTO

1.1 Antecedentes de la Industria

El Sistema Eléctrico Nacional (SEN) de Chile incluye las instalaciones de generación eléctrica, transmisión y consumo que abarcan el territorio desde las regiones de Arica - Parinacota (Norte) hasta la Décima Región (Isla Grande de Chiloé, Sur). Este sistema es el más grande de los tres existentes en Chile (35.501 km de líneas de transmisión en 3.100 kilómetros de territorio) y proporciona energía eléctrica al territorio chileno, con una capacidad instalada a través de 25.284 MW y una cobertura de suministro que alcanza alrededor del 97% de la población (Araneda, 2021).

En la Figura 1 se pueden identificar los tres sistemas eléctricos en Chile. El SEN marcado en verde, además de los Sistemas Eléctricos aislados de Aysén y Magallanes, destacados en color naranja y azul respectivamente.

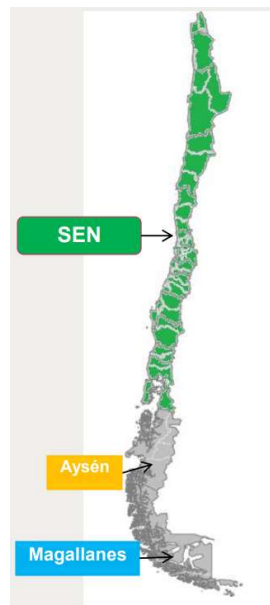


Figura 1: Sistemas eléctricos en Chile. SEN, Aysén y Magallanes

Fuente: https://www.cigre.cl/wp-content/uploads/2021/07/Sistema-Electrico_CHILE__e_RIAC_2021.pdf

Dentro del SEN, se encuentra el Sistema de Transmisión de voltaje de Alta Tensión de Chile, al cual pertenece STS, que cuenta con diferentes tipos de torres con niveles de tensión o voltaje para transportar la energía y finalmente para transformarla bajando su voltaje en las S/E, como se puede observar en la Figura 2:

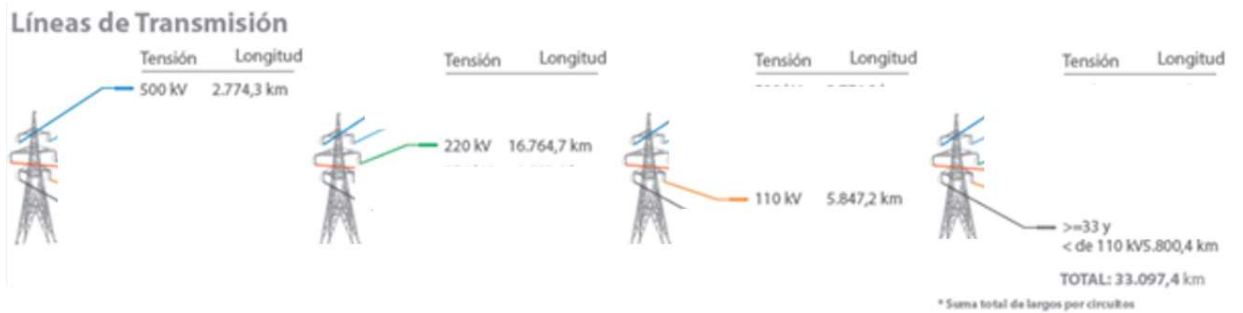


Figura 2: Torres y líneas de Transmisión eléctrica y sus niveles de tensión (voltaje)

Fuente: Elaboración propia

En la Figura 3 se muestra el ejemplo de una S/E, indicando sus componentes.

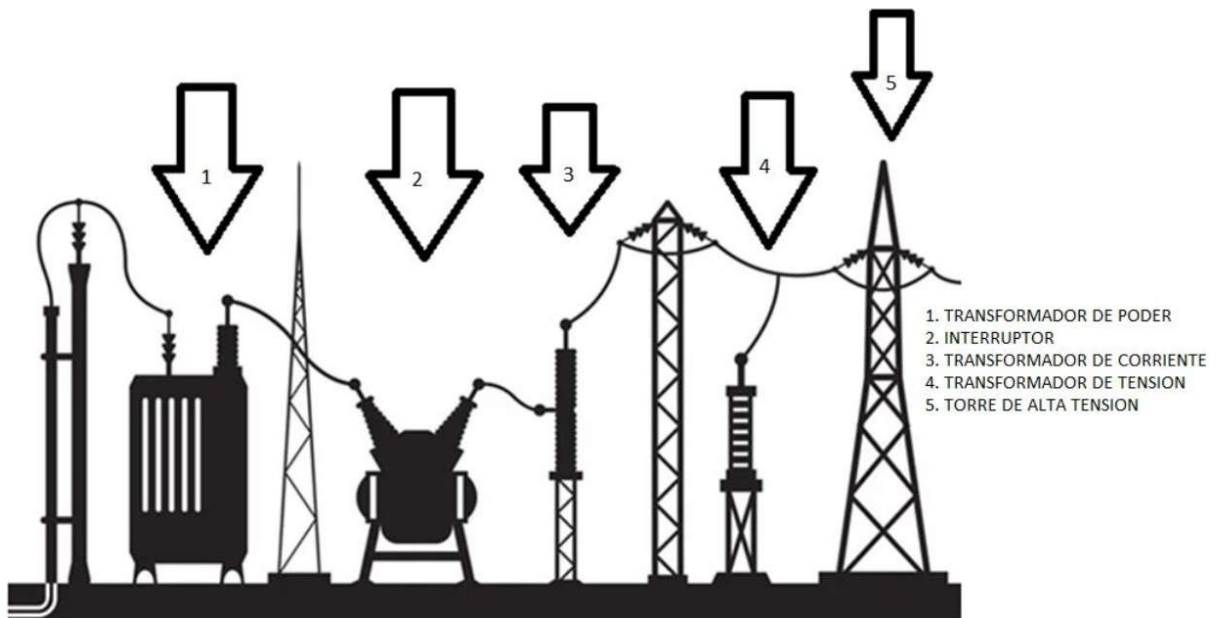


Figura 3: Ejemplo de una Subestación de Transmisión eléctrica y sus componentes

Fuente: Elaboración propia

Cada uno de estos activos son necesarios en una S/E de Transmisión eléctrica para transformar la energía que llega desde las Centrales de generación, a un voltaje más bajo, el cual puede ser utilizado por las empresas Distribuidoras eléctricas en las diferentes industrias y hogares.

1.2 Descripción General de la Empresa

STS pertenece al Grupo SAESA, es un conglomerado de empresas eléctricas constituido el año 2001. Cuenta con 236 colaboradores, 1.753 contratistas, y aporta al EBITDA del Grupo SAESA un monto de \$52.581 millones de pesos.

STS es una empresa con 2.139 KM de redes de transmisión eléctrica y sus actividades claves son el transporte y transformación de energía eléctrica. Presta servicios de asesorías en el diseño, construcción y mantenimiento de líneas de Tx y S/E. Además, gracias a su avanzada tecnología, realiza la operación en tiempo real de los sistemas eléctricos propios y de sus clientes. Sus principales ingresos, se deben a contratos de los servicios ofrecidos y el cobro de peajes a compañías generadoras que tienen contratos de suministros con distribuidoras.

Sus competidores son otras empresas de transmisión eléctrica como TRANSELEC, TRANSNET, TEN, CELEOREDES, entre otras repartidas a lo largo del país con tramos más cortos y una suma menor de KM.

1.2.1 Visión

Según indica el portal web oficial de STS, la visión de la empresa es: “Mejorar la calidad de vida de las personas y contribuir al desarrollo sustentable del país, entregando energía confiable y segura. Nuestro trabajo se fundamenta en un profundo compromiso con nuestros clientes, el cuidado y desarrollo de nuestros trabajadores y la modernización de la industria eléctrica en Chile. Nuestra visión tiene una perspectiva de largo plazo, que

busca crear valor para nuestros accionistas” (<https://web.gruposaes.cl/quienes-somos/mision-vision-y-valores>).

1.3 Acerca del problema y su justificación

Para crecer como compañía, STS debe adjudicarse licitaciones para aumentar los kilómetros de líneas de Tx eléctrica, sumando la construcción de nuevas S/E. También debe cumplir con la norma técnica exigida por Coordinador Eléctrico Nacional (autoridad fiscalizadora de Chile), que indica que la indisponibilidad del servicio eléctrico en el índice SAIDI, no debe superar los 3 minutos por cada falla.

Este proyecto resalta la importancia de abordar los problemas anteriormente descritos y se rescata de las entrevistas realizadas a los *stakeholders* de la empresa, que STS tiene un interés en encontrar soluciones innovadoras y modernas. En los últimos años la compañía ha realizado inversiones en una serie de proyectos que involucran nuevas tecnologías de reportería, automatización, supervisión y control en tiempo real. Esto ha dado como resultado el contar con unas robustas, pero complejas, arquitecturas de Tecnologías de la Información (TI) y Tecnologías Operacionales (TO). Los recientes desarrollos que existen en la empresa han impulsado la necesidad de contar con nuevas capacidades para trabajar con una cantidad considerable de variables difíciles de procesar y de diversas fuentes de datos estructurados y no estructurados. Por esta razón, desde el año 2021 han nacido un conjunto de iniciativas de pilotos y tesis Universitarias relacionadas con Data Analytics, como son el pronosticar demanda, analizar patrones de curva para trasposos de carga y machine learning.

Para el desafío mencionado, este proyecto de tesis disminuye los COMA, evitando los mantenimientos correctivos o de curso forzoso, cambiándolos por predictivos. Por otro lado, y para evitar las multas, se consigue aportar en la reducción de minutos en la indisponibilidad anual del índice SAIDI de los transformadores de poder de 1,08 horas del año 2021 a 0,3 horas al año 2025, anticipándose a sus fallas.

El presente proyecto de grado tiene el propósito de utilizar todas las tecnologías y datos disponibles para rediseñar el proceso de estudios y calidad de servicio, integrando un modelo de predicción para las desconexiones de curso forzoso en los Transformadores de

Poder. El objetivo es disminuir los índices COMA y SAIDI, debido a que los Transformadores son el activo más caro y complejo de reparar en una S/E de Tx. Se entregan algoritmos de machine learning como producto mínimo viable (MVP), para predecir las desconexiones de curso forzoso y pasarlas a desconexiones programadas, y así anticiparse a posibles fallas.

Este proyecto es supervisado y validado por el área de Estudio y Calidad de Servicio de Transmisión de la compañía.

1.4 Objetivos y Resultados Esperados del Proyecto

1.4.1 Objetivo General

Para disminuir los COMA, y bajar los índices de disponibilidad SAIDI, se rediseña un proceso, utilizando una metodología de Data Analytics, que logra anticiparse a los mantenimientos de curso forzoso de los transformadores de poder. El Objetivo General del proyecto se define como:

“Rediseñar el proceso de estudios y calidad de servicio, integrando un modelo de predicción para las desconexiones de curso forzoso en los Transformadores de Poder”

1.4.2 Objetivos Específicos

Los objetivos específicos del proyecto que son necesarios para lograr el objetivo General son los siguientes:

- Proponer un rediseño a los procesos existentes, o diseñar un proceso faltante en los estudios de calidad de servicio, utilizando modelos predictivos.

- Levantar requerimientos, definir roles y analizar los procesos involucrados para solucionar problemas con Data Analytics.
- Proponer un modelo predictivo con la metodología CRISP-DM para identificar las variables de entrada que necesita el algoritmo de predicción.
- Programar algoritmos de Machine Learning para predecir las desconexiones de curso forzoso.

1.4.3 Resultados esperados

- Entregar un rediseño al proceso de estudios y calidad de servicio, que utilice Data Analytics.
- Etapas, roles y funciones definidas para el rediseño del proceso, que permitan desarrollar nuevas capacidades de Data Analytics para problemas de predicción.
- Entregar las variables de entrada necesarias para correr los algoritmos de entrenamiento supervisado..
- Entregar un Producto Mínimo Viable (MVP) y con las predicciones de los algoritmos, obtener una disminución de los COMA que aporten a alcanzar un EBITDA de un 41,7% al 2021 a un 42,8% al 2023 y además aportar con disminuir el índice SAIDI de 1,08 hrs. obtenidos el año 2021 a un 0,42 hrs. para el año 2023.

1.5 Alcance

Este proyecto se desarrolla en la Subgerencia de Operación y Mantenimiento de Transmisión Eléctrica de la compañía, perteneciente a la Gerencia de Operaciones. Es supervisado por el área de Estudio y Calidad de Servicio, la cual actualmente no utiliza técnicas de Data Analytics, pero sí realiza sus estudios apoyados en diferentes *softwares* especializados, crea informes y calcula flujos potencia. Como parte del desarrollo del proyecto se rediseña el proceso de estudios y calidad de servicio y se entrega un MVP con modelos predictivos con técnicas de Data Analytics. Además, se contará con la colaboración de los ingenieros de esta área, para entender el problema de negocio, obtener retroalimentación y validar los resultados obtenidos.

1.5.1 Exclusiones

Quedan fuera de esta tesis el rediseño de otras actividades de la cadena de valor que puedan influir en disminuir los COMA y la disminución de los índices SAIDI, como son los procesos de operación, supervisión y monitoreo en tiempo real y las planificaciones de mantenimientos preventivos y correctivos.

1.6 Riesgos Potenciales

La Gestión del Cambio según el Modelo CHESS (Conciencia, Holístico, Experiencial, Sistémico y Sustentable), el cual es un modelo Integral de Liderazgo y Gestión del Cambio (Olguín, Crawford, & Soto, 2016), se considera como riesgo, porque las personas por naturaleza tienden a resistirse a nuevas formas de hacer las cosas, por lo tanto, es un riesgo la Gestión del Cambio si es que no se considera para este proyecto un periodo de capacitación y adaptación, ya que el rediseñar un proceso integrando un modelo predictivo afectará la forma actual de pensar y trabajar. En resumen, existe el riesgo de que los modelos de machine learning entregados queden obsoletos, porque no existen los profesionales con competencias adecuadas para depurarlos.

En STS, como también en la Industria Eléctrica, las TI y las TO se encuentran separadas, con visiones y prioridades distintas, por lo tanto, se administran los recursos y consideran tiempos de respuesta ante contingencias, con un sentido de urgencia diferente. Esto es un riesgo porque se necesitan ambas tecnologías TI/TO para disponer de datos de calidad, para correr los modelos de predicción, y si no se trabaja en conjunto en ambas áreas tecnológicas, con acuerdos de servicios negociados anticipadamente, surgirán lagunas de información y, en consecuencia, los *data set* pueden ser de calidad insuficiente para aplicar técnicas de Data Analytics.

En este proyecto de tesis se construye una matriz de riesgos utilizando dos tablas de referencia. La Tabla 1 es utilizada por STS para clasificar la probabilidad de ocurrencia de los riesgos identificados en una tarea o en un proyecto.

Tabla 1: Probabilidad de ocurrencia de los riesgos.

| | CATEGORÍA | VALOR | DESCRIPCIÓN |
|---------------------------------------|-----------------------|-------|---|
| PROBABILIDAD DE OCURRENCIA (P) | Casi certeza | 5 | Riesgo cuya probabilidad de ocurrencia es muy alta, es decir, se tiene un alto grado de seguridad que éste se presente (90% a 100 %) |
| | Posible | 4 | Riesgo cuya probabilidad de ocurrencia es alta, es decir, se tiene entre 66% a 89% de seguridad que éste se presente. |
| | Moderado | 3 | Riesgo cuya probabilidad de ocurrencia es media, es decir, se tiene entre 31% a 65% de seguridad que éste se presente. |
| | Improbable | 2 | Riesgo cuya probabilidad de ocurrencia es baja, es decir, se tiene entre 11% a 30% de seguridad que éste se presente. |
| | Muy Improbable | 1 | Riesgo cuya probabilidad de ocurrencia es muy baja, es decir, se tiene entre 1% a 10% de seguridad que éste se presente. |

La siguiente Tabla 2 fue construida por STS, para medir el impacto de los riesgos.

Tabla 2: Impacto de los riesgos

| | CATEGORÍA | VALOR | DESCRIPCIÓN |
|--------------------|---------------------|-------|--|
| IMPACTO (I) | Catastrófico | 5 | Riesgo cuya materialización puede generar pérdidas financieras (\$) que tendrán un impacto CATASTRÓFICO en los resultados y/o comprometen TOTALMENTE la imagen pública de la empresa. Su materialización dañaría gravemente el desarrollo del proceso y el cumplimiento de los objetivos, impidiendo finalmente que éstos se logren. |
| | Mayores | 4 | Riesgo cuya materialización puede generar pérdidas financieras (\$) que tendrán un impacto IMPORTANTE en los resultados y/o comprometen FUERTEMENTE la imagen pública de la empresa. Su materialización dañaría SIGNIFICATIVAMENTE el desarrollo del proceso y el cumplimiento de los objetivos, impidiendo TOTAL O PARCIALMENTE que se desarrollen en forma normal. |

| | | | |
|--|------------------------|---|---|
| | Moderados | 3 | Riesgo cuya materialización puede generar pérdidas financieras (\$) que tendrán un impacto MODERADO en los resultados y/o comprometen MODERADAMENTE la imagen pública de la empresa. Su materialización causaría un DETERIORO en el desarrollo del proceso, dificultando o retrasando el cumplimiento de los objetivos, impidiendo PARCIALMENTE que se desarrollen en forma normal. |
| | Menores | 2 | Riesgo cuya materialización puede generar pérdidas financieras (\$) que tendrán un impacto MENOR en los resultados y/o comprometen de forma MENOR la imagen pública de la empresa. Su materialización causaría un BAJO DAÑO en el desarrollo del proceso y NO AFECTARIA el cumplimiento de los objetivos. |
| | Insignificantes | 1 | Riesgo cuya materialización NO GENERA pérdidas financieras (\$) ni comprometen de ninguna forma la imagen pública de la empresa. Su materialización puede tener un pequeño o nulo efecto en el desarrollo del proceso y no afectaría el cumplimiento de los objetivos. |

La Tabla 3 que se presenta a continuación, contiene los riesgos de este proyecto de grado. Identifica los riesgos, permite mitigarlos, llevar un control en el tiempo y, si estos ocurren, la empresa se encuentra preparada para cuantificar sus consecuencias.

Tabla 3: Matriz de Riesgos.

| Riesgo Crítico | Causa | Probabilidad (P) | | Impacto (I) | |
|--|--|------------------|-------|---------------|-------|
| | | Clasificación | Valor | Clasificación | Valor |
| No contar con el apoyo de los Ingenieros del área de Estudios y Calidad de Servicio. | Personas - Estructura organizacional | Improbable | 2 | Moderados | 3 |
| No contar con datos para alimentar el modelo. | Personas - Infraestructura Tecnológica | Posible | 4 | Moderados | 3 |

| | | | | | |
|--|--|---------|---|-----------|---|
| Tener datos de mala calidad o lagunas de información. | Personas - Infraestructura Tecnológica | Posible | 4 | Moderados | 3 |
| Que no usen el modelo por falta de capacitación y/o adaptación | Personas - Estructura organizacional | Posible | 4 | Moderados | 3 |

CAPÍTULO 2. MARCO TEÓRICO

2.1 Metodología de Ingeniería de Negocios

En esta tesis se utiliza la metodología de Ingeniería de Negocios, la cual tiene pasos secuenciales con un enfoque en la estrategia, su posicionamiento y objetivos estratégicos. Como se puede apreciar en la Figura 4, cada una de las etapas ayuda a identificar y definir el modelo de negocio, permite diseñar y rediseñar los procesos, además de implementar una herramienta de apoyo TI con gestión del cambio en la organización (Barros, 2015).

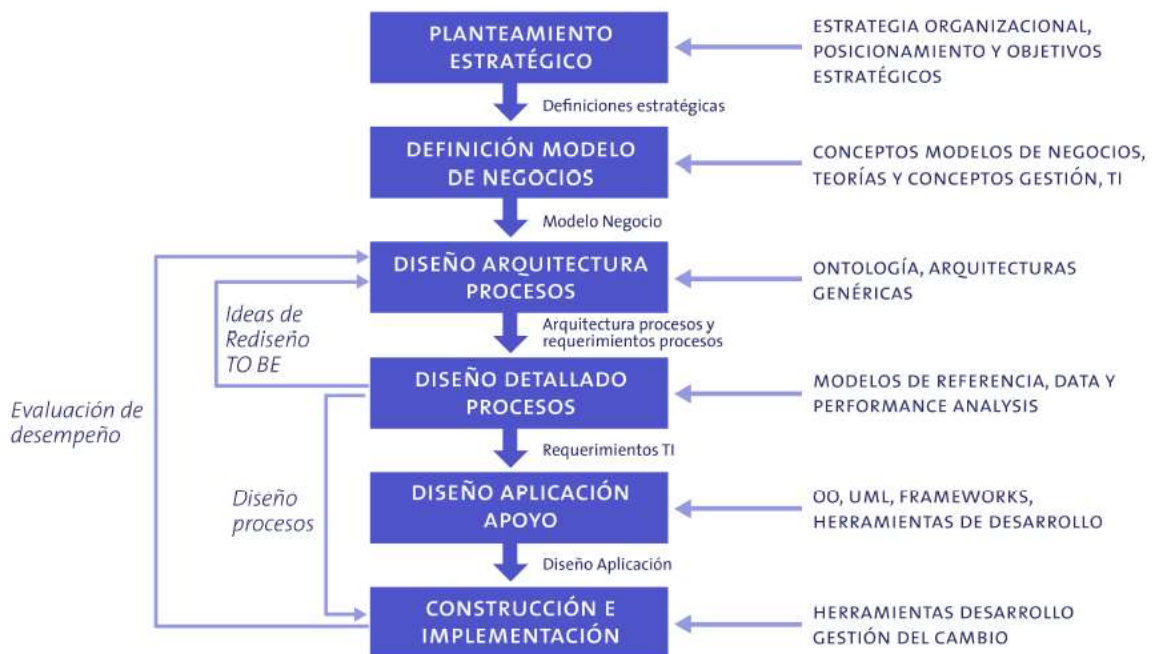


Figura 4: Metodología Ingeniería de Negocios

Fuente: Barros, 2015

En la figura anterior, se pueden observar las fases de la metodología del modelo. Cada una de estas se explica a continuación en forma general:

- i) Planteamiento estratégico: En esta etapa se deben identificar los ejes esenciales que sustentan el planteamiento estratégico y que impulsan u orientan al modelo de

negocio. Además, se identifican los *drivers* estratégicos para su crecimiento, competencia en el mercado, su relación con el cliente y sus índices de medición en el tiempo.

- ii) Definición del modelo de Negocio: Conceptos modelos de negocio, teorías y conceptos de gestión, TI. Se define en esta etapa el valor que se le entrega al cliente (Productos o servicios y problema que se resuelve), además la arquitectura que genera los ingresos de forma sostenida en la empresa.
- iii) Diseño y arquitectura de procesos: Ontología, Arquitecturas genéricas. Aquí se organizan y entienden todos los procesos de la empresa (As-Is), que generan el valor a los clientes.
- iv) Diseño detallado de procesos: Modelos de referencia, data y performance análisis. Con filosofía Lean se eliminan los desperdicios que no generan valor al cliente y se rediseñan los procesos de la empresa (To-Be).
- v) Diseño aplicación de apoyo: OO, UML, frameworks, herramientas de desarrollo. Utilizando las Tecnologías de la Información se proponen herramientas de apoyo para los nuevos procesos.
- vi) Construcción e implementación: Herramientas de desarrollo gestión del cambio. Se desarrollan productos mínimos viables y pilotos que pueden ser probados en la empresa, los cuales entregan resultados en un corto plazo y con bajo presupuesto.

Para modelar los procesos internos de la organización se utilizan los patrones de Macroprocesos (Barros, 2015). Como se muestra en la Figura 5, existen cuatro macroprocesos que ayudan a identificar los procesos a intervenir.

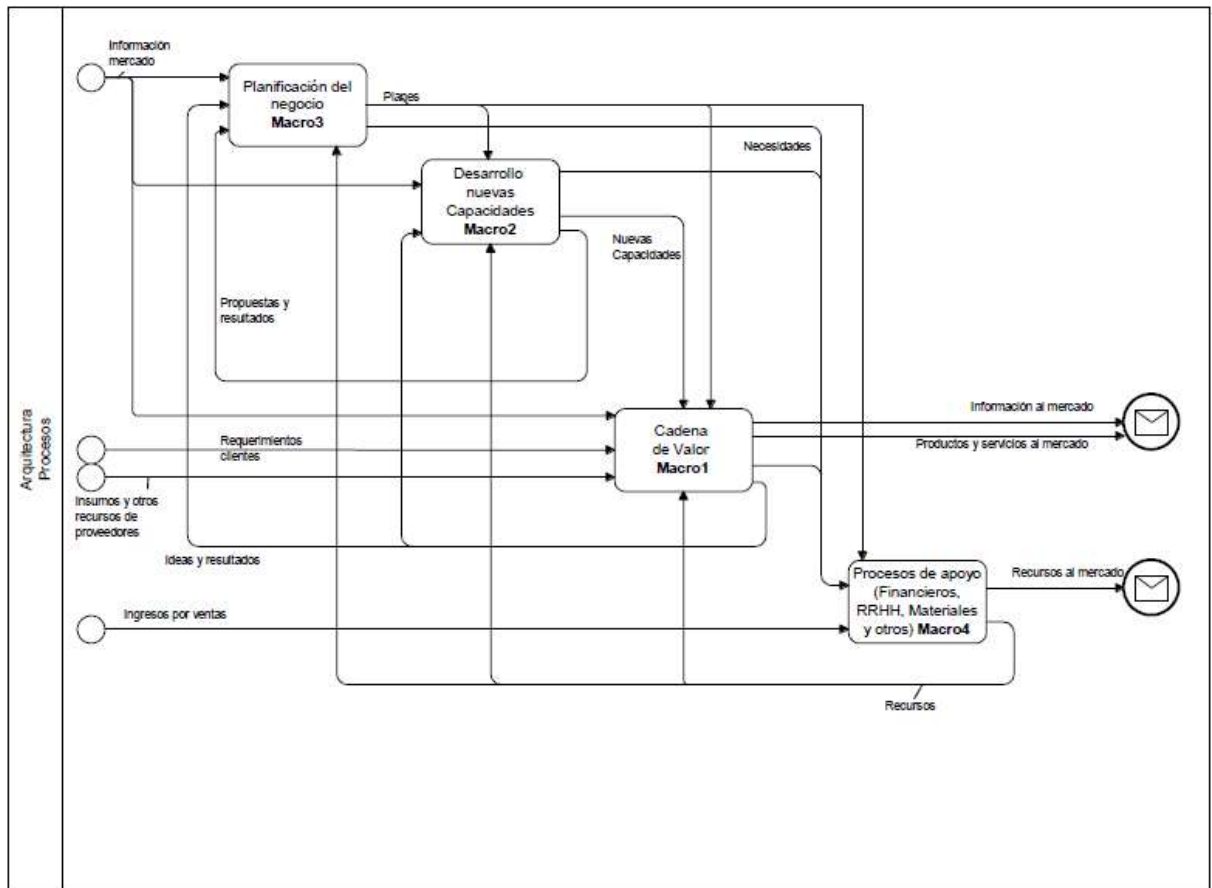


Figura 5: Macroprocesos de Barros

Fuente: Barros, 2015

Como se puede apreciar en la figura anterior, la **Macro1** es la Cadena de Valor de las organizaciones. Tiene como *input* externo la información del mercado, los requerimientos de los clientes, los insumos y otros recursos de los proveedores. Como *input* interno, está alineada con la **Macro3**, que es la Planificación estratégica del negocio. Además, **Macro1**, retroalimenta y recibe de la **Macro2** el desarrollo de nuevas capacidades.

También desde la **Macro1**, se consideran los productos y servicios al mercado. Tiene como *output* requerimientos que solicita a la **Macro4**, que es la habilitadora que presta apoyo con recursos al resto de las tres Macros y recibe los ingresos por ventas de los servicios prestados.

2.2 Marco Teórico CRISP-DM

En la Figura 6 se muestra el ciclo CRISP-DM, que es una metodología que presenta un proceso de tareas enfocadas a la aplicación de técnicas de minería de datos a empresas de diferentes industrias (Chapman et al., 2000). Esta metodología es un trabajo iterativo de mejora continua que en base a la experimentación se aprende y logran mejores resultados.

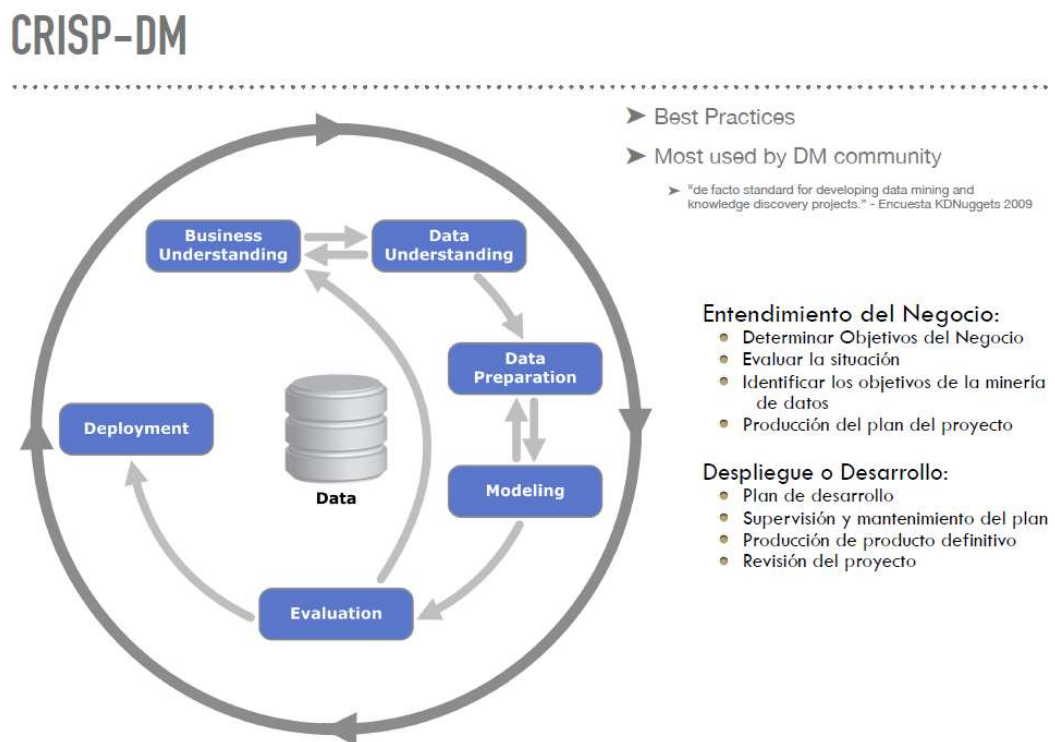


Figura 6: Metodología CRISP-DM

Fuente: CRISP-DM 1.0 <http://www.crisp-dm.org/download.htm>

Del modelo anteriormente expuesto, se pueden enumerar sus fases y cada una de estas se explican a continuación:

- i) Entender el Negocio: En esta etapa se determinan los objetivos del negocio, analizando los antecedentes, objetivos y criterios del éxito comercial de la empresa. Además, se debe evaluar su situación actual, los recursos disponibles,

requerimientos, suposiciones y restricciones. Se analizan los riesgos y contingencias, sus costos y beneficios. Aquí es donde se seleccionan los objetivos y criterio de éxito del trabajo de minería de datos, finalizando con la creación del plan de proyecto con la evaluación inicial de herramientas y técnicas a utilizar.

- ii) Entender los Datos: Se deben recopilar los datos iniciales. Se describen los datos con un informe de descripción, se exploran con un reporte y se verifica su calidad.
- iii) Preparar los Datos: Se crean conjuntos de datos con sus descripciones. Se seleccionan los datos justificando la inclusión y exclusión de estos. Se limpian datos con vacíos o errores y existe la posibilidad de construir nuevas variables derivadas de diferentes atributos.
- iv) Modelar: Se selecciona la técnica de modelado considerando suposiciones de modelado. Se generan diseños de prueba. Se configuran parámetros para el modelo en construcción. Finalmente, en esta fase se evalúa el modelo, revisando los parámetros.
- v) Evaluar: En este proceso se deben alinear los resultados obtenidos de la minería de datos, con el criterio de éxito del negocio. Se revisan los modelos aprobados y se determinan los próximos pasos con un listado de posibles acciones para la toma de decisiones.
- vi) Desplegar: En este último paso se crea un plan de implementación, seguimiento y mantenimiento. Se debe construir un informe y su presentación final, para complementar la documentación de la experiencia.

Existen también otras metodologías de Data Analytics que se pueden utilizar, como son KDD y SEMMA, entre otras, pero basándonos en las estadísticas que muestra la Figura 7, para este proyecto, se trabaja con la metodología CRISP-DM de minería de datos, que tiene más aprobación entre los profesionales con investigaciones empíricas.

EVOLUCIÓN DE METODOLOGÍAS

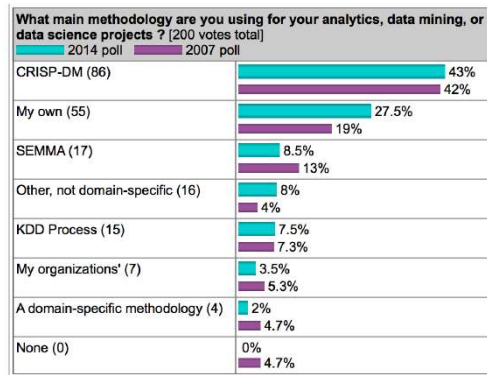


Figura 7: Evolución metodologías DataScience

Fuente: Ramo Introducción Data Science “CAP_1_Conceptos_Fundamentales.pdf”

2.3 Algoritmos de clasificación y aprendizaje supervisado

En este trabajo de tesis, se utilizan 3 tipos de algoritmos para entrenar los datos: Nearest Neighbor (KNN), Naive Bayes (NB) y Random Forest (RF).

Los costos de los algoritmos son medidos por:

- Accuracy: Corresponde a las predicciones correctas. Verdaderos positivos (TP) + Verdaderos Negativos (TN) / Total de predicciones Positivas (P) + Negativas (N).
 - $P+N = TP+TN+FP+FN$
- Precision = $TP/TP+FP$. Aquí se consideran los falsos positivos. Verdaderos positivos en relación con todos los positivos, es decir, ~~o sea~~ en relación con la suma de todos los verdaderos positivos y falsos positivos de la predicción.
- Recall (sensibilidad) = $TP/TP+FN$ Aquí se consideran los falsos negativos, en relación a todos los positivos actuales.

2.3.1 K Nearest Neighbor (KNN)

Es un algoritmo de *machine learning* de clasificación, que busca los puntos de datos “más similares” (distancia cercana).

En la Figura 8 se muestra un ejemplo de clasificación k-NN

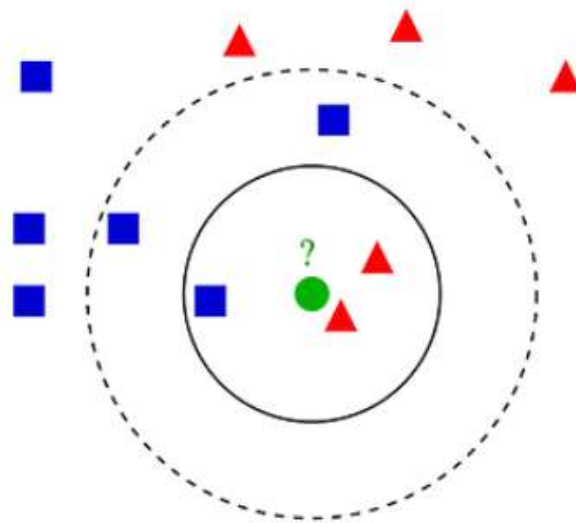


Figura 8: Descripción general de K-vecinos más cercanos (KNN)

Fuente: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

“La muestra de prueba (punto verde) debe clasificarse en cuadrados azules o triángulos rojos. Si $k = 3$ (círculo de línea continua), se asigna a los triángulos rojos porque hay 2 triángulos y solo 1 cuadrado dentro del círculo interior. Si $k = 5$ (círculo de línea discontinua), se asigna a los cuadrados azules (3 cuadrados frente a 2 triángulos dentro del círculo exterior).

Ventajas:

- Se puede utilizar tanto para tareas de regresión como de clasificación, a diferencia de otros algoritmos de aprendizaje supervisado.
- Es muy preciso y fácil de usar. Es fácil de interpretar, comprender e implementar.

- c) No hace suposiciones sobre los datos, lo que significa que puede usarse para una amplia variedad de problemas.

Desventajas:

- a) Almacena la mayoría o la totalidad de los datos, lo que significa que el modelo requiere mucha memoria y es computacionalmente costoso. Los grandes conjuntos de datos también pueden hacer que las predicciones tarden mucho tiempo.
- b) Demuestra ser muy sensible a la escala del conjunto de datos y puede verse afectado por características irrelevantes con bastante facilidad en comparación con otros modelos.” (Nelson, 2020)

2.3.2 Naive Bayes (NB)

Es un algoritmo de *machine learning* de clasificación probabilístico. Usa variables predictoras independientes entre sí.

En la siguiente Figura 9 se ejemplifica la probabilidad inicial de que sea $P(A)$, luego la probabilidad de que sea R dado A , $P(R|A)$. La probabilidad de que sea $P(R)$ y la probabilidad posterior de que sea A dado R , $P(R|A)$.

En el otro evento existe la probabilidad inicial de que sea $P(B)$, luego la probabilidad de que sea R dado B , $P(R|B)$. La probabilidad de que sea $P(R)$ y la probabilidad posterior de que sea B dado R , $P(R|B)$.

Al resultado posterior se deben normalizar las probabilidades, dividiéndolas por la suma de ambas.

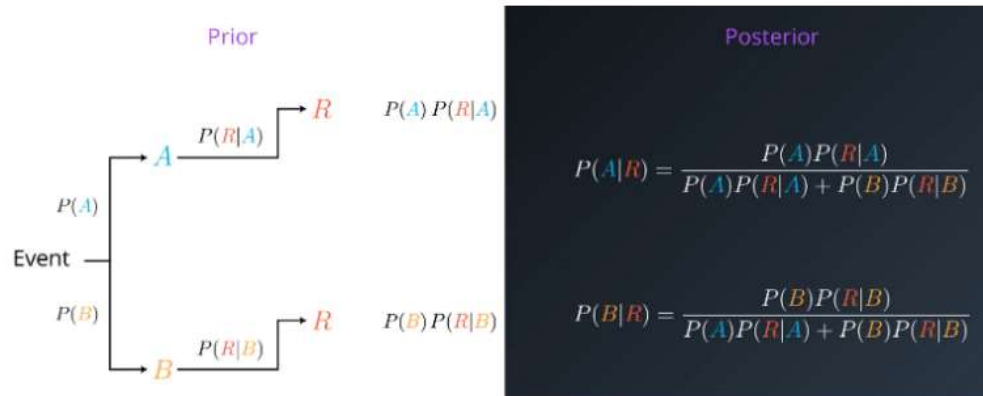


Figura 9: Diagrama ejemplificando la Probabilidades de predecir de Naive Bayes

Fuente: <https://medium.com/datos-y-ciencia/algoritmos-naive-bayes-fundamentos-e-implementaci%C3%B3n-4bcb24b307f>

A continuación, se explican los pasos del algoritmo:

- Convertir el conjunto de datos en una tabla de frecuencias.
- Crear una tabla de probabilidad calculando las correspondientes de que ocurran los diversos eventos.
- La ecuación Naive Bayes se usa para calcular la probabilidad posterior de cada clase.
- La clase con la probabilidad posterior más alta es el resultado de la predicción.

Ventajas:

- Un manera fácil y rápida de predecir clases, para problemas de clasificación binarios y multiclase.
- En los casos en que sea apropiada una presunción de independencia, el algoritmo se comporta mejor que otros modelos de clasificación, incluso con menos datos de entrenamiento.

- c) El desacoplamiento de las distribuciones de características condicionales de clase significa que cada distribución puede ser estimada independientemente como si tuviera una sola dimensión. Esto ayuda con problemas derivados de la dimensionalidad y mejora el rendimiento.

Desventajas:

- a) Aunque son unos clasificadores bastante buenos, los algoritmos Naive Bayes son conocidos por ser pobres estimadores. Por ello, no se deben tomar muy en serio las probabilidades que se obtienen.
- b) La presunción de independencia Naive probablemente no reflejará cómo son los datos en el mundo real.
- c) Cuando el conjunto de datos de prueba tiene una característica que no ha sido observada en el conjunto de entrenamiento, el modelo le asignará una probabilidad de cero y será inútil realizar predicciones. Uno de los principales métodos para evitar esto, es la técnica de suavizado, siendo la estimación de Laplace una de las más populares.” (Roman, 2019).

2.3.3 Random Forest (RF)

Algoritmo de *machine learning* de clasificación que utiliza múltiples árboles de decisión. Depende de los valores de un vector aleatorio probado independientemente y con la misma distribución. Su funcionamiento puede observarse en la Figura 10:

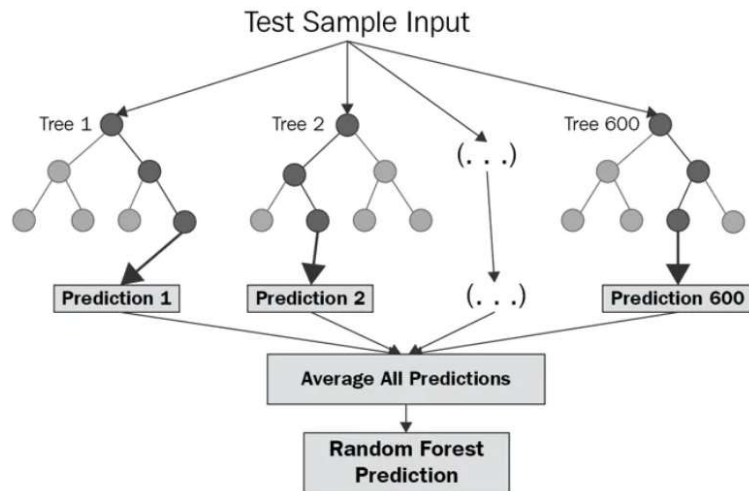


Figura 10: Técnica Random Forest

Fuente: <https://medium.com/swlh/random-forest-and-its-implementation-71824ced454f>

“Cada árbol en las clasificaciones toma información de muestras en el conjunto de datos inicial. Luego, las características se seleccionan al azar, y se utilizan para hacer crecer el árbol en cada nodo. No se debe eliminar ningún árbol hasta el final del ejercicio, cuando se alcanza la predicción de manera decisiva. De esta manera, el Random Forest permite que cualquier clasificador con correlaciones débiles cree un clasificador fuerte.

Ventajas:

- a) Random Forest presenta estimaciones de importancia variable, es decir, redes neuronales. También ofrece un método eficiente para trabajar con datos faltantes. Los valores faltantes se sustituyen por la variable que aparece más veces en un nodo en particular. Entre todos los métodos de clasificación disponibles, Random Forest proporciona la mayor precisión.

La técnica de Random Forest también puede manejar grandes datos con numerosas variables que llegan a miles. Puede equilibrar automáticamente conjuntos de datos cuando una clase es menos frecuente que otras. El método también maneja variables rápidamente, lo que lo hace adecuado para tareas complicadas.” (Corporate Finance Institute, 2021)

CAPÍTULO 3. PLANTEAMIENTO ESTRATÉGICO Y ANÁLISIS DE LA SITUACIÓN ACTUAL

3.1 Posicionamiento Estratégico

En la Figura 11 encontramos el modelo Delta, que permite analizar el posicionamiento estratégico de las empresas. Este *framework*, cuenta con 8 posiciones estratégicas, identificando las competencias propias del negocio (Hax, 2010).

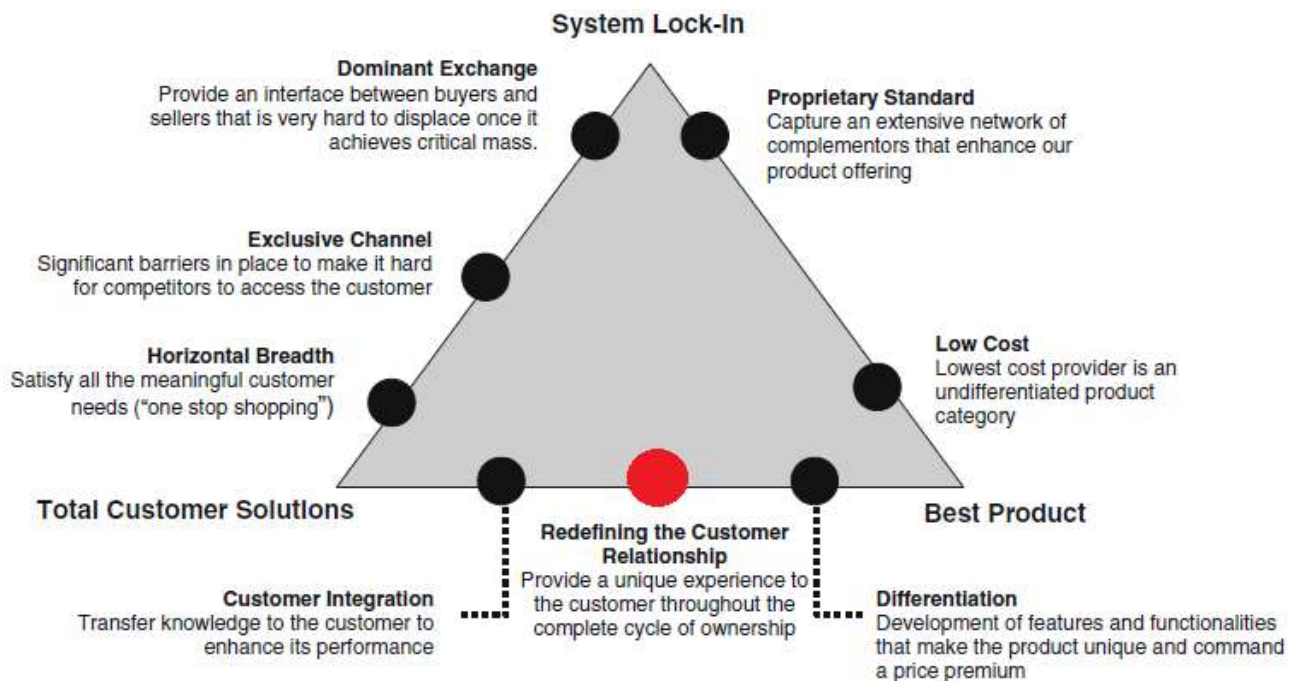


Figura 11: Posicionamiento estratégico de STS de acuerdo con el modelo Delta

Fuente: *The Strategic Tasks of the Delta Model*

Se analiza a STS en este modelo marco y en principio se identifica que, entre sus servicios, la compañía entrega a sus clientes la construcción, mantenimientos preventivos y correctivos de kilómetros de líneas de Tx, con sus respectivas Subestaciones (SSEE) de Tx. Además, en sus Centros de Control, opera en tiempo real estas instalaciones eléctricas, con el Sistema SCADA/EMS. Por todo lo anterior, en el modelo delta podemos posicionar a STS como "Redefiniendo la Relación con el Cliente" (marcado con rojo en la Figura

11). Esto debido a que la empresa, tal como menciona Hax, logra una proximidad cercana al cliente, y cuenta con un conjunto de servicios y conocimientos de los sistemas eléctricos que brindan una experiencia única; los atrae, satisface y retiene en el tiempo (Hax, 2010).

3.2 Modelo de Negocios

El modelo CANVAS (Osterwalder, 2009) permite en un solo lienzo visualizar y explicar los componentes de un modelo de negocio que crea, captura y entrega valor en las organizaciones. Como se puede observar en la Figura 12, el lienzo muestra en el extremo izquierdo a los socios y proveedores, al extremo derecho los segmentos de clientes. En el medio están las actividades claves, recursos de la empresa, propuesta de valor, relación y canales con los clientes. Debajo del modelo marco encontramos una capa financiera con los costos e ingresos de la compañía.










| | | | | |
|--|--|--|--|---|
| <p>Socios Clave </p> <p>Proveedores y socios de:</p> <ul style="list-style-type: none"> *Tecnología y telecomunicaciones. *Maquinarias y servicios de mantenimiento. *Instrumentos de medición. *Constructoras de torres de alta tensión y subestaciones. | <p>Actividades Clave </p> <ul style="list-style-type: none"> *Transporte y transformación de energía eléctrica. *Asesorías de diseño, construcción, mantenimiento y operación en tiempo real de sistemas electricos. <p>Recursos Clave </p> <ul style="list-style-type: none"> *Capital Humano: Personal altamente calificado. *Sistemas de Tecnología Operacional (TO), como son: Sistema SCADA, Sistema de Medición Inteligente. *Inversión para obras nuevas y de ampliación. | <p>Propuesta de Valor </p> <ul style="list-style-type: none"> *Entregar energía confiable y segura. *Profundo compromiso con el cliente. *Modernización de la industria en Chile. | <p>Relación con Clientes </p> <ul style="list-style-type: none"> *Asistencia y contrato de implementación y servicios a largo plazo. *Socio estratégico del cliente para la operación en tiempo real <p>Canales </p> <p>Se encuentran:</p> <ul style="list-style-type: none"> *Sitio Web *APP grupo SAESA *Call Center *Correo electrónico *RRSS del grupo SAESA | <p>Segmentos De Clientes </p> <ul style="list-style-type: none"> *Empresas Distribuidoras y Transmisoras: TRANSEMEL, CGE, ENEL y SAESA *Industrias minera, Lechera, Pesquera, Cemento Bio Bio, etc. *Generadoras eléctricas |
| <p>Estructura De Costos </p> <p>*Movilización de maquinaria, incumplimientos (gastos por multas por fallas), movilización de personal, gasto del personal, gasto de mantenimiento de máquinas. Mantenimiento de instalaciones eléctricas. Gastos de enlace de comunicaciones, soporte y mantenimiento tecnológico.</p> | | <p>Fuente De Ingresos </p> <ul style="list-style-type: none"> *Contratos (ventas) de los servicios ofrecidos por STS. *Pago por peajes de otras empresas que deban pasar la electricidad de AT por las redes de STS. | | |

Figura 12: CANVAS de STS

Fuente: Elaboración propia.

A continuación, se explica el modelo de negocio de STS que ilustra la Figura 12, en cada uno de los 9 módulos del *framework*:

3.2.1 Segmentos de Clientes

STS tiene clientes de diferentes industrias. Por un lado, las empresas de Generación eléctrica, que pagan peaje por el uso de las líneas de Tx. Por otro lado, la compañía tiene contratos de operación en tiempo real y de mantenimiento con otras empresas de Tx, como también relaciones contractuales con empresas de la industria minera, pesqueras, industria lechera y empresas de Distribución eléctrica.

3.2.2 Propuesta de valor

Es entregar energía confiable y segura a sus clientes, con la implementación de una moderna infraestructura eléctrica y TO de punta. Esto permite monitorear en tiempo real el Sistema eléctrico y optimizar los flujos de energía. STS tiene un profundo compromiso con el cliente.

3.2.3 Canales

Los canales de comunicación de STS son diversos y actualmente se está potenciando el uso de nuevas aplicaciones para sus clientes. Existe un portal *web*, servicios de Call Center, correo electrónico y las RRSS del Grupo Saesa.

3.2.4 Relaciones con el cliente

La compañía permitirá lograr una integración con el cliente, que permite ser un socio estratégico y colaborar con la operación de las instalaciones de sus clientes.

3.2.5 Fuentes de ingreso

Los ingresos son por los contratos de los servicios ofrecidos y por el cobro de peajes al usar las líneas y subestaciones de TX.

La Sociedad reconoce ingresos de las siguientes fuentes principales:

- Transmisión.
- Ingresos por venta al detalle de productos y servicios.
- Ingresos por construcción de obras a terceros.

3.2.6 Recursos clave

La empresa cuenta con profesionales especialistas, altamente calificados. STS tiene programas de capacitación continua para sus colaboradores, quienes pueden postular cada año a estos. La compañía el año 2018 invirtió en nuevo y moderno SCADA/EMS, que es de las TO más avanzadas que existen en el mercado. Además, todas sus obras nuevas son con infraestructuras eléctricas contemporáneas, construidas con los más altos estándares de la industria.

3.2.7 Actividades clave

La construcción y ampliación de nuevas instalaciones son parte de sus actividades clave. También lo son el realizar mantenimientos, supervisar y controlar en tiempo real todo el Sistema eléctrico.

3.2.8 Socios claves

STS tiene alianzas estratégicas con algunos proveedores y los hace participar en diferentes licitaciones para implementar proyectos. Se tienen proveedores para la construcción de SSEE de Tx, trabajos en obras civiles, implementación de equipos electrónicos

inteligentes de medición, conexión con empresas de telecomunicaciones y la integración de tecnologías SCADA/EMS.

3.2.9 Estructura de Costos

Los costos anuales por mantenimientos en STS ascienden a un monto de \$15.000 millones. Se distribuyen en costos de mantenimientos programados y correctivos a líneas de Tx y todos los equipos eléctricos dentro de las SSEE de Tx. También existen costos mensuales por los enlaces de comunicaciones, por el soporte a las TO implementadas, y a los equipos electrónicos inteligentes que permiten la automatización, supervisión y control remoto de las SSEE. Además, existen gastos por multas de indisponibilidad eléctrica, en caso de ser multados por el Coordinador Eléctrico Nacional. Finalmente, se realizan gastos por quehaceres administrativos, como contratación de personal y arriendo de inmobiliarios, entre otros.

3.3 Diagnóstico de la Situación Actual

3.3.1 Problema(s) Identificado(s) / Oportunidad(es) identificada(s)

STS tiene como foco 4 desafíos estratégicos:

- i) Aumentar su cobertura nacional y crecer en kilómetros de líneas Tx con sus respectivas SSEE.
- ii) Disminuir los Costos de Operaciones y Mantenimiento (COMA) para ganar nuevas licitaciones y con ello aumentar el EBITDA del Grupo SAESA de un 41,7% al 2021, a un 42,8% el 2023. Las proyecciones se grafican en la Figura 13.

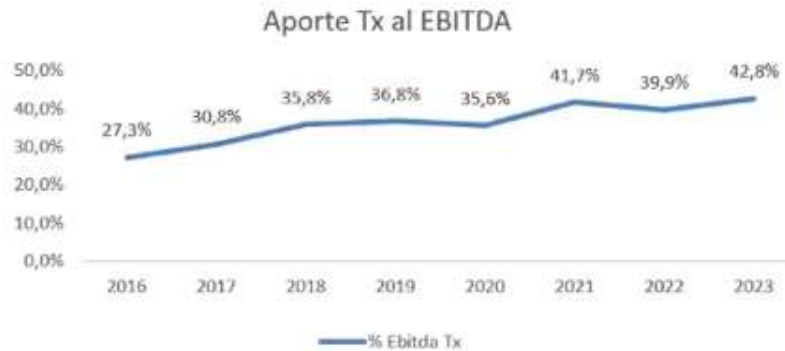


Figura 13: Proyecciones del aporte al EBITDA
Fuente: Proyección estratégica de STS

- iii) Disminuir el índice anual de indisponibilidad SAIDI de 1,08 horas del año 2021, a 0,3 horas al año 2025.



Figura 14: Proyección SAIDI
Fuente: Proyección estratégica de STS

- iv) Disminuir indicadores de Seguridad Tx.

En la Figura 14 se aprecia que, por motivos de mantenimientos preventivos y correctivos, hace 6 años atrás, se desconectaban los clientes en promedio 2,6 horas al año. La proyección para el 2025 es llegar bajo los 18 minutos, logrando estándares europeos.

Para este proyecto de tesis se realizaron una serie de entrevistas a los *stakeholders* de la empresa, principalmente profesionales del área de Calidad de Servicio y Estudios de Tx. Para seleccionar el proceso adecuado a rediseñar, se utiliza la metodología de Gestión de Negocios de Procesos centrada en el Valor (vBPM) de Franz y Kirchmer (2012). Este

proceso se encuentra alineado con los objetivos estratégicos, para que el proceso a intervenir capture, cree y entregue valor a la organización.

En la Figura 15 se enmarcan e identifican los Objetivos estratégicos de la compañía que tienen prioridad, según los profesionales entrevistados.



Figura 15: Objetivos estratégicos

Fuente: Elaboración propia

Dentro de los Objetivos estratégicos antes seleccionados, se califican y ponderan los diferentes *drivers* del negocio. Se destacan en la Figura 16, los drivers del negocio que son relevantes para los stakeholders entrevistados.

VALUE DRIVERS

| Asegurar un estándar operacional superior | Aumentar la productividad y disminuir gastos buscando eficiencias. | Estandarización de procesos y fortalecer la gestión operacional | Asegurar la implementación de proyectos críticos en tiempos establecidos | Implementar prácticas que generen relaciones de largo plazo |
|---|---|---|--|---|
| <ul style="list-style-type: none"> • Reducir a cero el índice SAIDI de fallas. • Implementación de la norma ISO 55001 – Gestión de activos. • Foco en la ingeniería, luego en la construcción y finalmente la operación y mantenimiento. • Ejecutar el mantenimiento preventivo, predictivo y correctivo a los activos físicos • Uso de herramientas digitales para monitoreo, telecontrol, operación y mantenimiento y mejoramiento continuo. • Ejecución de planes y softwares asociados a ciberseguridad. • Realización de auditorías internas y externas. | <ul style="list-style-type: none"> • Reducción de costos operacionales OPEX. • Mantener FTE permanente y temporal dentro de lo requerido. • Mantener la productividad de los técnicos por sobre la media. | <ul style="list-style-type: none"> • Levantamiento de la forma actual de operar. • Elaboración del mapa de procesos. • Identificación de procesos críticos que permitan la continuidad operacional. • Levantamiento de competencias del personal en obra y establecer una metodología de evaluación. • Levantamiento de procesos para la construcción de nuevas instalaciones • Incorporación de tecnología 2D y 3D que permita valorizar las obras y permitan reducir los errores de diseño. • Uso de software PLANNER. • Plan manejo de contingencias imprevistos. | <ul style="list-style-type: none"> • Plan de expansión • Licitación de obras • Obras urgentes • Oportunidades con Obras desiertas licitadas • Entregar en los tiempos planificados los proyectos a operación comercial | <ul style="list-style-type: none"> • Generación de encuestas de satisfacción al cliente. • Plataformas de pago de fácil acceso al cliente. • Evaluación de demanda anticipada del cliente. • Entregar calidad de servicio al cliente. |

Figura 16: Primeros Value Drivers

Fuente: Elaboración propia

En la Figura 17 a continuación también se pueden observar los *drivers* destacados y se agrega en esta figura, una tabla ordenada con los *drivers* que obtuvieron mayor calificación.

VALUE DRIVERS

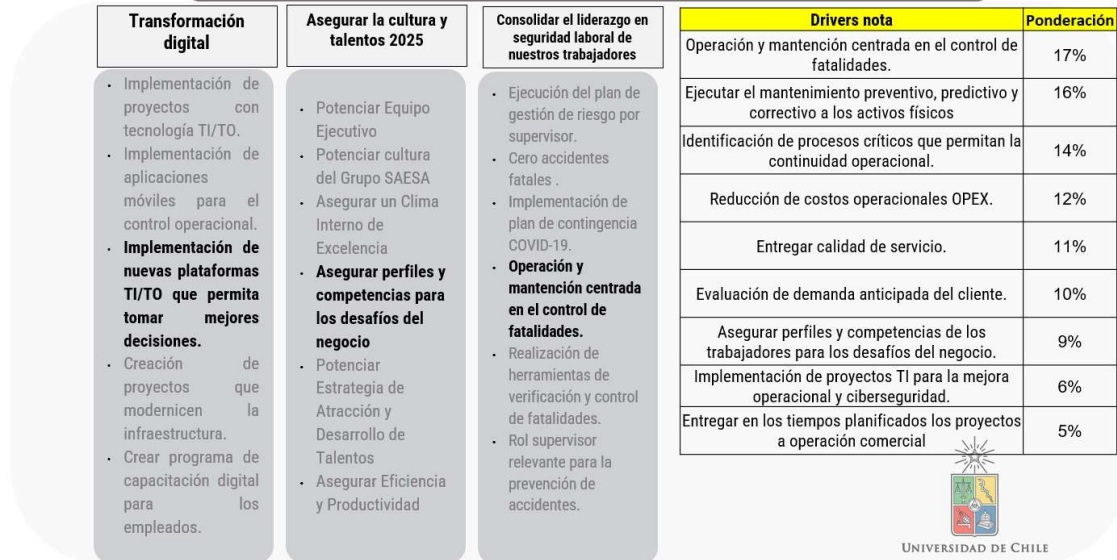


Figura 17: Segundos Value Drivers

Fuente: Elaboración propia

Siguiendo con la metodología, en el árbol de problemas de la Figura 18 se destacan los drivers de negocio que obtienen un corte mayor al 10%.

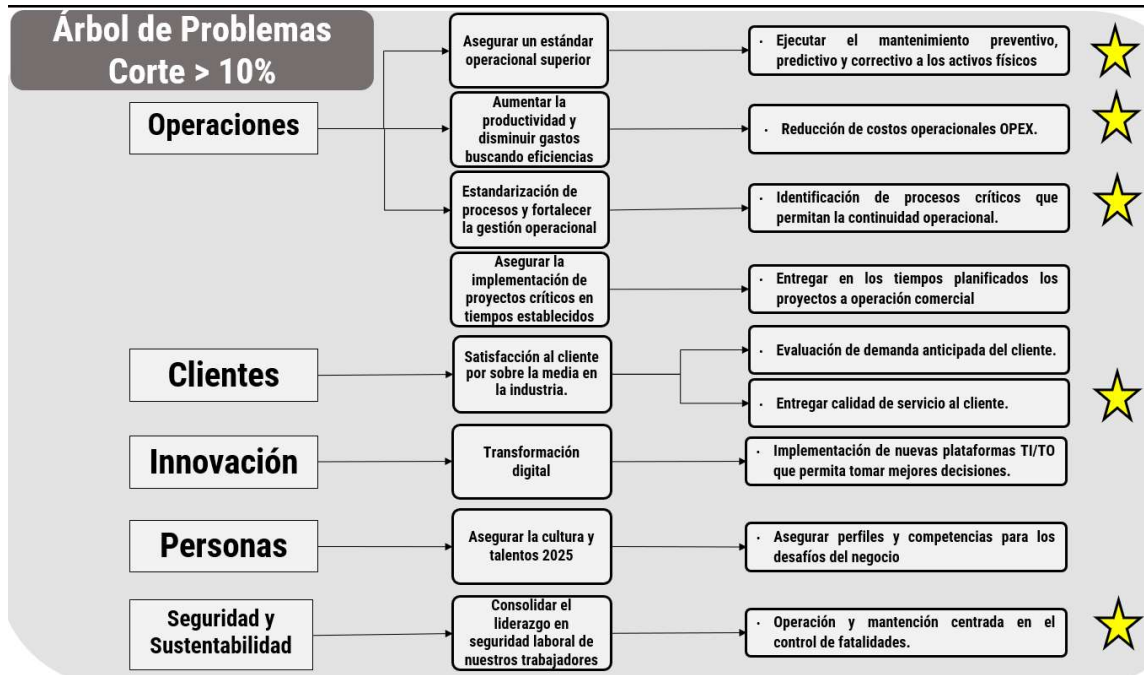


Figura 18: Árbol de Problemas

Fuente: Elaboración propia

La metodología recomienda usar *frameworks* de procesos para encontrar coincidencias con cada uno de los *drivers* de negocios seleccionados. Para el caso de las Utilities se utiliza el *framework* APQC que muestra la Figura 19.

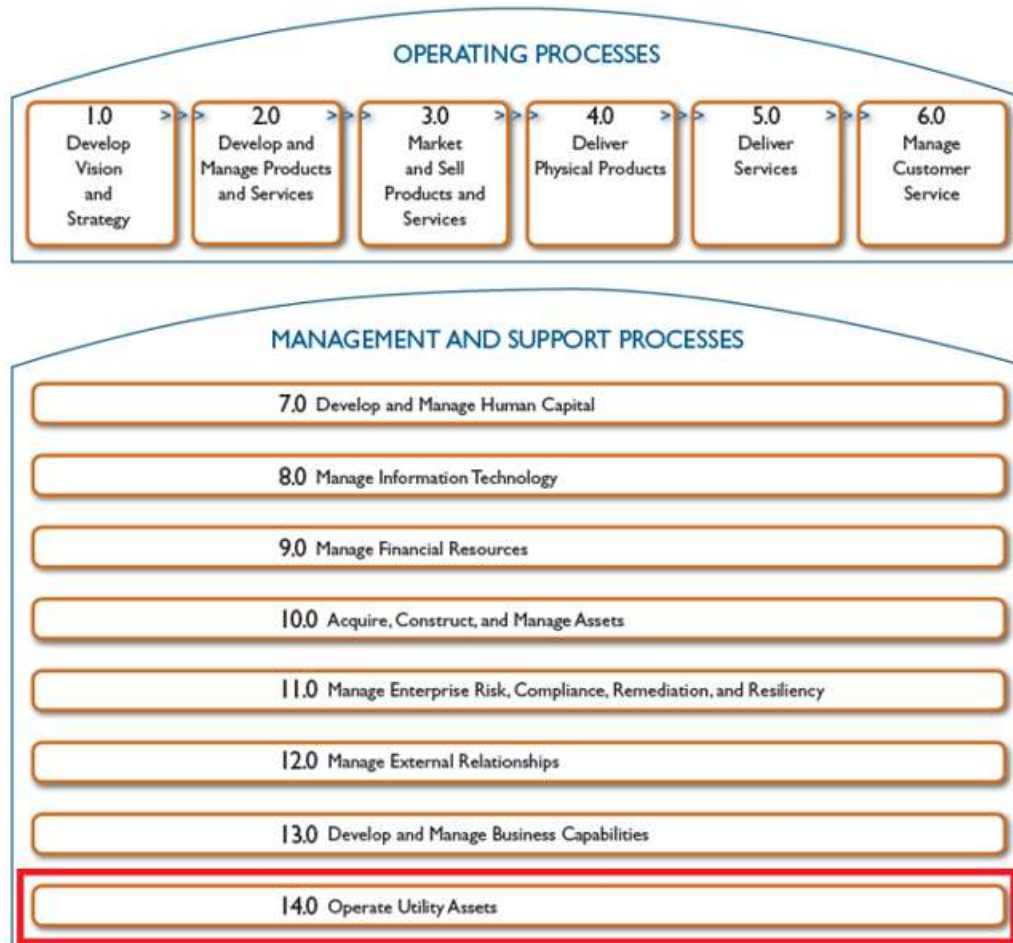


Figura 19: Framework APQC con enfoque en al Utilities

Fuente: Elaboración propia

En este *framework* se encuentran procesos recolectados de forma colaborativa por profesionales de diversas industrias. Para este proyecto de tesis, se trabaja en una matriz que contiene los procesos de APQC y se cruzan con los *drivers* de negocio antes seleccionados por los profesionales entrevistados, calificándolos y otorgándoles un peso según su relevancia estratégica. La importancia estratégica de los procesos APQC en relación con los *drivers* de negocio se evalúa en la Figura 20.

| APQC | Operación y mantención centrada en el control de fatalidades. | Ejecutar el mantenimient o preventivo, predictivo y correctivo a los activos físicos | Identificación de procesos críticos que permitan la continuidad operacional. | Reducción de costos operacionales OPEX. | Entregar calidad de servicio. | Ponderación |
|--|---|--|---|--|-------------------------------------|-------------|
| N3 | 0,21 | 0,21 | 0,2 | 0,2 | 0,18 | 1 |
| 1.4.2.1 Determinar el valor comercial para cada prioridad | 5 | 3 | 2 | 3 | 2 | 3,04 |
| 1.4.3.2 Comunicar iniciativas estratégicas a unidades de negocio y | 5 | 2 | 2 | 3 | 3 | 3,01 |
| 1.5.1.2 Asegurar las aprobaciones adecuadas (20947) | 5 | 2 | 3 | 2 | 2 | 2,83 |
| 10.1.2.2 Realizar el diseño y planificación de la construcción (11278) | 5 | 3 | 2 | 3 | 3 | 3,22 |
| 10.1.2.3 Realizar la construcción (11279) | 5 | 3 | 2 | 3 | 2 | 3,04 |
| 10.1.3.4 Negociar los términos de la instalación (10961) | 5 | 3 | 2 | 3 | 3 | 3,22 |
| 10.1.3.5 Gestionar la construcción (10962) | 5 | 3 | 2 | 3 | 3 | 3,22 |
| 10.2.3.5 Planificar recursos de construcción (19223) | 5 | 2 | 3 | 2 | 2 | 2,83 |
| 10.2.4.1 Programar trabajos de construcción (19230) | 5 | 2 | 3 | 2 | 3 | 3,01 |
| 10.2.5.1 Supervisar el desempeño del trabajo (19225) | 5 | 2 | 3 | 2 | 3 | 3,01 |
| 10.2.5.2 Llevar a cabo el control de calidad de la construcción | 5 | 3 | 2 | 3 | 2 | 3,04 |
| 10.2.5.4 Gestionar la seguridad, la protección y el acceso a los | 5 | 3 | 2 | 3 | 3 | 3,22 |
| 10.2.6.1 Programar la puesta en servicio (19235) | 5 | 2 | 3 | 2 | 2 | 2,83 |
| 10.2.6.2 Coordinar la energización con las operaciones de la red | 5 | 3 | 2 | 3 | 3 | 3,22 |
| 10.2.6.3 Capturar registros de puesta en marcha (19237) | 5 | 2 | 3 | 2 | 2 | 2,83 |
| 13.2.3.1.1 Identificar los requerimientos y objetivos de los | 5 | 2 | 3 | 2 | 2 | 2,83 |
| 13.2.3.1.2 Identificar los requerimientos de los recursos de los | 5 | 3 | 2 | 3 | 3 | 3,22 |
| 13.2.3.2.1 Definir roles y recursos (11123) | 5 | 3 | 2 | 3 | 2 | 3,04 |
| 13.2.3.2.2 Adquirir/asegurar los recursos del proyecto (20142) | 5 | 2 | 3 | 2 | 3 | 3,01 |
| 13.2.3.2.3 Identificar requerimientos de TI específicos (11124) | 5 | 3 | 2 | 3 | 3 | 3,22 |
| 13.2.3.3.1 Evaluar el impacto administrativo del proyecto (estrategia y proyectos) sobre las medidas y resultados (11131) | 5 | 2 | 3 | 2 | 3 | 3,01 |
| 13.2.3.3.2 Reportar el estatus del proyecto (16415) | 5 | 3 | 2 | 3 | 2 | 3,04 |
| 13.2.3.5 Cerrar el Proyecto (16418) | 5 | 4 | 3 | 2 | 3 | 3,43 |

Figura 20: Primeros Procesos con nota y ponderación

Fuente: Elaboración propia

En la Figura 21, se destacan en color rojo los procesos APQC en relación con los *drivers* de negocio, que obtienen mayor puntuación, con un corte superior a nota 4.

| APQC | Operación y mantención centrada en el control de fatalidades. | Ejecutar el mantenimient o preventivo, predictivo y correctivo a los activos físicos | Identificación de procesos críticos que permitan la continuidad operacional. | Reducción de costos operacionales OPEX. | Entregar calidad de servicio. | Ponderación |
|---|---|--|---|--|-------------------------------------|-------------|
| N3 | 0,21 | 0,21 | 0,2 | 0,2 | 0,18 | 1 |
| 14.1.1.1 Determinar las metas y objetivos operativos de la red | 5 | 3 | 2 | 3 | 2 | 3,04 |
| 14.1.1.2 Especificar y seleccionar las políticas operativas (19266) | 5 | 2 | 3 | 2 | 3 | 3,01 |
| 14.1.2.1 Especificar requisitos de operación (19268) | 5 | 3 | 2 | 4 | 2 | 3,24 |
| 14.1.2.3 Crear planes de reconfiguración de red / conmutación | 5 | 2 | 3 | 3 | 2 | 3,03 |
| 14.1.2.4 Simular / calcular la demanda y las condiciones de la red [voltaje eléctrico / presiones de agua y gas, etc.] (19271) | 5 | 3 | 3 | 2 | 3 | 3,22 |
| 14.1.3.1 Desarrollar / revisar los protocolos operativos de la red | 5 | 3 | 2 | 3 | 3 | 3,22 |
| 14.1.3.3 Recopilar información e inteligencia sobre el uso y la | 5 | 4 | 3 | 4 | 4 | 4,01 |
| 14.1.3.4 Producir informes de la condición de la red (19275) | 5 | 4 | 3 | 4 | 4 | 4,01 |
| 14.1.3.6 Monitorear e informar el desempeño operativo y | 5 | 4 | 4 | 4 | 3 | 4,03 |
| 14.1.4.1 Monitorear el estado actual del interruptor / válvula de la red y los datos de desempeño operativo (19279) | 5 | 4 | 3 | 3 | 4 | 3,81 |
| 14.1.4.2 Mantener el historial de estado de la red (19280) | 5 | 4 | 3 | 4 | 4 | 4,01 |
| 14.1.4.3 Supervisar alertas y alarmas (19281) | 5 | 3 | 4 | 3 | 4 | 3,8 |
| 14.1.5.1 Identificar fallas de red (19282) | 5 | 4 | 3 | 4 | 4 | 4,01 |
| 14.1.5.2 Ejecutar ubicación de fallas, aislamiento, restauración | 5 | 4 | 4 | 4 | 5 | 4,39 |
| 14.1.5.3 Coordinar las operaciones de reparación y restauración | 5 | 4 | 3 | 4 | 4 | 4,01 |
| 14.1.6.3 Programar operaciones de red planificadas (19288) | 5 | 3 | 2 | 3 | 3 | 3,22 |
| 14.1.6.5 Ejecutar conmutaciones y desvíos planificados (19290) | 5 | 4 | 3 | 4 | 4 | 4,01 |
| 14.1.6.6 Coordinar con operaciones de mantenimiento y | 5 | 5 | 5 | 5 | 5 | 5 |
| 14.1.6.7 Volver a poner la red en servicio (19292) | 5 | 2 | 3 | 3 | 3 | 3,01 |
| 14.1.7.3 Realizar controles de optimización de estabilidad, tensión, y voltaje de la red (19293) | 5 | 4 | 3 | 4 | 4 | 4,01 |
| 14.1.7.4 Mantener la configuración y los controles de protección | 5 | 2 | 3 | 3 | 2 | 3,03 |
| 14.1.7.6 Monitorear y evaluar el desempeño de la red (19298) | 5 | 4 | 3 | 4 | 4 | 4,01 |
| 14.1.8.1 Conducir planificación de respuesta a emergencias (19299) | 5 | 4 | 3 | 4 | 4 | 4,01 |
| 14.1.8.2 Identificar y gestionar las fallas de los clientes (19300) | 5 | 4 | 3 | 4 | 4 | 4,01 |
| 14.1.8.3 Gestionar y aplicar las normas de seguridad (19301) | 5 | 3 | 2 | 3 | 3 | 3,22 |

Figura 21: Primeros Procesos seleccionados con nota y ponderación

Fuente: Elaboración propia

Como se menciona en los párrafos anteriores. La metodología califica con visión de expertos, colocándose además en el rol del cliente. Los procesos seleccionados del *framework* APQC son nuevamente evaluados. La nueva matriz de la Figura 22, se construye con la puntuación de la estandarización posible de conseguir dentro de la organización, además de la importancia estratégica y sus costos.

| N° | Procesos | Estandarización | Importancia Estratégica | Costos o MM\$ |
|----|--|-----------------|-------------------------|---------------|
| 1 | 14.1.3.3 Recopilar información e inteligencia sobre el uso y la disponibilidad de la red (19274) | 18 | 13 | 3 |
| 2 | 14.1.3.4 Producir informes de la condición de la red (19275) | 16 | 16 | 2 |
| 3 | 14.1.3.6 Monitorear e informar el desempeño operativo y regulatorio general (19277) | 12 | 11 | 3 |
| 4 | 14.1.4.2 Mantener el historial datos de estado de la red (19280) | 11 | 11 | 3 |
| 5 | 14.1.7.3 Realizar controles de optimización de estabilidad, tensión, presión y carga (19296) | 17 | 14 | 2 |
| 6 | 14.1.5.2 Ejecutar ubicación de fallas, aislamiento, restauración (19283) | 17 | 13 | 2 |
| 7 | 14.1 .5.3 Coordinar las operaciones de reparación y restauración (19284) | 18 | 17 | 2 |
| 8 | 14.1.6.5 Ejecutar conmutaciones y desvíos planificados (19290) | 17 | 15 | 3 |
| 9 | 14.1.5.1 Identificar fallas de red (19282) | 19 | 18 | 1 |
| 10 | 14.1.7.6 Monitorear y evaluar el desempeño de la red (19298) | 18 | 14 | 2 |
| 11 | 14.1.8.1 Conducir planificación de respuesta a emergencias (19299) | 15 | 16 | 1 |
| 12 | 14.1.8.2 Identificar y gestionar las fallas de los clientes (19300) | 17 | 18 | 1 |

Figura 22: Matriz de estandarización

Fuente: Elaboración propia

En la figura anterior se destacan en color verde los procesos que obtienen un máximo de puntuación. Para su mejor comprensión, el resultado de la tabla anteriormente expuesta se grafica en la Figura 23, donde la importancia estratégica será el eje Y, siendo representado el eje X como la estandarización posible de lograr.



Figura 23: Gráfico matriz de estandarización

Fuente: Elaboración propia

El resultado de todos los procesos que entran en el cuarto cuadrante del gráfico de la Figura 23, se triangulan como muestra la siguiente Figura 24, y se evalúan con una puntuación del 1 al 3. Para la capacidad BPM, la máxima puntuación para mejoras en el proceso es 3, aplicando Six Sigma o cualquier otra metodología de la filosofía Lean. También se considera lo imperativo del negocio, puntuando con nota 3 lo que es realmente necesario a intervenir en la empresa. Ver en Anexo A la triangulación realizada a todos los procesos seleccionados.

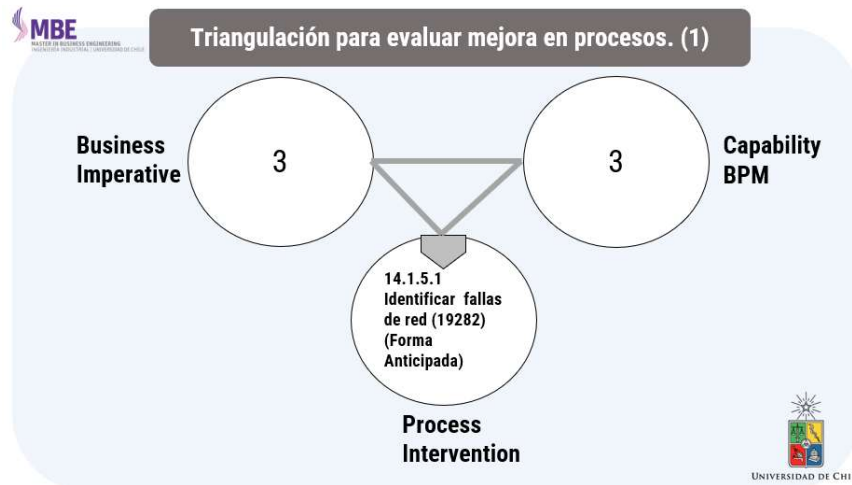


Figura 24: Triangulación para evaluar primer proceso seleccionado.

Fuente: Elaboración propia

De la triangulación explicada, se concluye que la metodología vBPM realiza una “depuración” de los procesos en varias etapas para seleccionar solo uno. En la Figura 25 se muestran los 5 procesos que se consideran clave y destacado en verde el seleccionado.

| N° | Procesos | Factibilidad | Urgencia | Importancia Estratégica | Impacto al negocio |
|----|---|--------------|----------|-------------------------|--------------------|
| 9 | 14.1.5.1 Identificar fallas de red (19282) | 3 | 3 | 18 | 21 |
| 11 | 14.1.8.1 Conducir planificación de respuesta a emergencias (19299) | 3 | 1 | 17 | 17 |
| 12 | 14.1.8.2 Identificar y gestionar las fallas de los clientes (19300) | 2 | 2 | 18 | 20 |
| 7 | 14.1.5.3 Coordinar las operaciones de reparación y restauración (19284) | 3 | 1 | 17 | 18 |
| 2 | 14.1.3.4 Producir informes de la condición de la red (19275) | 2 | 2 | 16 | 18 |

Figura 25: Tabla Matrix Impacto / Esfuerzo.

Fuente: Elaboración propia

Finalizando esta metodología, en la Figura 26, el proceso a intervenir se grafica en una Matriz de Impacto/ Esfuerzo, que permite visualizar el proceso que tiene mayor factibilidad y efecto dentro de la organización.

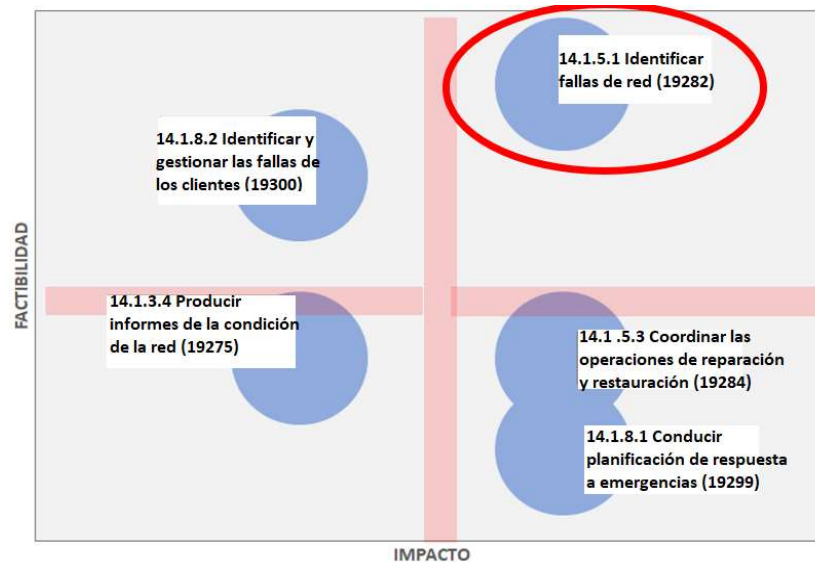


Figura 26: Gráfico Matrix Impacto / Esfuerzo.

Fuente: Elaboración propia

El proceso seleccionado, impulsa la innovación y creatividad para identificar las fallas de red utilizando estudios con nuevas capacidades. En este trabajo se redefine al proceso APQC con el nombre de “Identificar **con predicciones** fallas de red”.

El proceso está relacionado con los desafíos de disminuir los COMA y los SAIDI. Se identifica en un análisis de causa raíz (Anexo B: Selección del Problema), que no existen las herramientas que contribuyan a predecir una falla en los equipos eléctricos. Es imprescindible contar con una serie de *dataset* de fallas históricas de STS y otras empresas eléctricas, que se desconocen y deben ser identificadas.

Con los colaboradores se concluye que es necesario colocar foco en un solo elemento eléctrico en particular y se decide trabajar con los Transformadores de Poder que se encuentran dentro de una S/E de Tx, considerando que es un activo clave para la operación.

Esta tesis supone que, si se integra con éxito un modelo predictivo para identificar los mantenimientos programados y de curso forzoso de los transformadores de poder, se

podrán usar en un futuro cercano, las metodologías aplicadas en este trabajo para construir nuevos modelos de predicción para otros activos eléctricos. Como muestra la Figura 27, un transformador de poder es un equipo de grandes dimensiones y compleja reparación.



Figura 27: Cambio de transformador en falla en S/E Melipulli de la empresa STS

Fuente: Falla en S/E Melipulli de STS

En la figura 28 se muestran los recursos, con las dificultades para trasladar y reemplazar los transformadores de poder.

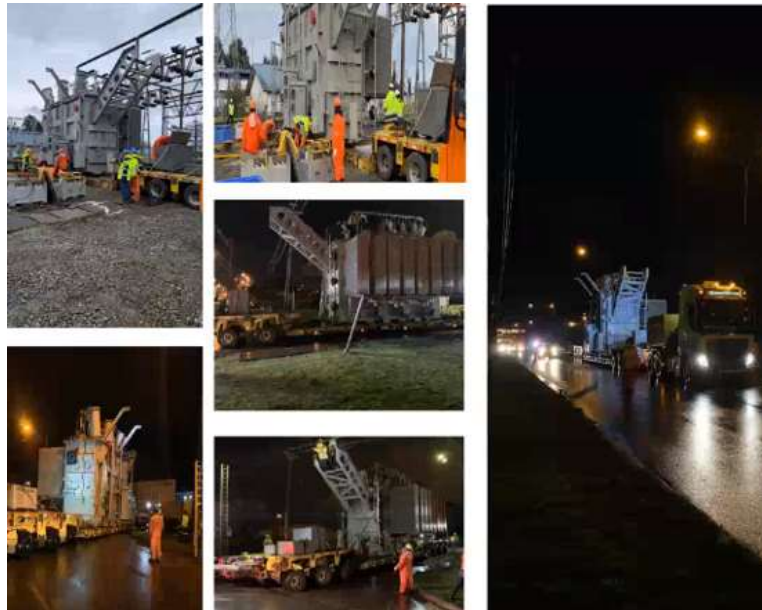


Figura 28: Traslado de nuevo transformador hacia S/E Melipulli de la empresa STS

Fuente: Falla en S/E Melipulli de STS

Se utiliza un camión de grandes dimensiones y de giro amplio. Esto requiere la implementación de una compleja logística de traslado en las ciudades, afectando los tiempos de respuesta para el reemplazo del equipo. Para el reemplazo de un transformador se requieren más de 3 semanas de trabajo. Existen aproximadamente 106 Transformadores en la compañía y es el activo más caro de las SSEE de Tx.

Como se mencionó anteriormente, uno de los problemas en STS, es que en la actualidad no existe un método que permita anticiparse a una falla de un transformador de poder (ver Anexo B: Selección del Problema).

De las entrevistas realizadas, se pueden nombrar las siguientes dificultades, cuando falla un Transformador de Poder:

- La falla de un transformador afecta la calidad de servicio al cliente. La calidad de servicio la fiscaliza la Superintendencia de Electricidad y Combustible (SEC), la cual solicita a la empresa eléctricas indicadores de voltaje mínimos y máximos que se entregan a los clientes.
- Si falla un transformador, afecta los índices de disponibilidad, que no pueden superar los 3 minutos en una falla para el cliente. Estos índices son exigidos por el Coordinador Eléctrico Nacional y existe riesgo de multas si los superan.
- Si falla un transformador aumentan los costos de mantenimiento, ya que obligaría en algunos casos al reemplazo del equipo, lo que tiene un costo de USD 3 millones por cada activo nuevo. Si se repara reemplazando partes, el costo puede disminuir a USD 1 millón.

Entonces, si falla un transformador, esto implica costos altos. Se debe movilizar gran cantidad de personal de trabajo, equipos técnicos especialistas propios de la empresa, como también sumar diferentes contratistas para el apoyo técnico y de logística.

3.3.2 Arquitectura de Procesos AS IS (Patrones)

El primer análisis de los macroprocesos para STS concluye que el Coordinador Eléctrico Nacional, además de ser la autoridad fiscalizadora, es en parte una cadena de valor

extendida. Esto debido a que ambas organizaciones comparten datos en tiempo real y supervisan las instalaciones eléctricas, para cumplir con operar todo el Sistema Eléctrico Nacional (SEN). Todo esto se puede observar en la Figura 29.

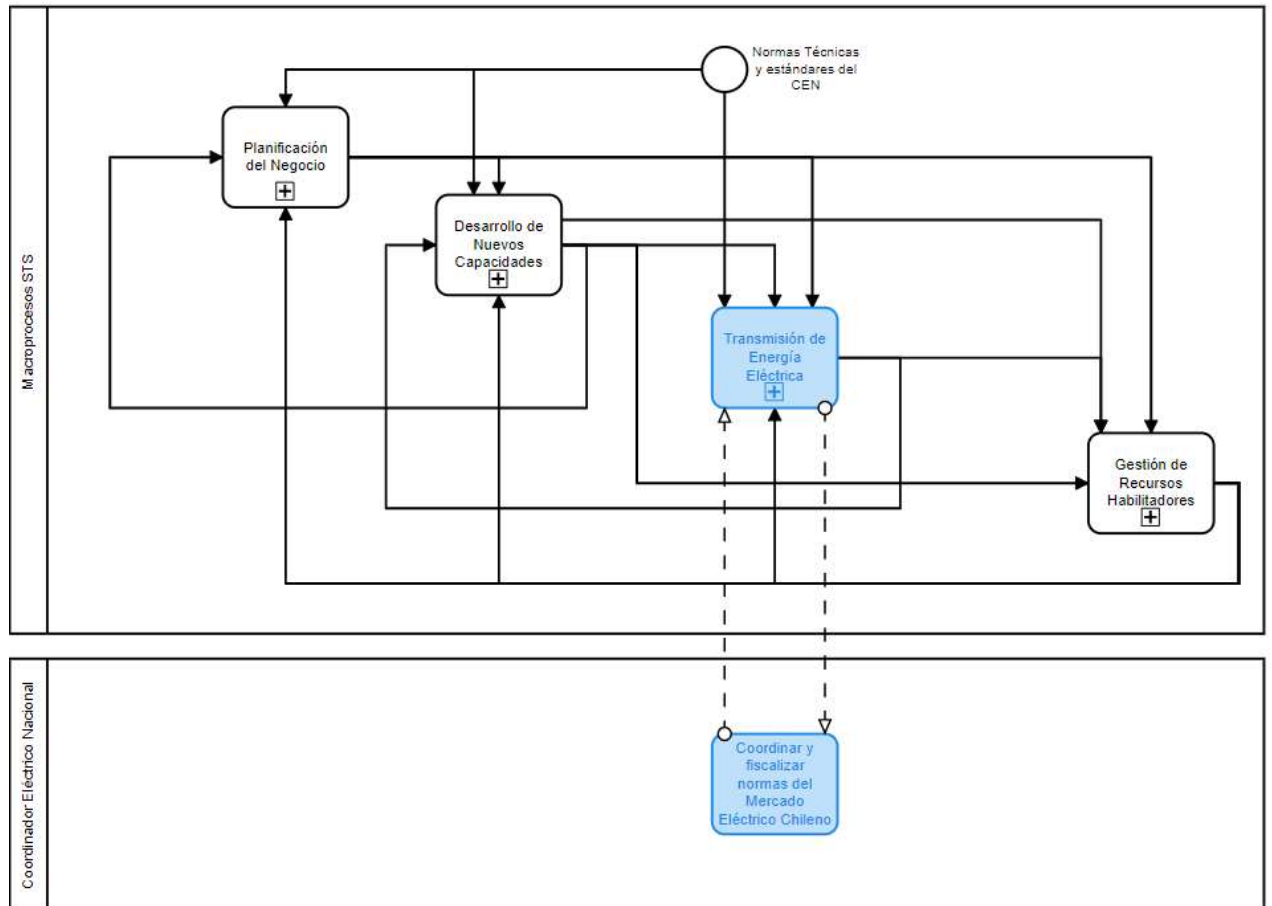


Figura 29: Análisis de Macroprocesos de STS con el Coordinador Eléctrico Nacional

Fuente: Elaboración propia

Continuando con el modelamiento de los procesos a rediseñar en esta Tesis, se describe la Macro1 en la Figura 30.

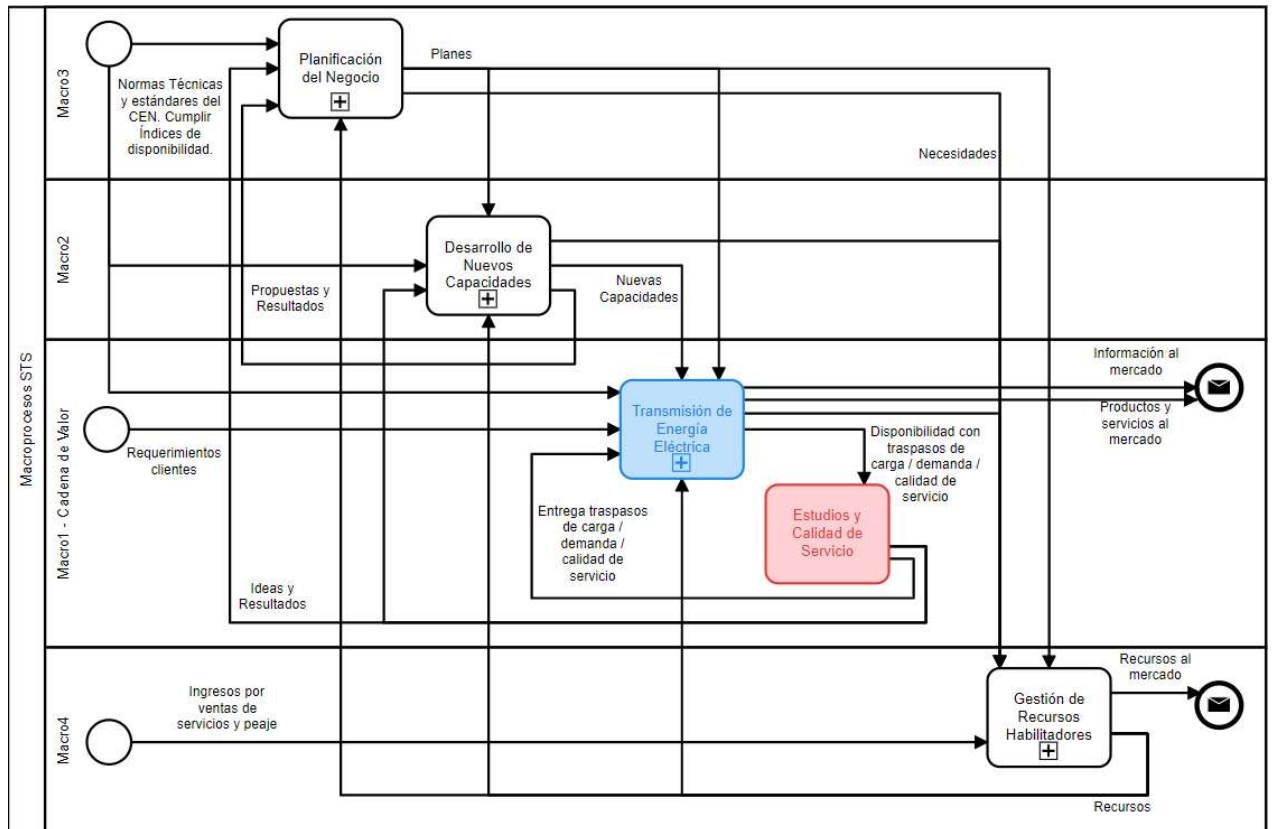


Figura 30: Macro 1 STS

Fuente: Elaboración propia

Esta figura expone a los Macroprocesos de STS y se enfoca en el rediseño de los procesos para la “Transmisión de Energía Eléctrica” y los procesos de “Estudios y Calidad de Servicio”.

La Transmisión de Energía Eléctrica es la Cadena de Valor de STS, y dentro de este macroproceso se encuentra el área de Estudio y Calidad de Servicio. Esta área tiene como *input* requerimientos para asegurar la disponibilidad y calidad del servicio eléctrico. Como *output* aporta con sus estudios a los traspasos de carga eléctrica, para evitar las desconexiones de los equipos del sistema. También realiza estudios de demanda y calidad de servicio, para asegurar la entrega de potencia y voltaje, que debe estar dentro de los rangos exigidos por la autoridad.

En la siguiente Figura 31 se ilustra el segundo nivel de la Macro1 y se apertura el proceso de “Transmisión de Energía Eléctrica”.

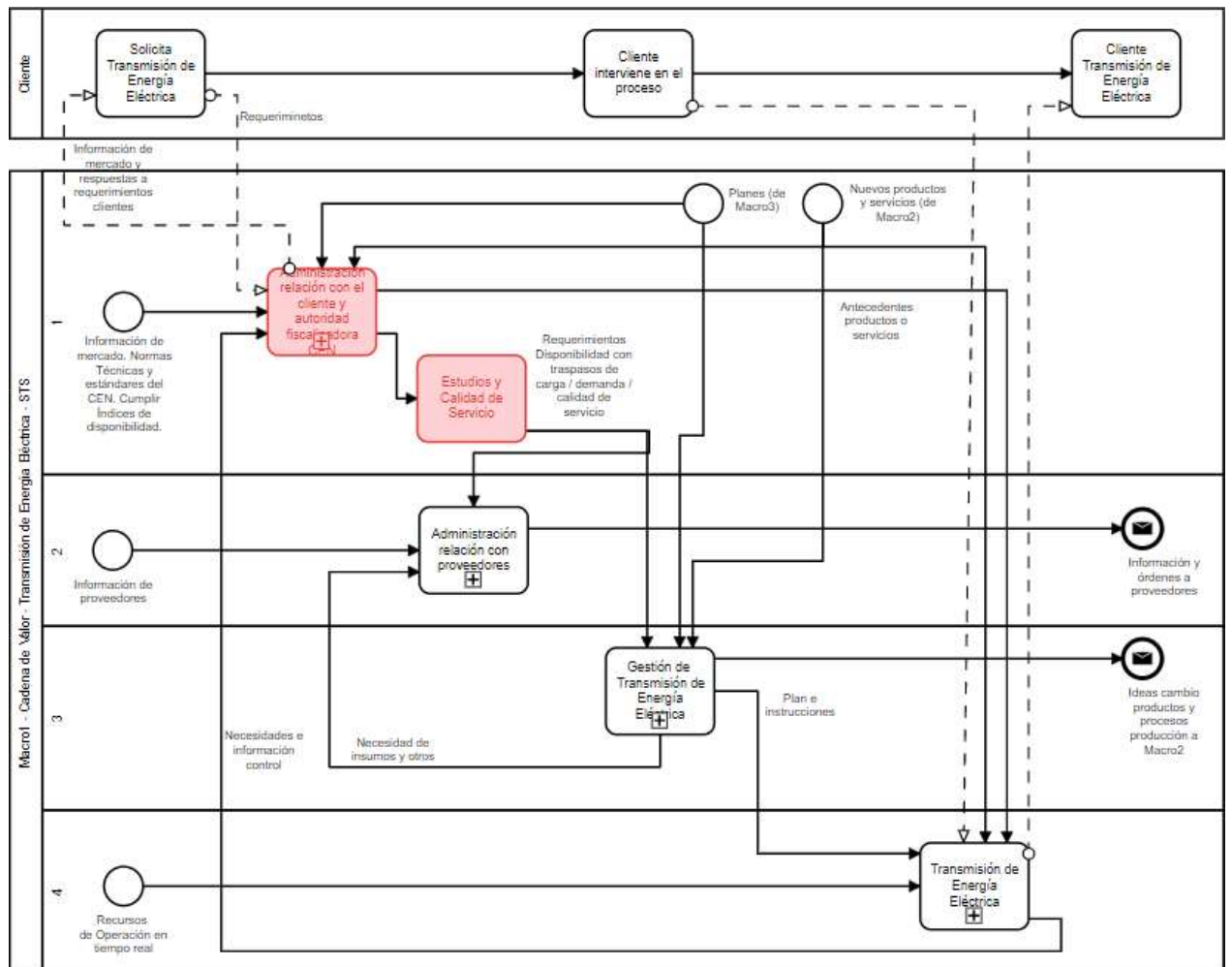


Figura 31: Macro 1 Transmisión de Energía Eléctrica de STS

Fuente: Elaboración propia

En esta apertura se aprecia que los procesos de “Administración relación con el cliente y autoridad fiscalizadora”, se encargan de cumplir con los requerimientos de los clientes y todas las exigencias de la Norma Técnica.

En la Figura 32, que se encuentra a continuación, se presenta la apertura realizada a los procesos de “Administración relación con el cliente y autoridad fiscalizadora”, y se muestran los procesos de “Estudios y Calidad de Servicio”.

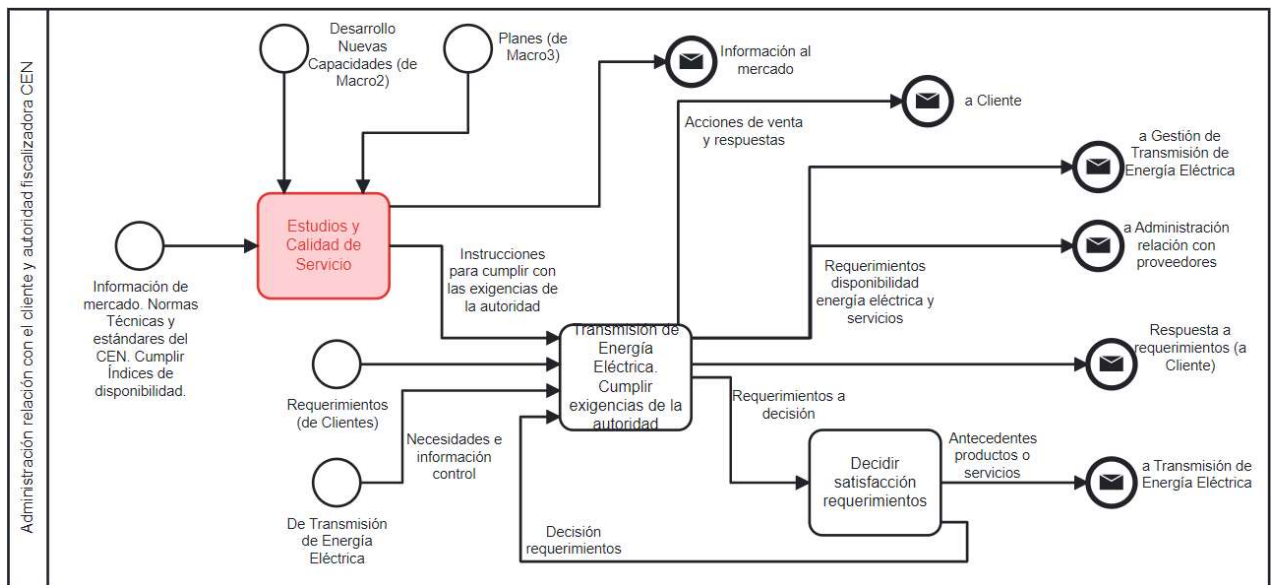


Figura 32: Macro 1 Administración con el cliente y autoridad fiscalizadora CEN

Fuente: Elaboración propia

Respecto a la apertura de los procesos señalados en la Figura 33, que se encuentra a continuación, estos entregan estudios e instrucciones para definir acciones para la continuidad del servicio eléctrico. Este análisis se profundiza con un diagrama BPMN en el siguiente capítulo 3.3.3.

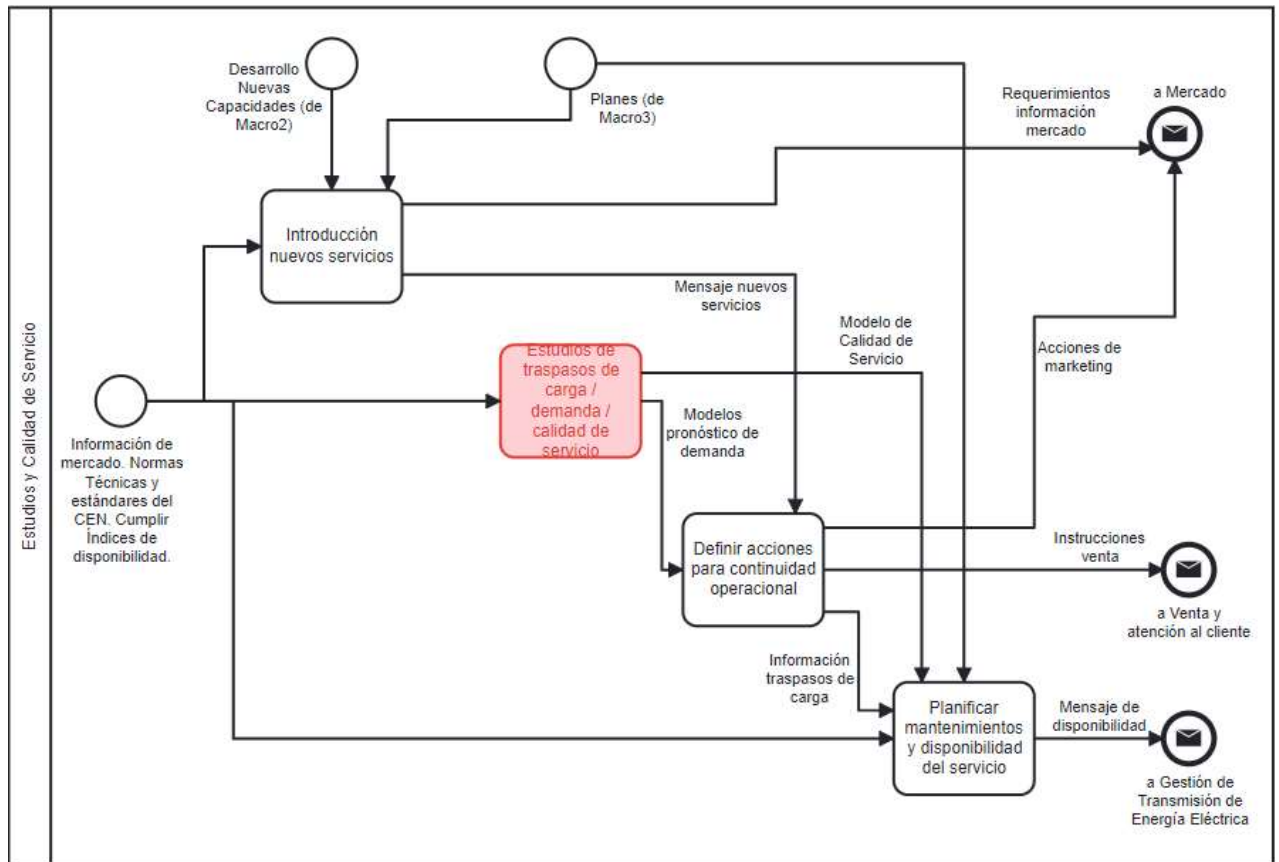


Figura 33: Macro 1 Estudio y Calidad de Servicio

Fuente: Elaboración propia

3.3.3 Modelamiento Detallado de Procesos AS IS (BPMN)

En este capítulo se detallan con un diagrama BPMN, los procesos que existen actualmente en el área de Estudios y Calidad de Servicio.

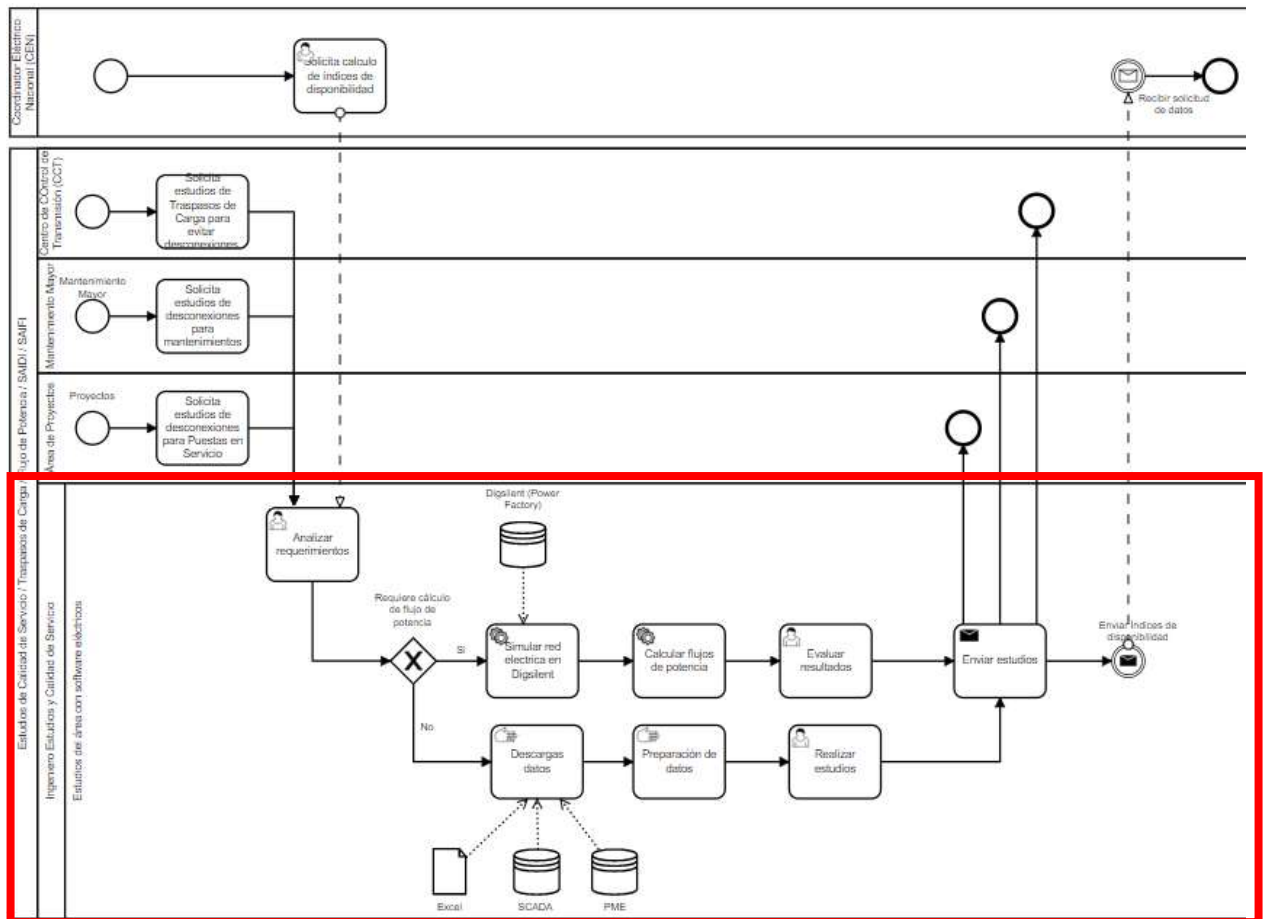


Figura 34: BPMN Área de Estudios y Calidad de Servicio

Fuente: Elaboración propia

En la Figura 34 anterior, se puede apreciar la llegada de requerimientos de diferentes áreas. El Centro de Control de transmisión (CCT), solicita estudios para traspasar cargas por distintas rutas de la red eléctrica y evitar una desconexión. También llegan solicitudes de estudios desde el área de Mantenimiento Mayor, para desconectar algún elemento de la red y con esto realizar un mantenimiento correctivo o preventivo. Existen, además, requerimientos desde el área de Construcción para desconectar equipos y realizar las Puestas en Servicios (PES) de algún nuevo activo en la red eléctrica.

Los procesos ilustrados son dos, y dependen del tipo de estudio a realizar. Por un lado, para simular distintos escenarios de la red eléctrica y realizar cálculo de flujos de potencia,

es necesaria la utilización de software (SW) especializados, como Digsilent o SCADA/EMS. Por otro lado, se realizan otros estudios eléctricos, como son por ejemplo los pronósticos de demanda y trasposos de carga.

3.4 Cuantificación del Problema u Oportunidad

Debido a que el problema es de predicción de mantenimientos programados y de curso forzoso, se detalla la siguiente cuantificación del problema.

Se poseen costos de los mantenimientos realizados a los transformadores (TTFF). Es un costo mínimo aproximado de \$2.000.000.- por cada uno de los TTFF, los que se gastan en inspecciones a los equipos y mantenimientos preventivos. Anualmente se asigna un presupuesto aproximado de \$120.000.000.- para mantener e inspeccionar algunos TTFF. Este monto corresponde al 5% del total de mantenimientos realizados a las instalaciones y resto de los equipos eléctricos. Pero, cuando ocurre una falla en un TTFF, se puede llegar a gastar, en el peor de los escenarios, de 8 a 10 veces el monto anual de los mantenimientos preventivos, solo por un equipo.

El año 2019 ocurrió efectivamente una falla en el TTFF 44 de la subestación Melipulli en Puerto Montt, llegando a gastar en el reemplazo del equipo \$900.000.000.-

La oportunidad es anticiparse a los mantenimientos de curso forzoso de los TTFF para evitar las fallas, convirtiendo los actuales mantenimientos en predictivos.

CAPÍTULO 4. PROPUESTA DE DISEÑO DE PROCESOS

4.1 Direcciones de cambio y alcance

El actuar de STS en los últimos años se ha basado en realizar mantenimientos preventivos y correctivos a sus instalaciones y equipos eléctricos. Desde hace 7 años, gracias a su estrategia de negocio y a las inversiones en tecnología, ha logrado disminuir los índices de disponibilidad SAIDI de 2,06 horas en el año 2015, a 1,08 horas al año 2021. Como se mencionó en el capítulo 3.3.1, STS tiene pronosticado llegar al 2025 a índices europeos, que no superen los 18 minutos de desconexión al año. Además, es necesario ser eficientes en el uso de los COMA y utilizar todas las herramientas y recursos disponibles para aportar a este desafío. Estos focos pueden ser conseguidos con la ayuda de los tratamientos y procesamientos de datos, transformándolos en información, para tomar decisiones en forma predictiva.

Mientras se escribe esta tesis, la compañía se encuentra trabajando en su Plan Estratégico de Gestión de Activos (PEGA). El objetivo es certificarse en la norma ISO 55.001, que permite mejorar la toma de decisiones en los gastos del ciclo de vida de un activo productivo. El PEGA en la Gerencia de Tx, tiene como objetivo Operacional y Mantenimiento, implementar una gestión predictiva de los activos operativos al año 2024.

En resumen, la dirección que se toma con este trabajo de título, es rediseñar los estudios en base a la eficiencia de los recursos actuales, anticipándose a posibles desconexiones de curso forzoso. Además, se apunta a complementar y aportar al plan estratégico, con la entrega de un MVP, con tres modelos predictivos de clasificación de Machine Learning. El objetivo este trabajo es anticiparse a las desconexiones de curso forzoso de los transformadores de poder y llevarlas a desconexiones programadas, ya que estas últimas no afectan al indicador SAIDI y, además, al ser programadas, no afectan en forma considerable a los COMA.

Para esto, se realizan análisis de variables de cambio con la metodología de Ingeniería de negocios de Barros (2015) y la metodología CRISP-DM para solucionar problemas de Data Analytics.

4.1.1 Estructura de la Empresa y Mercado

Variable de cambio con mayor impacto, si es que se consideran cambios a nivel estratégico, el modelo de negocio o de arquitectura de procesos.

Tabla 4: Estructura Empresa y Mercado.

| a | Variable de Diseño: Estructura Empresa y Mercado | Actual | Propuesta |
|-----|--|---|---|
| a.1 | Servicio Integral al Cliente | No. Redefiniendo la relación con el cliente | Fortalecer los mantenimientos en forma predictiva, para no afectar la calidad de servicio entregado al cliente. |
| a.2 | Lock-in Sistémico | No | Mantener situación actual |
| a.3 | Integración con Proveedores | No | Mantener situación actual |
| a.4 | Estructura Interna: Centralizada o Descentralizada | Centralizada | Mantener situación actual |
| a.5 | Toma de Decisiones: Centralizada o Descentralizada | Centralizada | Mantener situación actual |

4.1.2 Anticipación

Esta variable analiza la capacidad de anticiparse a eventos dentro de la organización.

Tabla 5: Anticipación.

| b | Variable de Diseño: Anticipación | Actual | Propuesta |
|-----|---|---|--------------------------------------|
| b.1 | Planificación de mantenimientos preventivos y correctivos | Si, se programan por calendario y por eventos | Programar mantenimientos predictivos |

| | | | |
|-----|---|----|---|
| b.2 | Modelos predictivos para anticiparse a las desconexiones de curso forzoso | No | Incorporar modelos de Machine Learning que apoyen en la planificación |
|-----|---|----|---|

4.1.3 Coordinación

Variable que identifica la necesidad de reformular las relaciones entre las áreas directivas y ejecutivas de la empresa.

Tabla 6: Coordinación.

| c | Variable de Diseño: Coordinación | Actual | Propuesta |
|-----|-------------------------------------|--|--|
| c.1 | Reglas | Se asignan mantenimientos predictivos sin apoyo de las TI y los mantenimientos correctivos son gatillados por eventos. | Incorporar nuevas reglas con el apoyo de Machine Learning para planificar mantenimientos predictivos |
| c.2 | Jerarquía | Jerarquía administrativa tradicional | Mantener situación actual |
| c.3 | Colaboración | Si | Mantener situación actual |
| c.4 | Participación | Por requerimientos | Mantener situación actual |

4.1.4 Prácticas de Trabajo

Variable de cambio que visualiza como ejecutar las tareas, para cumplir con el rediseño del proceso propuesto.

Tabla 7: Prácticas de Trabajo.

| d | Variable de Diseño: Prácticas de Trabajo | Actual | Propuesta |
|-----|--|---|--|
| d.1 | Lógica de negocio: Automatizada o semiautomatizada | No es automatizada. Es en base a la experiencia. | Nueva lógica con un proceso automatizado, que se actualizará cada 3 meses. |
| d.2 | Lógica de apoyo a actividades tácticas | No | Automatizar con los modelos predictivos la toma de decisiones |
| d.3 | Procedimientos de comunicación e integración | Existe iniciativa para es certificarse en la norma ISO 55.001, que permite mejorar la toma de decisiones en los gastos del ciclo de vida de un activo productivo. Tiene como objetivo Operacional y Mantenimiento, implementar una gestión predictiva de los activos operativos al año 2024 | Mantener situación actual |
| d.4 | Lógica y procedimientos de medición de desempeño y control | Si, con la certificarse en la norma ISO 55.001 | Mantener situación actual |

4.1.5 Integración de Procesos Conexos.

Variable que estudia la cantidad de procesos y macroprocesos que interactúan con el Proyecto.

Tabla 8: Integración de Procesos Conexos.

| e | Variable de Diseño: Integración de Procesos Conexos | Actual | Propuesta |
|-----|---|--------|---|
| e.1 | Proceso aislado | No | Incorporar un nuevo proceso que tiene impacto en los procesos utilizados en Estudio y Calidad de Servicio (Macro 1) |
| e.2 | Todos o la mayor parte de los procesos de un macroproceso | No | Mantener situación actual |
| e.3 | Dos o más macros que interactúan | No | Mantener situación actual |

4.1.6 Mantención Consolidada de Estados.

Esta variable identifica las necesidades de información, que requieren los procesos involucrados con el rediseño propuesto en el Proyecto.

Tabla 9: Mantención Consolidada del Estado.

| f | Variable de Diseño: Mantención Consolidada del Estado | Actual | Propuesta |
|-----|---|--------|--|
| f.1 | Datos Propios | No | Se propone trabajar en una cultura de datos. Tener datos técnicos de los transformadores de poder y datos meteorológicos |
| f.2 | Integración con datos de otros sistemas de la empresa | Si | Trabajar con datos de diferentes Sistemas de TI y TO. |
| f.3 | Integración con datos de sistemas de otras empresas | Si | Trabajar con datos públicos disponibles en el Coordinador Eléctrico Nacional. |

4.2 Propuesta de Solución

4.2.1 Arquitectura de Procesos To Be (Patrones)

En este capítulo se propone la integración de un proceso en el área de Estudios y Calidad de Servicio. En la Figura 35 se integra un proceso que contiene los modelos que predicen las desconexiones de curso forzoso y desconexiones programadas.

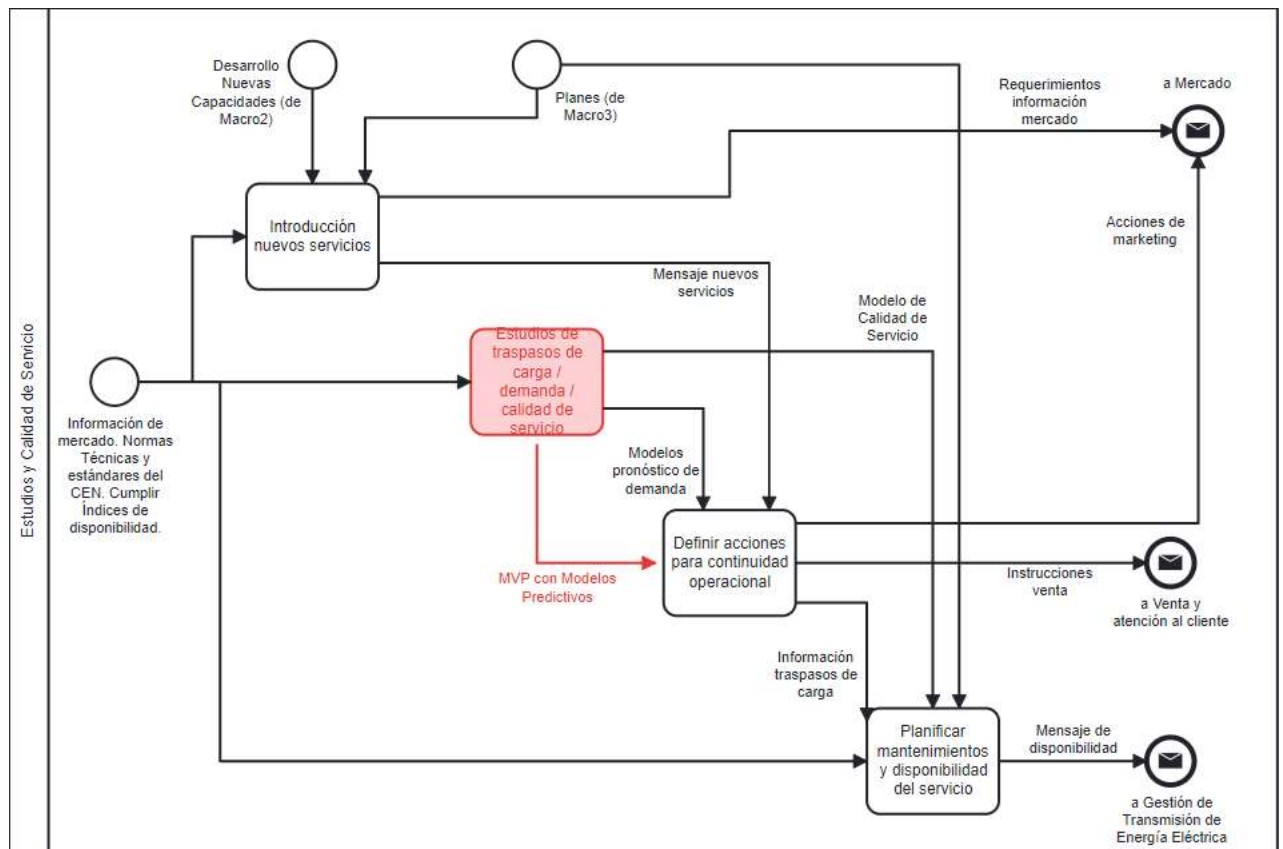


Figura 35: Proceso rediseñado integrando un modelo de predicción de mantenimientos programados y de curso forzoso

Fuente: Elaboración propia

Luego, en la apertura de “Estudios de traspaso de carga/demanda/calidad de servicio”, la estructura integra un nuevo proceso, denominado “Desarrollar y Actualizar Modelos

Predictivos”, cuyo foco es la Predicción de desconexiones de curso forzoso y la creación de mantenimientos predictivos. Este proceso es totalmente nuevo y de diseño elaborado por el autor del proyecto. La estructura del proceso rediseñado se observa en la Figura 36.

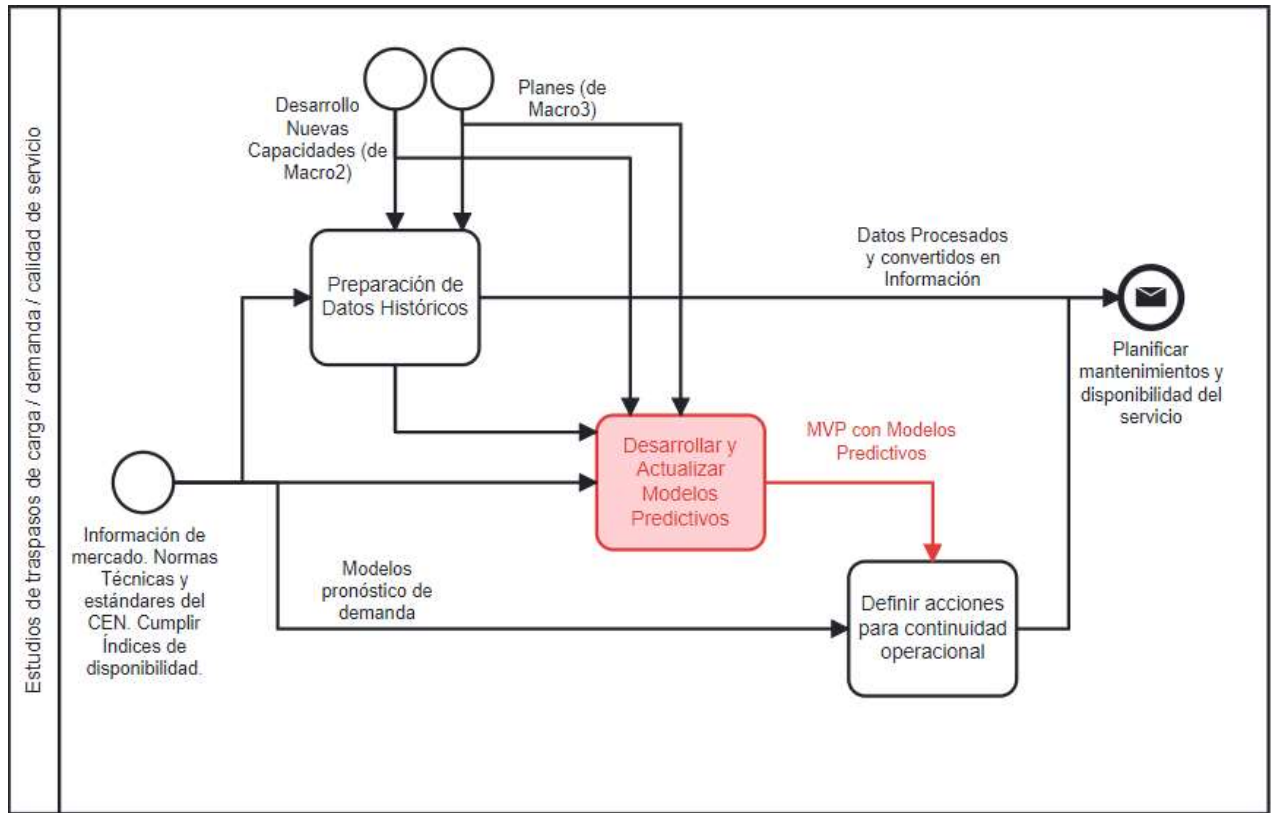


Figura 36: Desarrollar y Actualizar Modelos Predictivos, Rediseño.

Fuente: Elaboración propia

La integración del modelo permite como salida proponer un mantenimiento predictivo a los equipos eléctricos, evitando las desconexiones de curso forzoso y aportando con definir acciones para la continuidad operacional. Estas acciones apuntan a programar todos los mantenimientos con anticipación y crear planes de seccionamiento eléctrico para no afectar a los clientes en el suministro.

4.2.2 Modelamiento Detallado de Procesos To Be (BPMN)

A continuación, se detalla el proceso a integrar con el MVP. Este es un proceso de mejora continua, que permite según las variables de entrada, entrenar tres algoritmos de clasificación. Los algoritmos son de entrenamiento supervisado de clasificación, que se programan en Python. Cada modelo predice las desconexiones de curso forzoso y desconexiones programadas. Al completar el entrenamiento de los algoritmos, se conocen los parámetros de entrada y como resultado se clasifican los transformadores de poder.

En la Figura 37, se ilustra el nuevo proceso con un diagrama BPMN para mejorar el entendimiento.

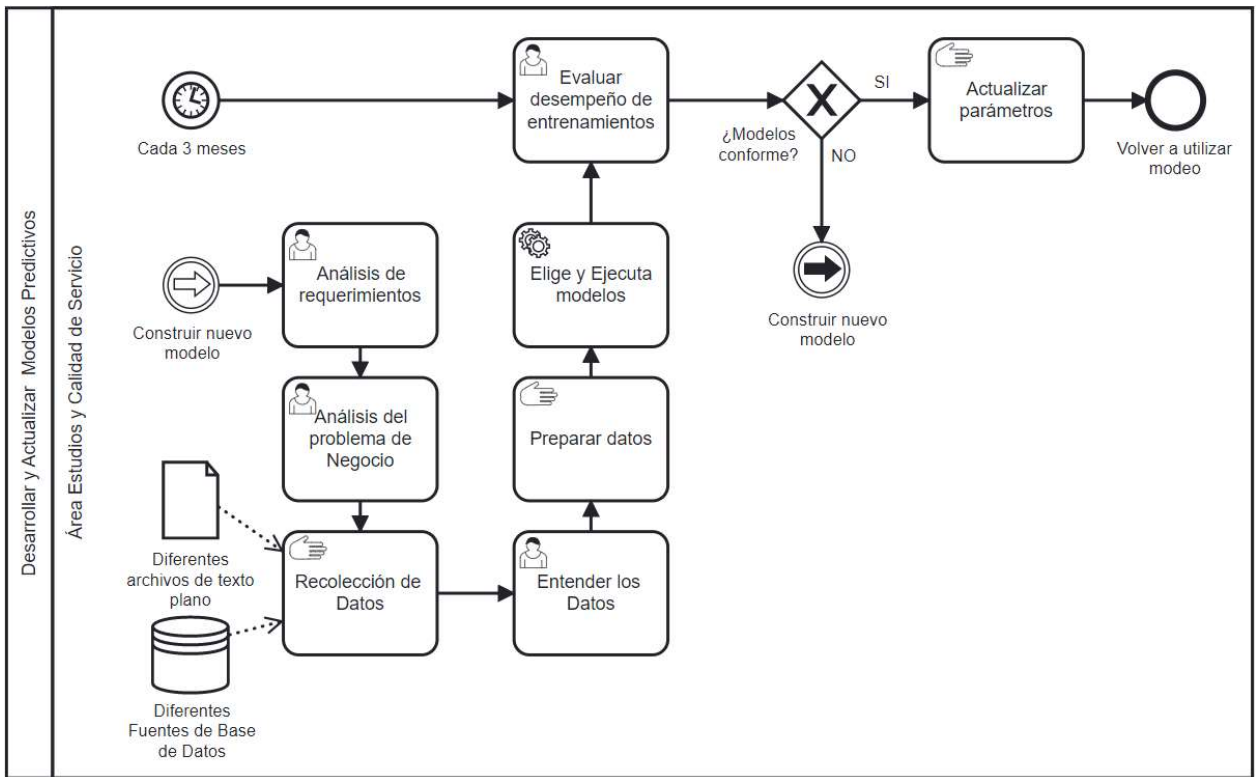


Figura 37: Proceso rediseñado integrando un modelo de predicción de desconexiones programadas y de curso forzoso

Fuente: Elaboración propia

Este nuevo proceso incorpora la variable tiempo, que se ejecuta cada tres meses. Es necesario evaluar con criterio experto, el desempeño de los algoritmos y su comportamiento con el nuevo conjunto de datos. Si el resultado no es conforme, se debe construir un nuevo modelo, utilizando la metodología CRISP-DM.

4.2.3 Diseño de Lógica de Negocios

En el diseño de Lógica de Negocios, usaremos la metodología CRISP-DM que se explicó con detalle en el capítulo 2.2. Entendemos el negocio, entendemos los datos, preparamos los datos, modelamos y visualizamos.

4.2.3.1 Entendimiento de los datos

Se utiliza un Data Set público de 21.009 filas y 37 columnas (variables), del Coordinador Eléctrico Nacional. Este contiene el registro de las desconexiones históricas de los Transformadores de Poder de las Empresas que participan en la Industria Eléctrica, desde el año 2014. (Ver Anexo C: Algoritmos de Entrenamiento Supervisado con Python).

En la Figura 38, podemos identificar las variables que contiene el set de datos.

```
In [11]: print(data.columns.values) #nombres de Las columnas
['Chequeo' 'Clave' 'ClaveEO' 'Fuente' 'ID Infotecnica' 'Tipo (2D/3D)'
'Fecha Entrada' 'Fecha Hoy' 'Edad Transformador' 'Nombre Infotecnica'
'Empresa' 'Modulo' 'Correlativo' 'Origen' 'Tipo Programación numero'
'Tipo Programación texto' 'Tipo' 'Fecha INICIO_falla' 'Fecha FIN_falla'
'Duracion_horas_falla' 'Duracion_minutos_falla' 'Mes_falla' 'Temporada'
'Fecha Inicial modificada' 'Fecha Fin Estimada o Mofificada'
'Observación' 'EO' 'Htot' 'Ffor' 'Horas sobre periodo'
'Horas bajo periodo' 'Comentario' 'Causa Presunta' 'Causa Definitiva'
'Traslape' 'Dfor' 'Dpro']
```

Figura 38: Variables del set de datos (21.009 filas y 37 columnas (variables)).

Fuente: Elaboración propia

Se analizan sus atributos: El significado de cada variable, su propósito, y tipo de datos. Junto con ello, se utilizan herramientas estadísticas descriptivas básicas y avanzadas para seleccionar variables numéricas.

Para ver la distribución de los datos se grafican las variables como muestra la Figura 39.

```
In [22]: sns.countplot(x='Tipo (2D/3D)',hue='Tipo Programación')
#La variable Tipo (2D/3D) tiene un valor unico en este
Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x19a9afd39>
```

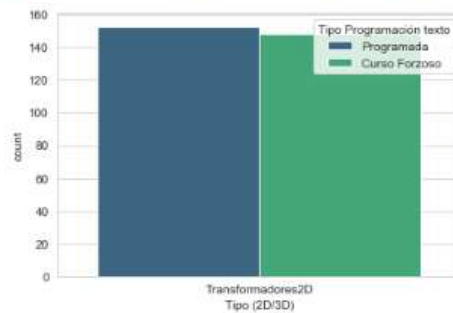


Figura 39: Identificación del Target.

Fuente: Elaboración propia

En esta etapa de entendimiento de los datos, se identifica el posible Target. Además, en los gráficos se encuentra información que puede ser utilizada para entrenar los modelos de clasificación. Por ejemplo, la temporada del año, como muestran la Figura 40 y 41.

```
In [25]: #Estudiamos variables relativas a La temporada y Tipo de Programación si es forzoso o Programada
g = sns.FacetGrid(data, col='Temporada')
g.map(plt.hist, 'Tipo Programación numero', bins=10)
```

```
Out[25]: <seaborn.axisgrid.FacetGrid at 0x267adaec788>
```

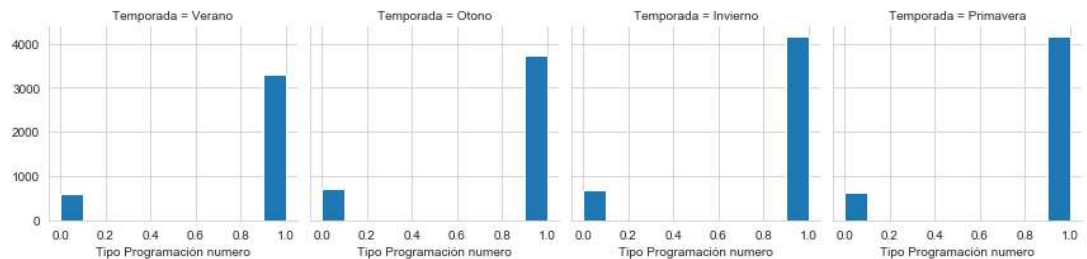


Figura 40: Grafico de variables distribuidas por temporada.

Fuente: Elaboración propia

```
In [50]: data['Temporada'].hist()  
Out[50]: <matplotlib.axes._subplots.AxesSubplot at 0x267b77785c8>
```

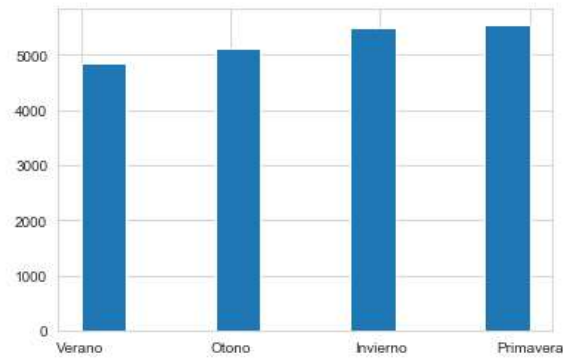


Figura 41: Grafico de variables distribuidas por temporada.

Fuente: Elaboración propia

Al mismo tiempo, en la Figura 42 se suma para entrenamiento la variable mes_falla.

```
In [51]: data['Mes_falla'].hist()  
Out[51]: <matplotlib.axes._subplots.AxesSubplot at 0x267b777cebc8>
```

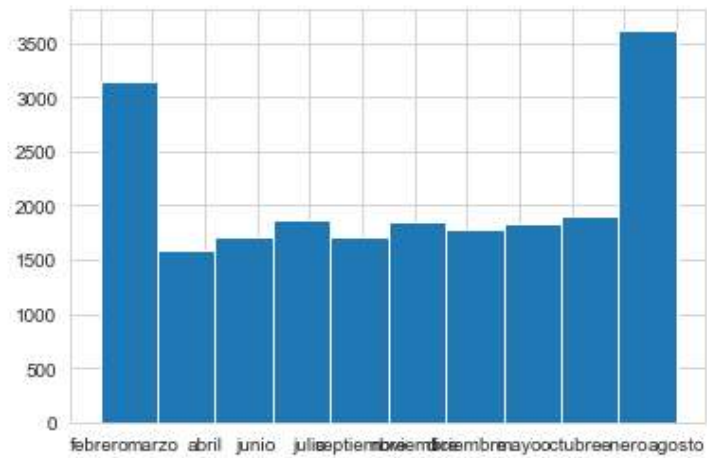


Figura 42: Grafico de variables distribuidas por meses.

Fuente: Elaboración propia

Se consideran, además, otras variables relevantes para entrenar los modelos, como son la “edad de los transformadores”, la cual se puede calcular teniendo en cuenta la fecha de PES.

Asimismo, se pueden identificar variables de importancia, como “Modulo” que separa en diferentes tipos las desconexiones. Sumado a esto, se obtiene la variable “Duración de las fallas en minutos y horas”, y de la misma manera, el “Origen”, entre otras que completan un total de 18 variables. Todas estas se explican con detalle en el “Anexo C: Algoritmos de Entrenamiento Supervisado con Python”.

4.2.3.2 Preparación de los datos

Entrando a la etapa de Limpieza, Diseño, Reducción y Preparación, se eliminan todas las variables que no se utilizan, y solo se dejan 18 columnas que son necesarias para entrenar los algoritmos de clasificación. En la Figura 43 se identifican las variables a utilizar y el tipo de dato.

```
In [60]: print(data.dtypes) #tipos de datos
```

| | |
|--------------------------|---------|
| ClaveEO | object |
| ID Infotecnica | int64 |
| Tipo (2D/3D) | object |
| Edad Transformador | object |
| Modulo | object |
| Origen | object |
| Tipo Programación numero | float64 |
| Tipo | object |
| Duracion_horas_falla | int64 |
| Duracion_minutos_falla | int64 |
| Mes_falla | object |
| Temporada | object |
| EO | object |
| Comentario | object |
| Causa Presunta | object |
| Causa Definitiva | object |
| Dfor | int64 |
| Dpro | int64 |
| dtype: | object |

Figura 43: Variables seleccionadas para el entrenamiento.

Fuente: Elaboración propia

Existen variables numéricas que, si se requieren para entrenar los algoritmos, pero estas contienen datos vacíos o nulos (NaN) y para no eliminarlos se reemplazan con la moda. Por otro lado, como se aprecia en la Figura 44, se aprecian en la gráfica demasiados datos vacíos, en dos columnas: “Causa presunta” y “Causa definitiva”. Por lo tanto, se toma la decisión de eliminar estos campos, ya que no es posible reemplazarlos, porque esas columnas son de texto no estructurado y requieren un tratamiento más complejo de minería de datos.



Figura 44: Muestra gráficamente que se limpian y no hay espacios vacíos en los datos.

Fuente: Elaboración propia

En esta etapa se transforman las variables antes seleccionadas a numéricas, ya que los modelos de Machine Learning a utilizar no aceptan datos tipo String (Object), Figura 45.

```

Data columns (total 43 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   ID Infotecnica                             21009 non-null  int64
1   Edad Transformador                          21009 non-null  float64
2   Tipo Programación numero                   21009 non-null  float64
3   Duracion_horas_falla                       21009 non-null  int64
4   Duracion_minutos_falla                    21009 non-null  int64
5   DFor                                       21009 non-null  int64
6   Dpro                                       21009 non-null  int64
7   Baja Reserva Primaria (SING)               21009 non-null  int64
8   Desconexión / Intervención                 21009 non-null  int64
9   Informe Fallas                             21009 non-null  int64
10  Limitaciones                               21009 non-null  int64
11  Otro                                        21009 non-null  int64
12  Prueba Operacional                         21009 non-null  int64
13  Pruebas de Puesta en Servicio              21009 non-null  int64
14  Invierno                                    21009 non-null  int64
15  Otono                                       21009 non-null  int64
16  Primavera                                   21009 non-null  int64
17  Verano                                       21009 non-null  int64
18  abril                                       21009 non-null  int64
19  agosto                                       21009 non-null  int64
20  diciembre                                   21009 non-null  int64
21  enero                                       21009 non-null  int64
22  febrero                                       21009 non-null  int64
23  julio                                       21009 non-null  int64
24  junio                                       21009 non-null  int64
25  marzo                                       21009 non-null  int64
26  mayo                                       21009 non-null  int64
27  noviembre                                   21009 non-null  int64
28  octubre                                       21009 non-null  int64
29  septiembre                                   21009 non-null  int64
30  DF                                           21009 non-null  int64
31  DN                                           21009 non-null  int64
32  DP                                           21009 non-null  int64
33  DP_E                                         21009 non-null  int64
34  DRO                                          21009 non-null  int64
35  FE                                           21009 non-null  int64
36  P                                             21009 non-null  int64
37  PO                                           21009 non-null  int64
38  RO                                           21009 non-null  int64
39  externo                                       21009 non-null  int64
40  interno                                       21009 non-null  int64
41  Desconexión                                 21009 non-null  int64
42  Intervención                                 21009 non-null  int64
dtypes: float64(2), int64(41)
memory usage: 6.9 MB

```

Figura 45: Variables seleccionadas y transformadas a numéricas para entrenar los modelos de predicción.

Fuente: Elaboración propia

En la figura anterior se puede apreciar que, al convertir las 18 columnas seleccionadas, pasan a aumentar su cantidad a 43 variables numéricas, las cuales servirán para entrenar los modelos. Además, cabe destacar que la edad del transformador, la duración de las desconexiones, la temporada del año, el mes, el tipo de programación, el origen y el módulo, son datos que aportan a la clasificación del *target*.

4.3 Resultados Obtenidos

4.3.1 Modelamiento de los datos

Según el criterio experto de Don Jorge Muñoz (ex- Gerente de Mantenimiento y actual Asesor externo de la empresa) y la retroalimentación de Italo Chiarella (Ingeniero de Estudio), las variables de entrada no sirven para predecir fallas de los Trafos, ya que no existen variables técnicas fundamentales para conocer el estado de salud de un transformador, como son los niveles de aceite, temperatura del aceite, temperatura del Trafo, cantidad de operaciones de CTBC (Cambiador Tomas Bajo Carga), análisis de gases y estudios de polimerización del papel dentro de un transformador. De lo anterior expuesto y como se ha explicado a lo largo de esta Tesis, se utilizará entrenamiento supervisado con Machine Learning utilizando otras variables que permiten predecir y disminuir las desconexiones de curso forzoso, que tienen un alto costo y afectan al índice SAIDI.

Con el lenguaje de programación Python, obtenemos un MVP en donde se modelan tres algoritmos de entrenamiento supervisado. Como se explicó en el Capítulo 2, Marco teórico, se utilizan KNN, NB y RF. Se obtienen los primeros resultados y luego se realizan 5 iteraciones, cambiando los hiperparámetros, con Grid Search y Random Search. También se modifica desde la tercera iteración la selección de atributos, con lo cual se logran mejores resultados y que estos no se encuentren sobre ajustados.

4.3.1.1 Resultados KNN, NB, RF

De todo lo anterior expuesto, primero se identifica el target y hacen correr los tres algoritmos con el *dataset* ya preparado:

- Target = Predecir
 - Desconexiones Programada = 1
 - Desconexiones de Curso Forzoso = 0

El *dataset* utilizado para entrenar los modelos es rescatado del Informe de fallas entregado por el Sistema NeoMante, publicado por el Coordinador Eléctrico Nacional. Considera el universo de todos los transformadores de las empresas eléctricas y mineras a nivel nacional, registrando las fallas entre los años 2014 y 2022. Cuenta con 27.000 filas.

Como se mencionó anteriormente, los mejores resultados fueron en la tercera iteración, con una selección de atributo con método embebido, y el mejor algoritmo fue RF. ¿Por qué es mejor el resultado del RF, en la tercera iteración? Porque logra reducir los Falsos Positivos (FP), que es fundamental disminuir. En la Tabla 10 se explican los resultados de TP, FP, TN y FN.

Tabla 10: Entradas y salidas obtenidas por los algoritmos.

| Algoritmos Clasificando | | | | |
|-----------------------------|----------------|---------|-----------|--------|
| Desconexión Programada = | | | | 1 |
| Desconexión Curso Forzoso = | | | | 0 |
| TP | True Positive | Entrada | Clasifica | Salida |
| | | 1 | | 1 |
| FP | False Positive | Entrada | Clasifica | Salida |
| | | 0 | | 1 |
| TN | True Negative | Entrada | Clasifica | Salida |
| | | 0 | | 0 |
| FN | False Negative | Entrada | Clasifica | Salida |
| | | 1 | | 0 |

Es posible determinar después de varios análisis, que para este modelo es muy importante reducir el FP, ya que si el resultado no identifica la desconexión de curso forzoso, ya que si el resultado no identifica la desconexión de curso forzoso, el mantenimiento del Trafo se programa demasiado tarde, por lo que no alcanza a evitar la desconexión de curso forzoso. Por ejemplo, un Trafo tiene programado su mantenimiento para octubre del próximo año y este debe ser adelantado 15 meses para evitar una desconexión de curso forzoso, la cual es hasta 10 veces más costoso.

En el “Anexo C: Algoritmos de Entrenamiento Supervisado con Python” se explican cada una de las iteraciones y como fueron cambiando los resultados según se modificaban la selección de atributos y los hiperparámetros con Grid Search y Random Search.

La matriz de confusión puede ayudarnos a tomar la determinación correcta. La tabla 11 ordena la matriz (2x2) de confusión con los TP, TN, FP y FN.

Tabla 11: Matriz de confusión 2x2. TP, TN, FP y FN

| | Condición Actual es +V | Condición Actual es -V |
|----------------------------|------------------------|------------------------|
| Condición Predictiva es +V | TP | FP |
| Condición Predictiva es -V | FN | TN |

La Figura 46 muestra el resultado de los tres algoritmos en donde RF obtiene un mejor desempeño, reduciendo los FP.

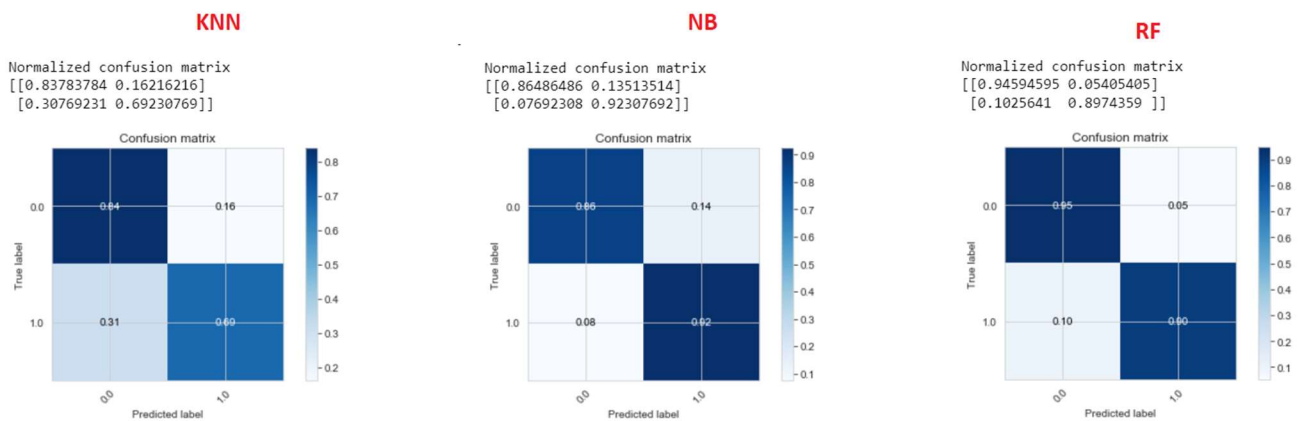


Figura 46: Matriz de confusión con los resultados de los algoritmos en la tercera iteración.

Fuente: Elaboración propia

La siguiente tabla 12, se construye para comparar los resultados de todas las iteraciones y se destaca en rojo la tercera iteración, ya que los FP obtuvieron el menor valor porcentual.

Tabla 12: Comparación matriz de confusión entre la segunda, tercera, cuarta y quinta iteración.

| Comparación de Matriz de confusión | | | | |
|------------------------------------|----|----|----|-----|
| KNN | TP | FP | FN | TN |
| Segunda Iteración | 85 | 15 | 21 | 79 |
| Tercera Iteración | 84 | 16 | 31 | 69 |
| Cuarta Iteración | 84 | 16 | 31 | 69 |
| Quinta iteración | 84 | 16 | 31 | 69 |
| NB | TP | FP | FN | TN |
| Segunda Iteración | 89 | 11 | 0 | 100 |
| Tercera Iteración | 86 | 14 | 8 | 92 |
| Cuarta Iteración | 86 | 14 | 8 | 92 |
| Quinta iteración | 86 | 14 | 8 | 92 |
| RF | TP | FP | FN | TN |
| Segunda Iteración | 85 | 15 | 0 | 100 |
| Tercera Iteración | 95 | 5 | 10 | 90 |
| Cuarta Iteración | 92 | 8 | 8 | 92 |
| Quinta iteración | 89 | 11 | 5 | 95 |

Para explicar a cabalidad por qué los resultados de la tercera iteración son mejores, a continuación, se ejemplifica asignando costos a los FP y FN.

Suponemos que:

- $X = \text{Costo mantenimiento} = 1$
- Desconexión de curso forzoso = hasta 10 veces más costosa que una desconexión programada.
 - Costo Total = $(10X * \text{cambio en FP\%}) + (X * \text{cambio en FN})$
 - Segunda iteración = $(10 * 0,15) + (1 * 0,0) = 1,5$
 - **Tercera iteración = $(10 * 0,05) + (1 * 0,10) = 0,6$**
 - Cuarta iteración = $(10 * 0,08) + (1 * 0,08) = 0,88$
 - Quinta iteración = $(10 * 0,11) + (1 * 0,05) = 1,15$

Las gráficas de la Figura 47, son fundamentales para determinar cuál algoritmo obtiene los mejores resultados. En ella, se aprecia que las curvas de ROC se encuentran sin sobreajustes. Sobre todo, mejoran los algoritmos NB y RF.

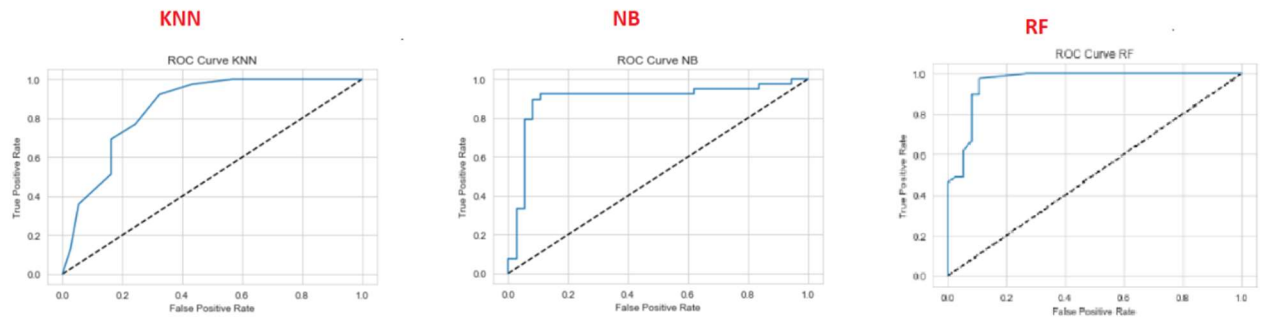


Figura 47: Curva de ROC de los algoritmos en la tercera iteración.

Fuente: Elaboración propia

En resumen, la tercera iteración obtiene los mejores resultados. En la tabla 12 se observa una reducción de los Falsos Positivos (FP) y además en la ecuación utilizada para interpretar la matriz de confusión, se obtienen costos muchos más bajos.

Como se mencionó, según el juicio experto es muy importante reducir los FP, porque si el resultado no se anticipa a una desconexión de curso forzoso, se actuaría tarde con el mantenimiento del Trafo, y una desconexión de curso forzoso es hasta 10 veces más costosa.

4.4 Retroalimentación juicio experto

En este capítulo, se expone la retroalimentación con juicio experto. Con entrevistas y en conjunto con los profesionales del área de estudios y calidad de servicio, se comparan los resultados con los objetivos planteados. En la siguiente Tabla 13 se detallan los resultados del ejercicio con los colaboradores.

Tabla 13: Retroalimentación Juicio Experto.

| Objetivos | Retroalimentación juicio experto |
|---|--|
| Objetivo General | |
| <p>“Rediseñar el proceso de estudios y calidad de servicio, integrando un modelo de predicción para las desconexiones de curso forzoso en los Transformadores de Poder”</p> | <p>El área de estudios y calidad de servicio evaluó y aceptó el nuevo proceso que se ejecuta cada 3 meses. Además, ejecuta y prueba MVP entregado en Jupyter notebook con éxito.</p> |
| Objetivos Específicos | |
| <p>Proponer un rediseño a los procesos existentes, o diseñar un proceso faltante en los estudios de calidad de servicio, utilizando modelos predictivos.</p> | <p>El área de estudios y calidad de servicio evaluó y aceptó el nuevo proceso que se ejecuta cada 3 meses.</p> |
| <p>Levantar requerimientos, definir roles y analizar los procesos involucrados para solucionar problemas con Data Analytics.</p> | <p>El área asume el desafío de capacitar a sus profesionales y contratar nuevos especialistas para cumplir con las etapas, roles y funciones definidas para el rediseño del proceso, que permitan desarrollar nuevas capacidades de Data Analytics.</p> |
| <p>Proponer un modelo predictivo con la metodología CRISP-DM para identificar las variables de entrada que necesita el algoritmo de predicción.</p> | <p>Durante el desarrollo de este proyecto se trabajó con profesionales de diversas áreas de estudio, mantenimiento y operación. Se analizaron e identificaron las variables necesarias para conocer el estado de salud de un transformador. Concluyendo que la empresa debe trabajar en una cultura de datos ya que los datos que son imprescindibles para entrenar los algoritmos no se encuentran disponibles dentro de la organización.</p> |

| | |
|---|--|
| <p>Programar algoritmos de Machine Learning para predecir las desconexiones de curso forzoso.</p> | <p>El Producto Mínimo Viable (MVP) se presenta en la empresa siendo aprobado por el área de estudios y calidad de servicios. Los entrenamientos y sus resultados anticiparon el mantenimiento al Trafo que falló en el 2020, por lo tanto, el índice SAIDI hubiese cumplido la meta para ese año. Además, los COMA hubiesen sido más bajos que el logrado, aportando a un mejor resultado del EBITDA</p> |
|---|--|

CAPÍTULO 5. PROPUESTA DE APOYO TECNOLÓGICO

5.1 Arquitectura Tecnológica

El Proyecto está basado en el desarrollo de modelos de Machine Learning. Se integra un modelo de predicción a los actuales procesos de Estudios y Calidad de Servicios, pasando a ser parte de las soluciones tecnológicas disponibles en la organización.

La arquitectura considera entregar el entorno de desarrollo Jupyter Lab como un MVP, el cual está basado en la *web*. Se presenta en un Notebook de trabajo, el código, los datos tabulados y gráficos amigables para el usuario final. El lenguaje de programación es Python y contiene variadas librerías para correr algoritmos de entrenamiento supervisado y no supervisado para trabajos en *data analytics*.

La arquitectura se divide en tres componentes:

- i. La fuente de los datos.
- ii. Un computador con capacidades suficientes para correr los modelos de predicción entregados.
- iii. La visualización de los resultados de los modelos con su *accuracy*, *precision*, *recall*, *f1-score* y *support*. Además, se visualiza en una gráfica la matriz de confusión y curva de ROC.

Se esquematizan en la Figura 48 los elementos que permiten comprender el entorno que rodea a la solución tecnológica.

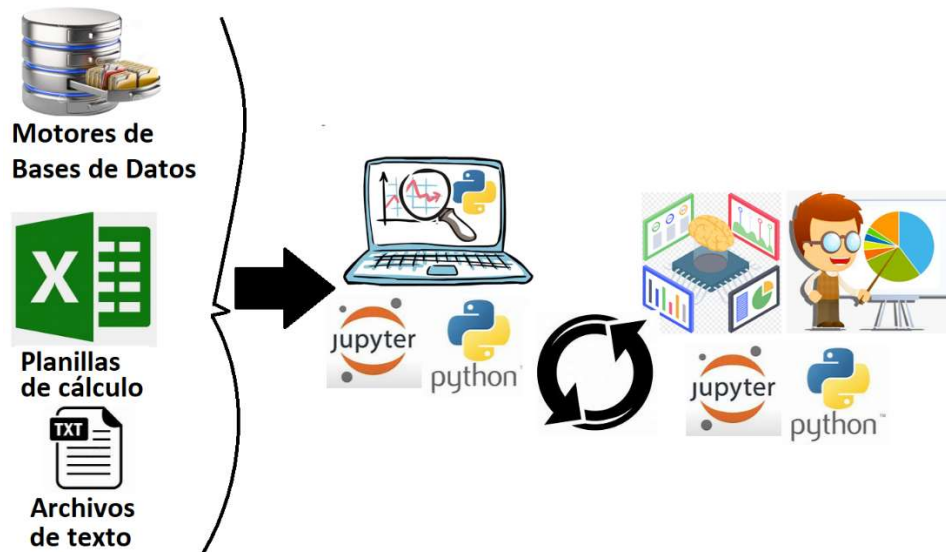


Figura 48: Arquitectura de Apoyo Tecnológico.

Fuente: Elaboración propia

5.2 Visualización de los datos

Tal como se presentan gráficamente los resultados en el capítulo de 4.2.1.4, la visualización de los modelos y todo el análisis de los datos es entregada en un Notebook de Jupyter, programado en Python.

En la siguiente Figura 49, se ilustran varios gráficos que ayudan en todo el proceso de la metodología aplicada en este trabajo de tesis. Gracias a estos, se pueden entender los datos de manera amigable para el usuario.

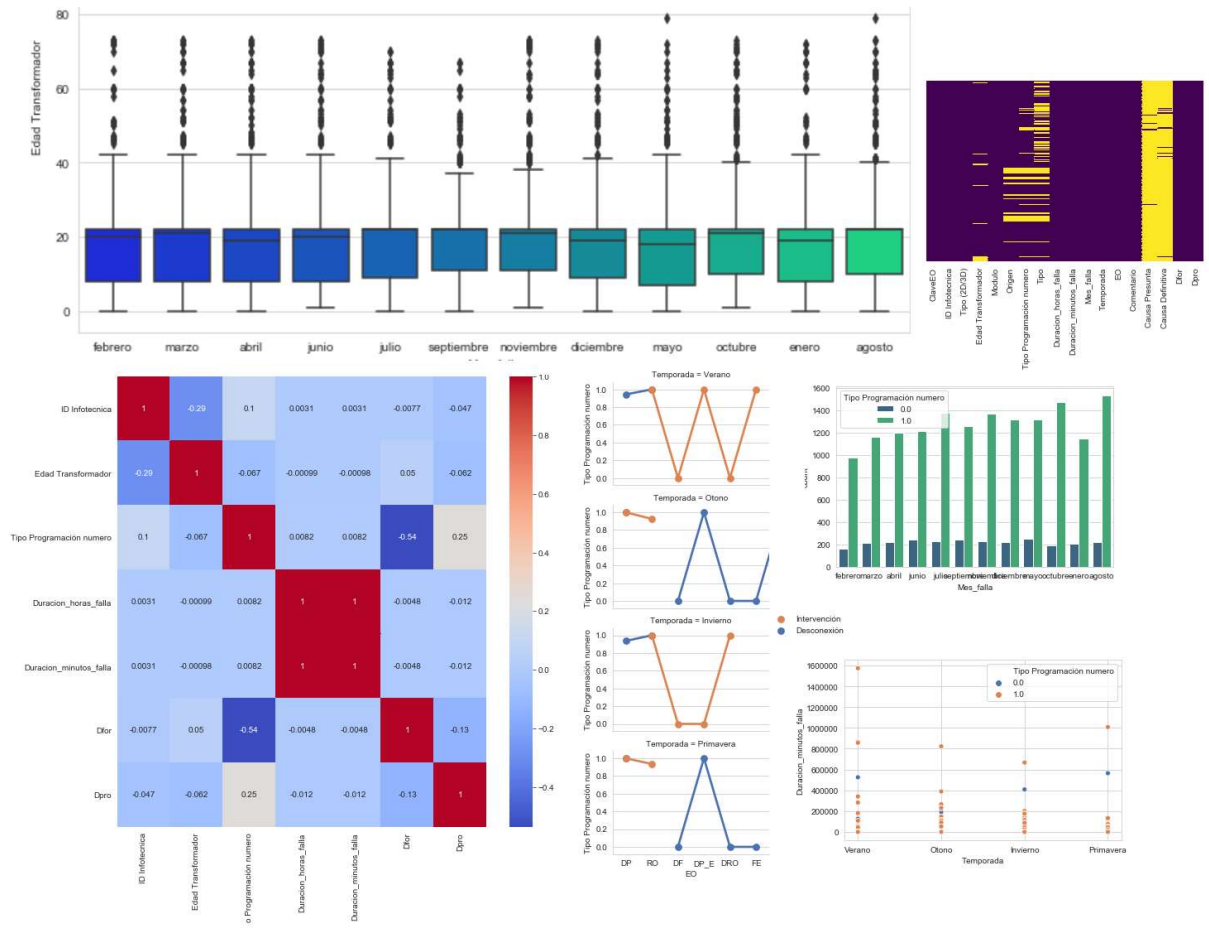


Figura 49: Visualización grafica del Jupyter Notebook.

Fuente: Elaboración propia

Los resultados de la figura anterior presentan distintos tipos de gráficos, que ayudan a inferir la distribución de los datos. Los gráficos que se presentan en la siguiente figura, son importantes para comprender los datos y seleccionar las variables que se utilizan en el entrenamiento de los modelos.

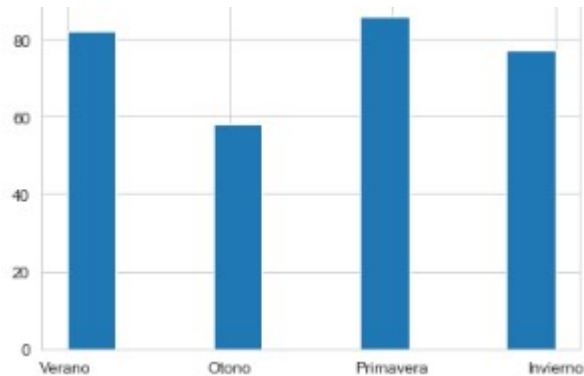


Figura 50: Gráfico de barras.

Fuente: Elaboración propia

En la Figura 50, se analiza el *dataset* con gráficos de barras, que muestran la relación entre cada valor tomado por la variable y la frecuencia de los datos en la temporada del año. Los distintos valores tomados por la variable “Temporada” están en el eje de las abscisas (eje x) y en cada valor se dibuja una barra de altura, igual a la frecuencia correspondiente. La altura de la barra aparece en la ordenada (eje y).

Para obtener la frecuencia absoluta de los datos en los meses, se usan los gráficos de histogramas formando intervalos en los datos, tal como se ilustra en la Figura 40.

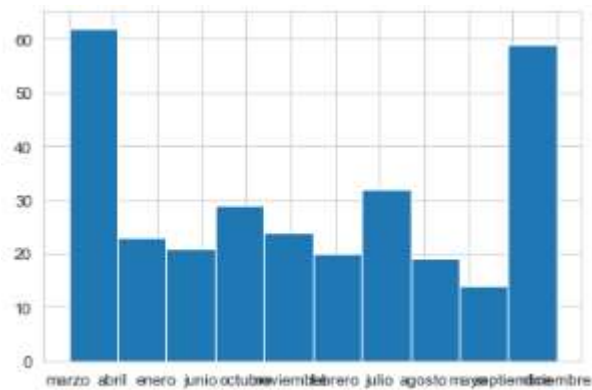


Figura 51: Gráfico de histograma.

Fuente: Elaboración propia

Como se mencionó anteriormente, se utilizan también gráficos de caja (box). Estos permiten comparar de manera simple varias distribuciones al mismo tiempo. En la Figura 52, que se muestra a continuación, se comparan la edad del transformador y el mes en donde ocurrieron las desconexiones. El 25% es el primer cuartil, 50% segundo cuartil (o mediana) y 75% tercer cuartil.

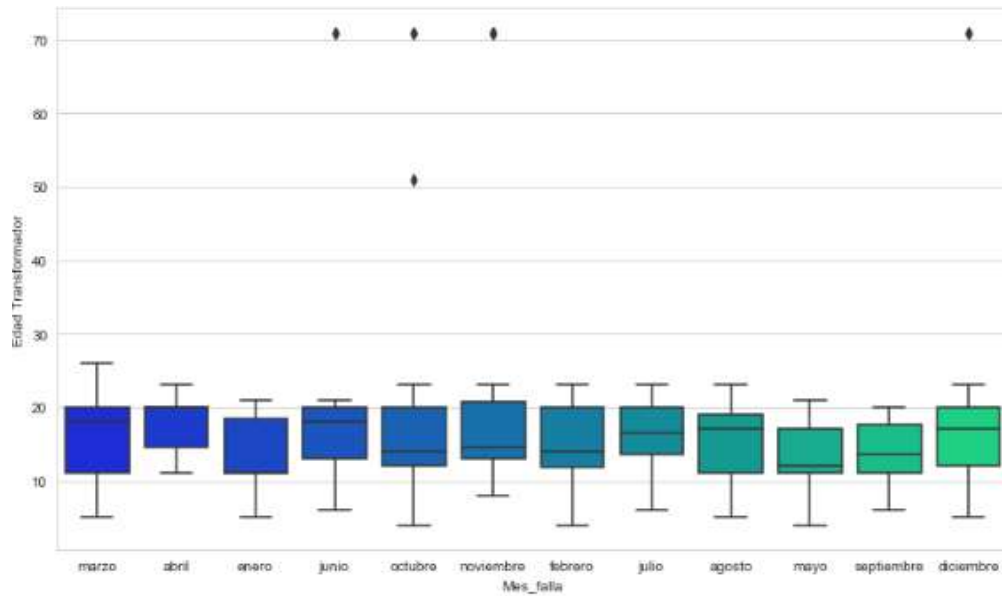


Figura 52: Gráfico de caja

Fuente: Elaboración propia

Se trabaja también con gráficos de dispersión, como se muestra en la siguiente figura:

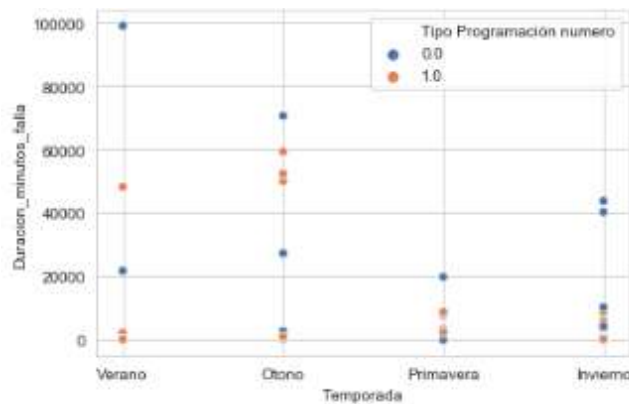


Figura 53: Gráfico de dispersión

Fuente: Elaboración propia

En la Figura 53, los gráficos de dispersión utilizan las coordenadas cartesianas para mostrar los valores de dos variables, temporada del año y la duración en minutos de las desconexiones.

Para finalizar de seleccionar los datos que entran en los modelos de machine learning, se programa una matriz de correlación entre los datos (Figura 54), porque esto permite ver el impacto que tienen unas variables sobre otras. Mientras más se acercan a la diagonal, mayor es la correlación entre los datos.

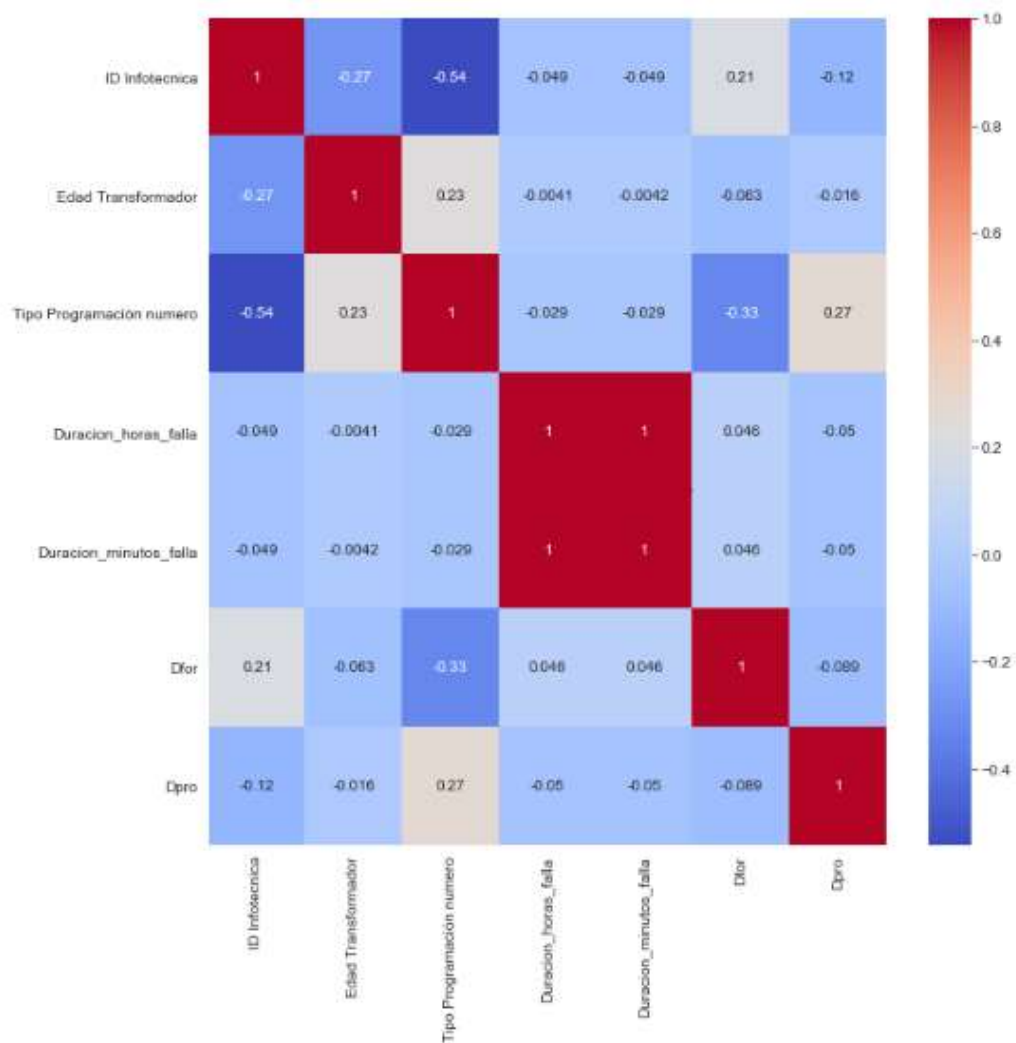


Figura 54: Matriz de correlación

Fuente: Elaboración propia

CAPÍTULO 6. PROPUESTA DE IMPLEMENTACIÓN

6.1 Contexto de la Empresa

STS es una empresa cuya función es relevante para el país, dado que se encarga de transmitir y transformar la energía eléctrica por grandes distancias en el territorio nacional, desde Antofagasta hasta Chiloé. Dentro de la organización, la Gerencia de Operaciones, tiene la responsabilidad de entregar al cliente continuidad del servicio eléctrico, sin cortes, y entregar calidad en los niveles de voltajes. En esta etapa del proyecto de tesis, es fundamental comunicar a la empresa y sus trabajadores para qué sirven los modelos predictivos, para así generar la aprobación, compromiso y motivar a los usuarios dentro de la compañía.

Una posible desventaja en la compañía es no poder predecir las fallas. En los últimos años y gracias al aumento de mantenimientos preventivos, inspecciones visuales, entre otras iniciativas, se han mejorado los índices de disponibilidad en forma considerable. A pesar de todas estas iniciativas, la empresa, según su estrategia de negocio, estima necesario continuar proponiendo soluciones innovadoras, como lo es la predicción. Con esta visión, es necesario la adopción de nuevas capacidades en los profesionales, como también la adopción de metodologías de trabajo con enfoque en Data Analytics, la cual tiene una manera de pensar distinta a la actual forma de trabajo reactivo y preventivo. Se trata de anticiparse a los problemas analizando casos de uso que entreguen valor, entender el negocio, seleccionar los datos y los algoritmos a modelar.

Como se expone en el párrafo anterior. La tesis propone una nueva metodología, que afecta la forma de trabajar y pensar en la organización, incorporando técnicas de Data Analytics para los problemas que afectan a la empresa, y que en este caso tienen relación con el estado de salud de los transformadores de poder. Esta incorporación de técnicas, atenta contra los actuales paradigmas en los cuales se basan parte de los estudios y mantenimientos programados por calendario. Por lo tanto, es necesario realizar un plan de gestión del cambio que permita a los colaboradores incorporar de manera exitosa estas nuevas prácticas.

6.2 Análisis de los Principios de Diseño

La base de este capítulo es el Modelo CHESS (Conciencia, Holístico, Experiencial, Sistémico y Sustentable), el cual es un modelo Integral de Liderazgo y Gestión del Cambio (Olguín, Crawford, & Soto, 2016). Este modelo enfatiza en las organizaciones la toma de conciencia por el cuidado de las personas y el impacto que tienen las empresas en la construcción de un futuro con sustentabilidad, incluyendo en su concepción una preocupación humana, medio ambiental, social y económica. El modelo comprende que la organización, las empresas y la sociedad son parte de un todo.

En los siguientes apartados se analizan 4 puntos del modelo, que se consideran imprescindibles para el éxito de este proyecto de tesis, en otras palabras, son puntos clave para integrar un modelo predictivo utilizando metodologías de Data Analytics.

6.2.1 Sentido y estrategia del proceso de cambio

Este proyecto de tesis, desarrollado por Giovanni Noceti Guzmán, es un trabajo que considera la participación de diferentes *stakeholders* dentro de la organización. Se utilizan dos metodologías de trabajo. Primero la de Ingeniería de negocio para el rediseño del proceso y, segundo, la metodología CRISP-DM para los modelos predictivos programados en Python.

Los profesionales fueron involucrados desde el principio, para que el Proyecto tenga sentido dentro de la empresa. Comunicar para qué sirve, generar la aprobación, compromiso y motivación necesaria para tener aprecio por el nuevo proceso dentro de la organización. Los colaboradores estuvieron presentes al analizar requerimientos, encontrar el problema, colaborar con los objetivos, comparar resultados y estuvieron presentes en las reuniones de avance, lo cual fue muy importante. Además, proporcionaron los datos para entrenar los modelos y costos requeridos para el logro del trabajo final.

6.2.2 Cambio y conservación

Actualmente en la empresa no existe una metodología para trabajar con modelos predictivos y aquí es donde comienzan los cambios en la organización. La forma de pensar, trabajar y almacenar los datos requiere un cambio. La compañía debe adoptar una nueva cultura de datos, donde estos deben ser almacenados históricamente, con nuevas variables y nuevos formatos para ser tratados en los modelos de Machine Learning entregados en esta tesis.

Del mismo modo que existe un cambio, se deben conservar todas las buenas iniciativas y prácticas que han dado buenos resultados a lo largo de la historia de STS. Lo que se conserva incluye los estudios, a los profesionales, las instalaciones, forma de relacionarse, y los privilegios del teletrabajo. Es muy importante aclarar que no hay despidos.

El cambio se aborda en base a integrar un modelo predictivo, que trabaja con metodología CRISP-DM y tres algoritmos de Machine Learning. Esto conlleva a pensar distinto, aprender nuevas habilidades y anticiparse a los problemas.

6.2.3 Gestión de las comunicaciones

El proyecto ha generado diversas instancias de comunicación y conversación, en donde los diferentes *stakeholders* han podido dar sus puntos de vista, orientar el trabajo y ser considerados dentro su desarrollo. Por lo tanto, el interés de los actores en realizar este proyecto sí ha sido considerado y revisado. Además, con el propósito de controlar las expectativas, se han explicado los alcances y exclusiones, es decir, lo que hace y no hace.

6.2.4 Gestión del aprendizaje y las habilidades

En este capítulo se aborda el proceso de aprendizaje necesario para usar el modelo de predicción y la metodología de trabajo CRISP-DM. Para aplicar la nueva metodología y

utilizar los modelos de Machine Learning es necesario aprender a pensar de forma distinta. Es también necesario desarrollar nuevas habilidades, como entender el negocio, analizar y depurar datos, aprender lenguajes de programación, potenciar las estadísticas y probabilidades, adquirir habilidades comunicativas para explicar los datos correctamente, ejercitar la mentalidad resolutiva y, finalmente, es clave desarrollar la capacidad de trabajar con equipos de profesionales multidisciplinarios.

Este proyecto se considera como la base para futuros trabajos relacionados con Machine Learning. Se puede decir, que es “el punta pie inicial”, para generar más estudios en torno a modelos predictivos y, en consecuencia, potenciar una nueva cultura de datos en la organización, en donde estos sean usados para abarcar y resolver problemáticas de la compañía.

CAPÍTULO 7. EVALUACIÓN DEL PROYECTO

Dentro del presente capítulo se desarrolla la evaluación económica del proyecto de tesis. En este se aborda el tipo de evaluación, se detalla también el escenario de la empresa con y sin proyecto, con sus respectivos resultados. Además, se construyen y desarrollan los costos del proyecto, sus beneficios, finalizando con el flujo de caja.

7.1 Producto Mínimo Viable (MVP)

Con el objetivo de obtener resultados reales, se entrenan los modelos predictivos con datos desde el 2014 al 2019 y se realiza una comparación de los mantenimientos programados y de curso forzoso, ocurridos en el año 2020.

7.1.1 Definición del MVP

La selección de los datos para el MVP se realiza en base a todas las desconexiones ocurridas en las empresas eléctricas a nivel nacional. El entrenamiento de los algoritmos es en base a las desconexiones programadas y de curso forzoso. Estas últimas contienen las fallas de los transformadores de poder.

Debido a la falta de datos dentro de la compañía, para entrenar a los algoritmos, es que se utiliza un *dataset* público del Coordinador Eléctrico Nacional, el cual contiene las desconexiones de los transformadores de la industria eléctrica, con un total de 27.000 registros.

7.1.2 Resultados Obtenidos

El MVP permite ingresar transformadores al modelo y obtener un plan de mantenimiento predictivo.

Por otro lado, se realiza una comparación de costos de los mantenimientos preventivos y correctivos. En el año 2020, por una falla en uno de los transformadores de poder, los costos de ese año triplicaron el presupuesto en comparación con los años 2018, 2019 y 2021, en conjunto, incluyendo el mantenimiento de los 136 transformadores de poder de la compañía. Como se aprecia en la Tabla 14, los mantenimientos se realizan cada año y solo con una falla se superó el presupuesto total de 3 años.

Tabla 14: Costos de mantenimientos TTFF y falla ocurrida el 2020

| ERP SAP PM Mto. Tx. | |
|------------------------------------|----------------------|
| Mantenimiento para los TTFF | Monto en \$ |
| 2019 | \$125.051.943 |
| 2020 | \$71.928.155 |
| 2021 | \$120.329.971 |
| 2022 | \$34.886.200 |
| Total | \$352.196.269 |
| Mantenimientos del T44 | Monto en \$ |
| 2018 al 2020 | \$1.954.027 |
| Falla del TTFF T44 2020 | Monto en \$ |
| 2020 | \$889.883.760 |
| 2021 | \$10.239.509 |
| Total | \$900.123.269 |

7.2 Definición de Variables y Criterios de Evaluación

7.2.1 Tipo de Evaluación

El principal ingreso de STS se debe al cobro de peajes por el uso de las líneas de transmisión y además por los servicios de construcción, mantenimiento y operación en tiempo real de las instalaciones eléctricas de sus clientes. El tipo de evaluación se focaliza en los costos en que incurre STS, y la disminución de estos evitando las desconexiones de curso forzoso, las cuales son provocadas por fallas en los equipos.

Como muestra la Tabla 15, los costos se deben en su mayoría al mantenimiento de las instalaciones y equipos eléctricos. (Deloitte, Auditores y Consultores Limitada, 2020 y 2019).

Tabla 15: Costos de STS

| Costos | 2019 | 2020 | 2021 |
|---|-----------------|-----------------|------------------|
| Servicios compartidos | \$ - | \$ - | \$3.902.949.000 |
| Operación y mantenimiento del sistema eléctrico | \$2.738.954.000 | \$2.280.178.000 | \$4.194.176.000 |
| Operación de vehículos, viajes y viáticos | \$167.947.000 | \$158.621.000 | \$203.361.000 |
| Arriendo maquinarias, equipos e instalaciones | \$2.192.000 | \$300.244.000 | \$107.666.000 |
| Provisiones y castigos | \$-1.484.000 | \$-135.743.000 | \$-5.992.000 |
| Gastos de administración | \$1.682.022.000 | \$1.681.226.000 | \$2.045.603.000 |
| Otros gastos por naturaleza | \$304.747.000 | \$426.290.000 | \$319.940.000 |
| Egresos por construcción de obras a terceros | \$ - | \$ - | \$261.750.000 |
| | \$4.894.378.000 | \$4.982.302.000 | \$11.029.453.000 |

7.2.2 Horizonte de Evaluación

Debido a la información entregada por STS, se decide trabajar con un horizonte de evaluación de 3 años.

7.2.3 Tasa de Descuento

Esta evaluación es privada, por lo tanto, no se incorporan los beneficios sociales, pero sí se aprecia una perspectiva de beneficio social, ya que, al evitar las fallas, los clientes obtienen beneficios en la calidad y disponibilidad del servicio eléctrico.

Debido a esto, se utiliza la tasa de descuento indicada en la Ley de Transmisión del 20 de julio del 2016, que se publicó en el Diario Oficial (Ley N° 20.936), que establece un Nuevo Sistema de Transmisión Eléctrica y crea un Organismo Coordinador Independiente del Sistema Eléctrico Nacional. La tasa de descuento utilizada para remunerar los costos de las instalaciones cambia desde una tasa fija real anual de 10% antes de impuestos, a una tasa con un piso de 7% real después de impuestos, considerando tanto el riesgo sistemático de las actividades propias de las empresas en relación con el mercado, como la tasa libre de riesgo y el premio por riesgo de mercado (Deloitte, Auditores y Consultores Limitada, 2021 y 2020). Esta última es la que utilizará para el ejercicio financiero de este proyecto de tesis.

7.2.4 Escenarios de Evaluación

El Escenario sin Proyecto representa la situación actual, donde se estima la cantidad de mantenimientos programados y correctivos por año y sus respectivas eficacias en base al tipo de mantenimiento.

El Escenario con Proyecto representa la incorporación del modelo predictivo y el rediseño del proceso que lo sustenta, donde la cantidad de desconexiones programadas considera

prioridad y las desconexiones de curso forzoso que contienen las fallas se ven disminuidas dada la incorporación del modelo.

7.2.5 Costos

La estructura de Costos de los Transformadores (TTFF) se puede observar en la Tabla 16. Este ejercicio está construido en base a los mantenimientos que se realizan anualmente a los TTFF, que son una proporción a los costos de Operación y mantenimiento del sistema eléctrico. Para el año 2019 corresponde al 5%, para el 2020 a un 3% y para el 2021 también un 3%.

Tabla 16: Costos de mantenimientos en los TTFF

| ERP SAP PM Mto. Tx. | |
|------------------------------------|----------------------|
| Mantenimiento para los TTFF | Monto en \$ |
| 2019 | \$125.051.943 |
| 2020 | \$71.928.155 |
| 2021 | \$120.329.971 |
| hasta mayo de 2022 | \$34.886.200 |
| Total | \$352.196.269 |

En los costos fijos no se considera lo relativo a la dotación de personas (considerando ajustes salariales y de IPC), ni tampoco otros costos de operación y mantenimiento, seguros involucrados y el nivel de inflación proyectado.

7.2.7 Inversión

En términos de inversión del proyecto, se considera necesaria la contratación de un Ingeniero que haga las labores de Data Scientist, utilice el modelo y trabaje a futuro en nuevos problemas de predicción y Machine learning. Se estima un costo de \$2.500.000 mensuales para el profesional contratado. Además, se solicita el uso de un computador con capacidades suficientes para correr modelos predictivos y grandes volúmenes de datos. Se estima un valor de \$1.200.000 para el computador. En resumen, la inversión total en un año es de \$31.200.000.-

7.2.8 Consideración Social de Beneficios

Si bien el beneficio social no es calculado ni evaluado en el flujo de caja, es importante destacar la importancia social que posee la línea de negocios de Transmisión Eléctrica en Chile. STS permite transmitir y transformar la energía eléctrica por grandes distancias a lo largo de Chile, y conecta la Generación eléctrica a la Distribución eléctrica para los hogares, instituciones del Estado y diferentes industrias nacionales. Los beneficios son concretamente las mejoras en la calidad de servicio (Voltaje) y la disponibilidad (sin cortes de energía eléctrica).

7.3 Flujo de Caja

Para el análisis del flujo de caja, se realiza un flujo por cada escenario, considerando la información de los reportes financieros de STS.

Para el escenario sin proyecto, se puede observar en la Tabla 17 el flujo (en millones de pesos). Este flujo implica una inversión en los años 2020 y 2021, que corresponde a la falla del transformador T44, la cual no se consideró como costo. Con este proyecto de tesis se espera anticipar a este tipo de fallas, por lo tanto, en el escenario con Proyecto, la inversión corresponde solo a lo indicado en el capítulo 7.2.7, en donde se necesita un Ingeniero de Data Analytics y un computador.

Tabla 17: Flujo de caja, escenario sin Proyecto

| | 0 | 1 | 2 | 3 |
|--|---|----------------|----------------|-----------------|
| Ingresos | | 37.005.000.000 | 34.047.769.000 | 61.112.458.000 |
| Costos operacionales | | -4.894.378.000 | -4.982.302.000 | -11.029.453.000 |
| Depreciación legal | | 0 | 0 | 0 |
| Intereses | | 0 | 0 | 0 |
| Ganancia de capital | | | | |
| Pérdida Recuperación Kt | | | | |
| Pérdida del ejercicio anterior | | | | |
| Utilidad antes de impuesto | | 32.110.622.000 | 29.065.467.000 | 50.083.005.000 |
| Impuesto (27%) renta de la primera categoría | | -8.669.867.940 | -7.847.676.090 | -13.522.411.350 |
| Utilidad después de impuesto | | 23.440.754.060 | 21.217.790.910 | 36.560.593.650 |
| Depreciación legal | | 0 | 0 | 0 |
| Pérdida de capital | | | | |
| Pérdida Recuperación Kt | | | | |
| Pérdida del ejercicio anterior | | | 0 | |
| Flujo operacional | | 23.440.754.060 | 21.217.790.910 | 36.560.593.650 |
| | | | | |
| Inversión | | | -889.888.760 | -10.239.509 |
| Valor residual | | | | |
| Capital de trabajo | | | | |
| Recuperación capital de trabajo | | | | |
| Préstamo | | | | |
| Amortización | | | | |
| Flujo Caja Capitales | 0 | 0 | -889.888.760 | -10.239.509 |
| | | | | |
| Flujo de caja neto | 0 | 23.440.754.060 | 20.327.902.150 | 36.550.354.141 |

Tabla 18: Indicadores, escenario sin Proyecto

| | |
|-------------------|-------------------------|
| R | 7,0% |
| VAN (Millones \$) | \$69.498.400.272 |

Los resultados del VAN en la Tabla 18 muestran que, a pesar de la falla del transformador, aún es rentable la compañía.

Para el resultado con proyecto que se observa en la Tabla 19, entrega un mejor flujo de caja.

Tabla 19: Flujo de caja, escenario con Proyecto

| | 0 | 1 | 2 | 3 |
|---------------------------------|-------------|----------------|----------------|-----------------|
| Ingreso por venta | | 37.005.000.000 | 34.047.769.000 | 61.112.458.000 |
| Costos operacionales | | -4.894.378.000 | -4.982.302.000 | -11.029.453.000 |
| Depreciación legal | | 0 | 0 | 0 |
| Intereses | | 0 | 0 | 0 |
| Ganancia de capital | | | | |
| Pérdida Recuperación Kt | | | | |
| Pérdida del ejercicio anterior | | | | |
| Utilidad antes de impuesto | | 32.110.622.000 | 29.065.467.000 | 50.083.005.000 |
| Impuesto (27%) | | | -7.847.676.090 | -13.522.411.350 |
| Utilidad después de impuesto | | 32.110.622.000 | 21.217.790.910 | 36.560.593.650 |
| Depreciación legal | | 0 | 0 | 0 |
| Pérdida de capital | | | | |
| Pérdida Recuperación Kt | | | | |
| Pérdida del ejercicio anterior | | | 0 | |
| Flujo operacional | | 32.110.622.000 | 21.217.790.910 | 36.560.593.650 |
| | | | | |
| Inversión | -31.200.000 | -30.000.000 | -30.000.000 | -30.000.000 |
| Valor residual | | | | |
| Capital de trabajo | | | | |
| Recuperación capital de trabajo | | | | |
| Préstamo | | | | |
| Amortización | | | | |
| Flujo Caja Capitales | -31.200.000 | -30.000.000 | -30.000.000 | -30.000.000 |
| | | | | |
| Flujo de caja neto | -31.200.000 | 32.080.622.000 | 21.187.790.910 | 36.530.593.650 |

Tabla 20: Indicadores, escenario con Proyecto

| | |
|----------------|-----------------------|
| R | 7,0% |
| VAN (Miles \$) | 78.276.772.905 |

El VAN de la Tabla 20 es superior al ejercicio financiero anterior, debido a que el proyecto de tesis detecta en los resultados de los entrenamientos de machine learning, la desconexión de curso forzoso del Trafo T44 en el año 2020. Los modelos no identifican el tipo de falla, pero si permite cambiar la prioridad para el mantenimiento programado.

7.4 Análisis de Sensibilidad

Como muestra la Tabla 21 se puede apreciar que el informe de año 2021, correspondiente a la indisponibilidad del año anterior, aumentó el índice SAIDI en 41,4 minutos en comparación con el informe del año 2020 y 2022. Se asume que la causa de este aumento fue en gran parte por la desconexión de curso forzoso del TTFF de T44 de Melipulli.

Tabla 21: Índice SAIDI por año

| Año de publicación del INDICE SAIDI | | |
|-------------------------------------|-------------------|---------------------|
| AÑO | Desconexión/horas | Desconexión/minutos |
| 2015 | 2,6 | 156 |
| 2016 | 1,77 | 106,2 |
| 2017 | 1,17 | 70,2 |
| 2018 | 1,18 | 70,8 |
| 2019 | 0,64 | 38,4 |
| 2020 | 0,39 | 23,4 |
| 2021 | 1,08 | 64,8 |
| 2022 | 0,42 | 25,2 |
| 2023 | 0,35 | 21 |
| 2024 | 0,34 | 20,4 |
| 2025 | 0,3 | 18 |

Por otro lado, la reparación del TTFF T44 de Melipulli fue considerada en el flujo de caja sin proyecto como una inversión, la cual afectó el estado de resultados de ese año. Por todo lo anterior expuesto, se estima que, si se repite una desconexión de curso forzoso de

otro TTF, cada año la compañía aumentará su índice SAIDI en 41 minutos aproximadamente, y además solicitará al Directorio de la Compañía un presupuesto de USD 1 millón de dólares adicional para el mantenimiento.

En la Tabla 22 se realiza un análisis What If en Excel, armando un escenario ficticio para ver el movimiento de los costos cuando aumentan las desconexiones de curso forzoso.

Tabla 22: Escenario ficticio donde se considera el porcentaje de desconexiones programadas, *versus* el porcentaje desconexiones de curso forzoso.

| | | | | | | |
|---|--|---------|---------|---------|---------|---------|
| Resumen del escenario ficticio | Se considera que una desconexión de curso forzoso en un Trafo. es 10 veces más que una desconexión programada. | | | | | |
| x falta de Información | Costo desconexión programada = 10 y Costo desconexión de curso forzoso = 100 | | | | | |
| Total de desconexiones al año = 100 | | | | | | |
| Desconexiones programadas % más alto | 50% | 60% | 70% | 80% | 90% | 100% |
| Desconexiones programadas | 50 | 60 | 70 | 80 | 90 | 100 |
| Desconexiones curso forzoso | 50 | 40 | 30 | 20 | 10 | 0 |
| Celdas de resultado: | | | | | | |
| Costo total por desconexiones | \$5.500 | \$4.600 | \$3.700 | \$2.800 | \$1.900 | \$1.000 |

Se comprueba, según los resultados de los entrenamientos supervisados, que las desconexiones de curso forzoso son anticipadas correctamente y, por lo tanto, disminuyen en el tiempo, pasando a ser desconexiones programadas, las cuales tienen un costo mucho menor y no afectan al índice SAIDI.

CAPÍTULO 8. CONCLUSIONES, LECCIONES APRENDIDAS, TRABAJOS FUTUROS.

El servicio que proporciona STS es de mucha importancia para el país, porque entrega energía eléctrica de calidad y en forma continua a sus clientes, velando por el cumplimiento de las normas exigidas por la autoridad nacional. La empresa tiene grandes desafíos para disminuir los índices de disponibilidad SAIDI, y además reducir sus costos de mantenimiento y operación, ser eficientes en el uso de sus activos, más competitivo y obtener mayor rentabilidad.

Se logra en este trabajo identificar las etapas, roles y funciones que permiten desarrollar nuevas capacidades de Data Analytics para problemas de predicción. También se logra identificar las variables necesarias para conocer el estado de salud de un transformador y anticiparse a las fallas. Se entrega un MVP con modelos predictivos que aportan con disminuir los COMA y el índice SAIDI.

Este estudio proporciona numerosas e importantes contribuciones para la compañía. La empresa tiene el conocimiento y experiencia en la Operación y Mantenimiento, pero no ha explorado las ventajas de trabajar con Data Analytics. De esto, nace la oportunidad de rediseñar el proceso de Estudios y Calidad de Servicio, incorporar tres modelos predictivos de Machine Learning para anticiparse a las desconexiones de curso forzoso en los transformadores de poder, y pasar de mantenimientos preventivos y correctivos a mantenimiento predictivos.

En el desarrollo de este proyecto, participaron diferentes stakeholders a partir de reuniones, entrevistas, documentos entregados por la organización. Todo lo anterior con el propósito de conocer de manera más detallada la estrategia del negocio, identificar los *drivers* de negocio, conocer los procesos y seleccionar los problemas. Con este trabajo inicial se levanta la arquitectura y estructura de procesos de la organización, y con foco en mejorar los índices que están asociados a los *drivers* de negocio seleccionados, se elige el problema a trabajar en esta tesis. El proyecto de tesis promueve un cambio conductual en la compañía, que exige una gestión del cambio, ya que son necesarias nuevas habilidades, que afectan la forma de pensar y trabajar.

En este trabajo se estudió sobre Data Analytics y cómo abordar los problemas de Machine Learning. Se programaron tres algoritmos de entrenamiento supervisado (clasificación), con datos públicos de la plataforma NEOMANTE del Coordinador Eléctrico Nacional, los cuales son datos externos a la organización, ya que la empresa no tenía la suficiente *data* para entrenar los modelos. Se iteró con datos históricos desde el año 2014 al 2019 comparando los resultados con el año 2020, con la información de las desconexiones programadas y de curso forzoso de todas las empresas eléctricas que participan en el mercado. Para obtener los resultados se ajustan los hiperparámetros de los modelos para que sea preciso, bajo criterios otorgados por los especialistas de la empresa y la literatura.

Una de las principales conclusiones del trabajo realizado, es que la Metodología de Ingeniería de Negocio del Magister, permite encontrar soluciones de rediseño de procesos que se encuentran alineadas a la estrategia del negocio, por lo tanto, pasa a solucionar un problema importante dentro de la compañía, con el apoyo de tecnologías de información. Se convierte en un aporte a la organización que entrega valor a los clientes, lo que genera al autor de esta tesis, y a los *stakeholders* de la organización, un incentivo para trabajar en el proyecto.

Otra conclusión, es que en la tercera iteración se obtuvo una buena aproximación en términos de predicción de mantenimientos programados y de curso forzoso. Se identifican las variables de entrada necesarias para anticiparse a las desconexiones de curso forzoso y con esto poder construir un modelo de mantenimiento predictivo. En esta línea, nace la necesidad en los *stakeholders*, de entrenar con estos tres algoritmos nuevos *dataset* para identificar variables de entradas más técnicas.

8.1 Lecciones Aprendidas

La metodología CRISP-DM y sus fases permite, de forma estructurada, conseguir un trabajo de calidad para solucionar problemas de Data Analytics. Con esta metodología se puede continuar trabajando en la organización, incorporar nuevos estudios y trabajos similares en otros tópicos.

La metodología de Ingeniería de Negocios con su enfoque en la estrategia y los procesos, tal como se mencionó en las conclusiones, crea valor alineando todos los esfuerzos en la dirección correcta de la organización. Es necesario volver a mencionar que la empresa comenzó un proceso grande y complejo de gestión de activos, y este proyecto de tesis, al estar alineado con los objetivos estratégicos, genera y entrega valor a la compañía, aportando incluso a este nuevo desafío a nivel organizacional. En síntesis, el proyecto de tesis no quedará obsoleto y servirá como base para nuevos proyectos de esta índole.

8.2 Trabajos Futuros

El trabajo desarrollado se basó en rediseñar el proceso de Estudios y Calidad de Servicio integrando modelos de entrenamiento supervisado de clasificación. Estos modelos fueron programados en base a la *data*, que en conjunto se acordó con los *stakeholders* de la organización.

Dentro del trabajo se presentaron nuevos requerimientos y oportunidades para generar proyectos que mejoren este rediseño y utilizarlos en otros problemas que tienen relación con otros activos eléctricos. Dentro de los nuevos requerimientos de mejoramiento encontrados se encuentran:

- Desarrollar una interfaz gráfica amigable para el usuario final, en donde puedan cargar datos de manera fácil y rápida.
- Incorporar variables técnicas al modelo, como son los niveles de aceite, temperatura, cantidad de operaciones de CTBC (Cambiador Tomas Bajo Carga), análisis de gases y estudios de polimerización del papel dentro de un transformador.
- Crear una cultura de los datos. Contar con formatos de datos tabular, para trabajar en algoritmos de Machine Learning. Actualmente, STS utiliza en sus inspecciones papel y formatos de texto no estructurado.
- Incorporar variables meteorológicas y volver a entrenar los modelos con el fin de analizar variables exógenas de los activos.

9. BIBLIOGRAFÍA

Araneda, JC. (2021). *Subgerente de Planificación, Coordinador Eléctrico Nacional*. Junio 10, 2022, desde https://www.cigre.cl/wp-content/uploads/2021/07/Sistema-Elctrico_CHILE_e_RIAC_2021.pdf

Franz, P., Kirchmer, M. (2012). *Value-Driven Business Process Management*.

Barros, O. (2015). *Ingeniería de Negocios: Diseño Integrado de Servicios, sus Procesos y Apoyo TI*.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*.

Hax, A. (2010). *The Delta Model, Reinventing your Business Strategy*. New York, USA: Springer.

Osterwalder, A., Pigneur, Y. (2009). *Business Model Generation (Generación de modelos de negocio)*.

Nelson, D. (2020). *What is a KNN (K-Nearest Neighbors)?*. Julio 15, 2022, desde <https://www.unite.ai/what-is-k-nearest-neighbors/>

Roman, V. (2019). *Algoritmos Naive Bayes: Fundamentos e Implementación*. Julio 15, 2022, desde <https://medium.com/datos-y-ciencia/algoritmos-naive-bayes-fudamentos-e-implementaci%C3%B3n-4bcb24b307f>

Corporate Finance Institute, (2021). *A combination of decision trees that can be modeled for prediction and behavior análisis*. Julio 15, 2022, desde <https://corporatefinanceinstitute.com/resources/knowledge/other/random-forest/>

Olguín, E., Crawford, B., & Soto, R. (2016). *Change Management in Technology Projects Using a Comprehensive Model of Change Management. Sistemas y Tecnologías de la Información, I*

Deloitte, Auditores y Consultores Limitada (2020). *Estados financieros por los años terminados al 31 de diciembre de 2020 y 2019.*
<https://web.gruposaes.cl/documents/20121/0/Memoria+Anual+Saesa+Transmisi%C3%B3n+S.A..pdf/c055baa1-1e52-39c9-1772-87171084be09?t=1617130368562>

Deloitte, Auditores y Consultores Limitada (2021). *Estados financieros por los años terminados al 31 de diciembre de 2021 y 2020.*
<https://web.gruposaes.cl/documents/20121/0/Estados+Financieros+STS+diciembre+2021.pdf/54df6b62-16c3-0531-ffff-1dcfd50dacf0?t=1648728193088>

Design Thinking (2023). *Pensamiento de diseño (design thinking en inglés), es una manera de trabajar en grupo que maximiza la creatividad colectiva.*
<https://www.ideo.com/>

ANEXOS

Anexo A: Extracto Metodología vBPM

Triangulación para mejorar procesos existentes

Según explica el libro vBPM, para triangular lo primero que se debe preguntar es: "¿Cuál es la necesidad operativa del negocio?"; Luego, "¿Cuáles son las fortalezas de la organización en términos de capacidad BPM?" y cuando estas dos preguntas se juntan, la tercera crea el tramo del triángulo a intervenir. En las siguientes figuras se evalúan los procesos nuevamente pero ahora con una puntuación del 1 al 3 en donde para la capacidad BPM la nota 3 tiene una alta capacidad de aplicar una mejora en el proceso aplicando Six Sigma, o cualquier otra metodología de la filosofía Lean. Y referente a lo imperativo del negocio la nota 3 es para calificar lo que es realmente necesario a intervenir en la empresa.

La Figura 55 muestra que el Proceso "Identificar fallas de red" tiene mayor urgencia para el negocio y una alta capacidad BPM. Este es el proceso que se selecciona y explica con detalle en el capítulo 3.3.1.

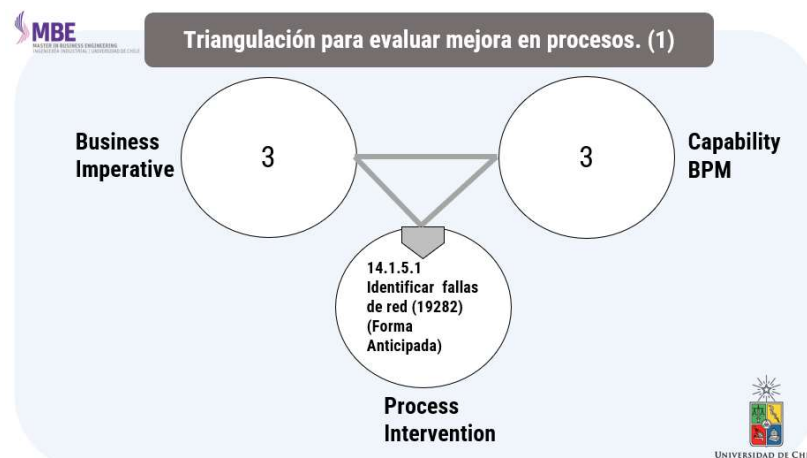


Figura 55: Triangulación para evaluar primer proceso seleccionado.

Fuente: Elaboración propia

La siguiente Figura 56 muestra que otro de los procesos seleccionados “Conducir planificación de respuesta a emergencias” tiene una alta capacidad BPM, pero en este caso puede ser aplazada su intervención, ya que la compañía tiene optimizado este proceso.

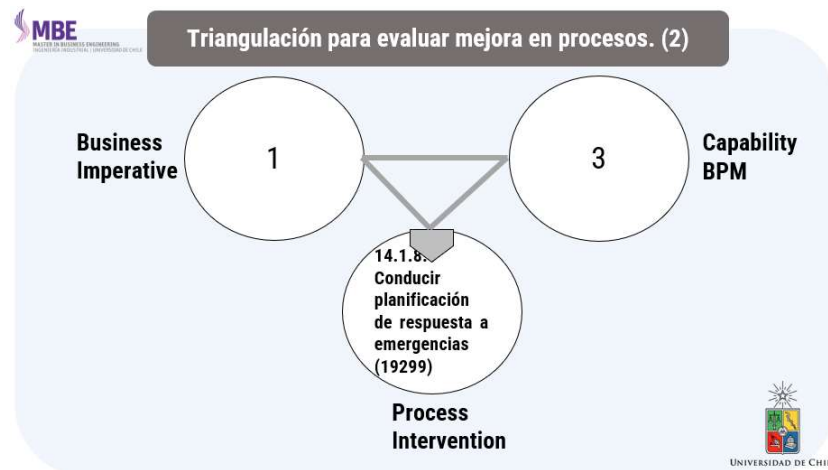


Figura 56: Triangulación para evaluar segundo proceso seleccionado.

Fuente: Elaboración propia

Continuando con la Figura 57, el proceso seleccionado “Identificar y gestionar las fallas de los clientes” no tiene una capacidad de intervención alta BPM y negocio. Esto debido a la complejidad de rediseñar este proceso, el cual afecta el servicio de los clientes.

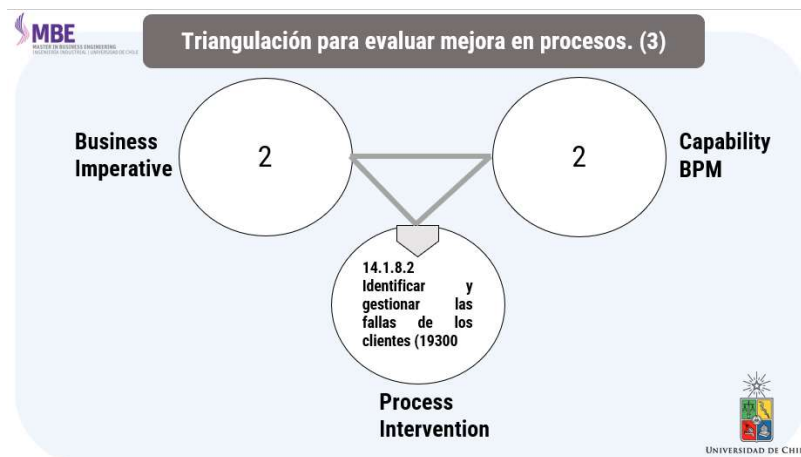


Figura 57: Triangulación para evaluar tercer proceso seleccionado.

Fuente: Elaboración propia

El cuarto proceso seleccionado a intervenir de la Figura 58 “Coordinar las operaciones de reparación y restauración”, es un proceso que tiene una baja necesidad de intervención, porque la empresa tiene resuelto desde hace unos años los procesos involucrados.

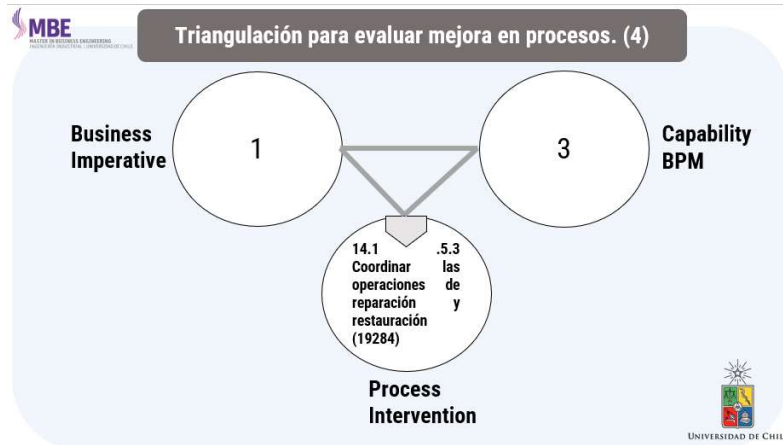


Figura 58: Triangulación para evaluar cuarto proceso seleccionado.

Fuente: Elaboración propia

La Figura 59 muestra el último proceso seleccionado a intervenir “Producir informes de la condición de la red”. Este proceso tiene mediana capacidad BPM y Negocio, debido a que existen otros frentes trabajando en estos procesos, con la ayuda de las Tecnologías de Información (TI).

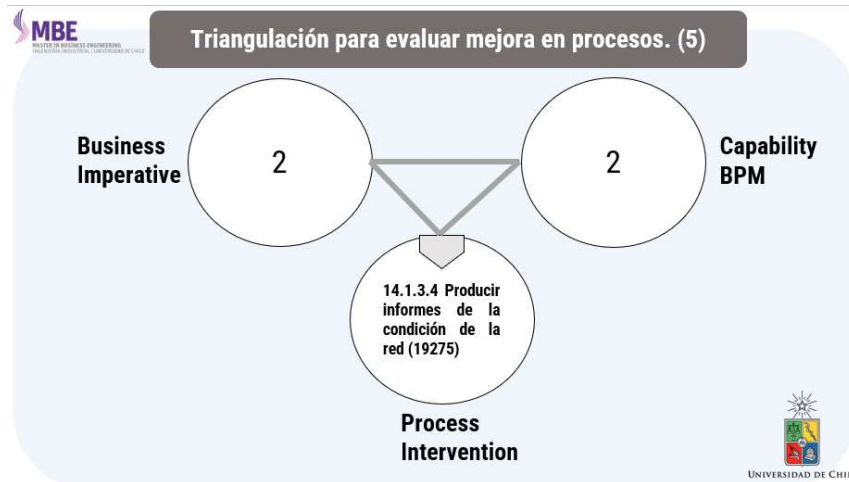


Figura 59: Triangulación para evaluar quinto proceso seleccionado.

Fuente: Elaboración propia

Anexo B: Selección del Problema

Este anexo contiene una serie de herramientas utilizadas para innovar creando valor a la empresa. Definir el problema es clave para encontrar la Causa Raíz y el ¿Por qué ocurren? Esta etapa involucra trabajo en equipo y con profesionales multidisciplinares. La Figura 60 destaca en la metodología Design Thinking, el momento en que se define el problema y su propuesta de solución.

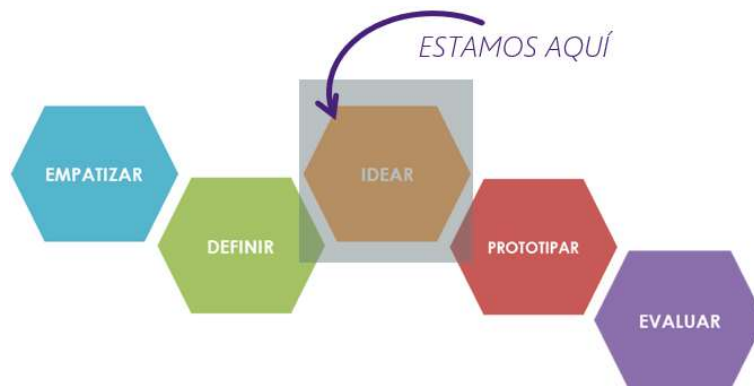


Figura 60: Design Thinking

Fuente: Proceso de Design Thinking, Ideo

Causa Raíz

En la Figura 61 se explica cómo se trabajó en conjunto con los stakeholders del área de Calidad y Estudios para identificar la causa raíz (Ishikawa, 1943).

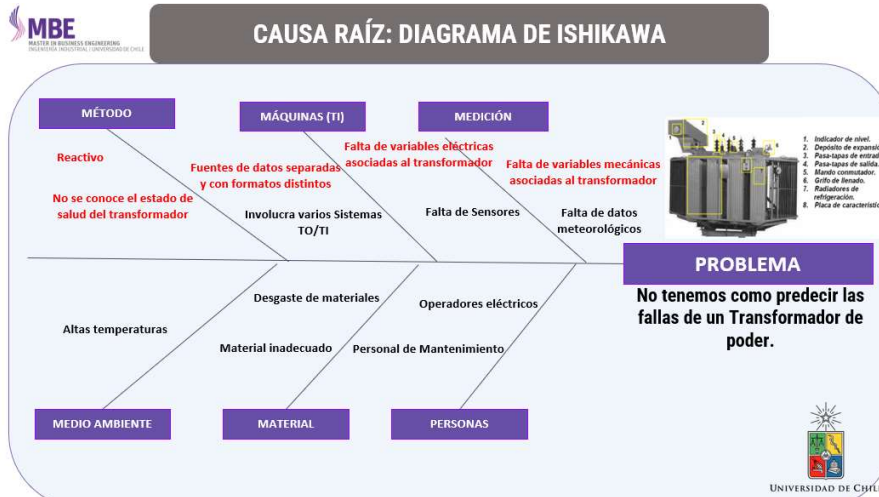


Figura 61: Causa Raíz de Ishikawa

Fuente: Elaboración propia

La Figura 61 fue trabajada en conjunto con los ingenieros del área de estudio y calidad de servicio con el diagrama de Ishikawa, el cual simula una espina de pescado. Se formula el problema a solucionar y se identifican las siguientes aristas, que ayudan a identificar la causa raíz del problema:

- Método. En la organización es reactiva y no se conoce el estado de salud de los transformadores.
- Máquinas (TI): Existen diversas fuentes de datos entre los sistemas TI y TO. Además, con los datos son con diferentes formatos, de los cuales no todos se pueden utilizar con tecnologías de Data Analytics.
- Medición: Faltan datos mecánicos de los transformadores y no se cuenta con información meteorológica de la ubicación de los equipos.
- Medio ambiente: Los equipos se encuentran en lugares de condiciones adversas climáticas, afectadas por altas temperaturas.
- Material: Los transformadores tienen desgaste de material, algunos en operación por largos años de funcionamiento y son exigidos a sus máximas capacidades para una operación continua.
- Personas: Los Operadores del Centro de Control y Mantenimiento son profesionales muy bien capacitados en sus áreas de especialidad, pero no tienen una cultura de datos, por lo tanto, la forma de guardar los datos es análoga.

En la Figura 62 se realiza un análisis causa raíz con los 5 ¿Por qué?.

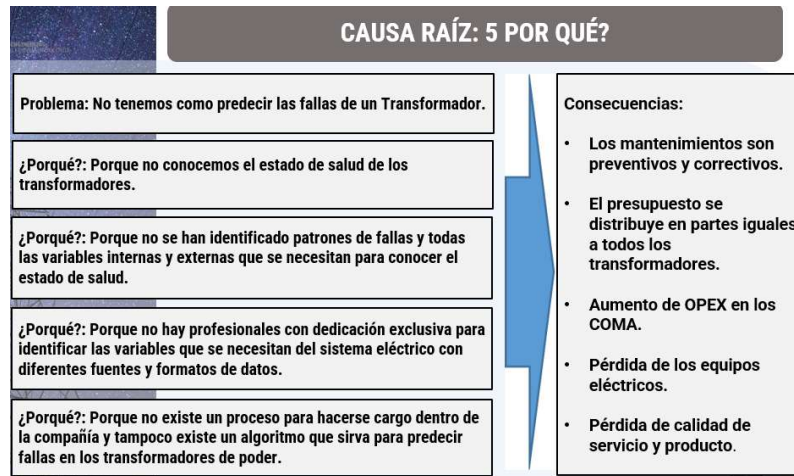


Figura 62: Causa Raíz de los 5 ¿por qué?

Fuente: Elaboración propia

Además, para buscar posibles soluciones, se utilizó el método de brainstorming como muestra la Figura 63.

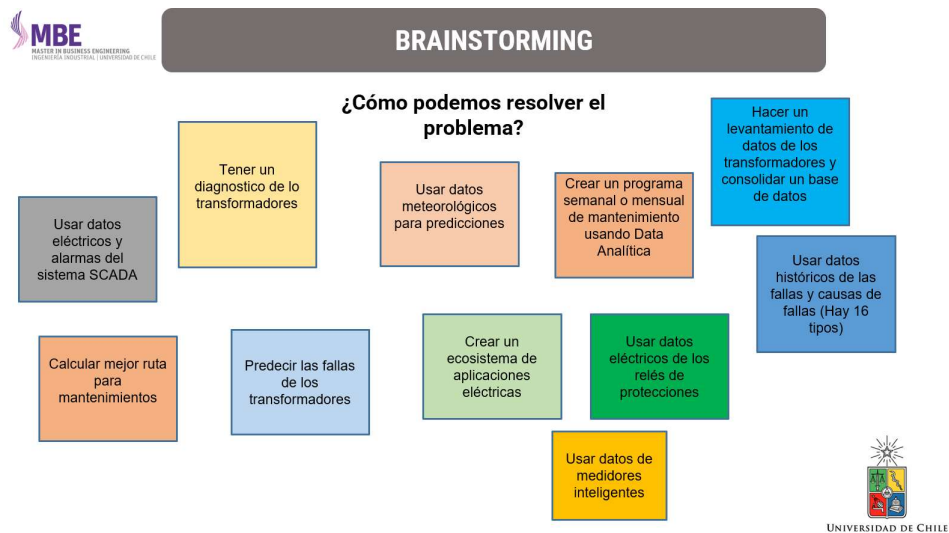


Figura 63: Brainstorming

Fuente: Elaboración propia

Para seleccionar y priorizar las posibles soluciones se creó una matriz de Impacto / Esfuerzo Figura 64.

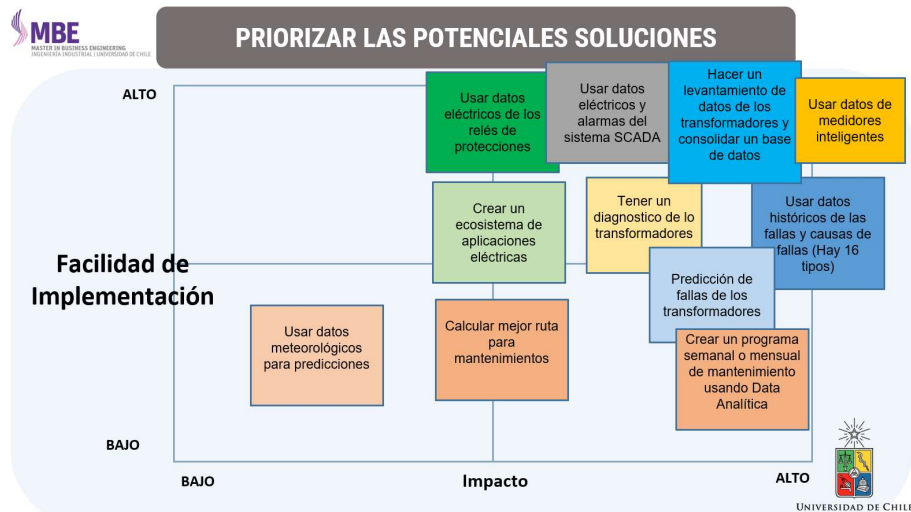


Figura 64: Matriz Impacto esfuerzo con potenciales soluciones

Fuente: Elaboración propia

De la matriz impacto/esfuerzo se selecciona trabajar en un modelo para predecir las fallas de los transformadores de poder y a continuación en la Figura 65 se propone crear un MVP producto mínimo viable.

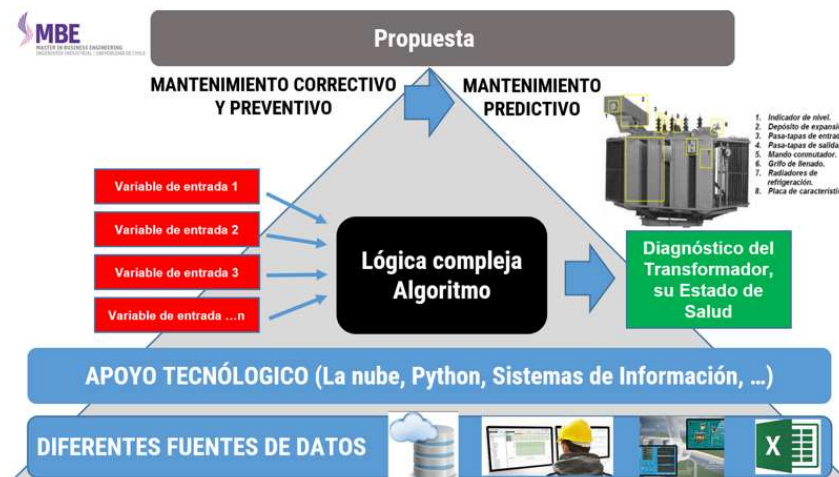


Figura 65: Propuesta de MVP

Fuente: Elaboración propia

Anexo C: Algoritmos de Entrenamiento Supervisado con Python

Archivos adjuntos de Jupyter Notebook programado en Python y Data Set utilizado

1. Tesis_Maching_Learning_Modelos_Clasificacion.ipynb
2. Data_Trafos_v4.csv

Primeros Resultados

Los primeros resultados son con el método de selección de atributos, con un 20% de test. Además, para correr varias iteraciones y cambiar los hiperparámetros, sumando a esto la limitación de recursos del computador, se disminuye la cantidad de registros del dataset, considerando los transformadores de la empresa STS. También se balancea el target para una mejor distribución normal de los datos.

Recordemos que nuestro target es:

- Target = Predecir
 - Desconexiones Programada = 1
 - Desconexiones de Curso Forzoso = 0

| KNN | | | | | NB | | | | | RF | | | | | | | |
|---------------------|--------------|--------|----------|---------|---------------------|-----------|--------------|----------|---------|---------------------|-----------|--------|--------------|---------|------|------|----|
| [[23 4] [7 27]] | | | | | [[24 3] [0 34]] | | | | | [[25 2] [0 34]] | | | | | | | |
| | precision | recall | f1-score | support | | precision | recall | f1-score | support | | precision | recall | f1-score | support | | | |
| | 0.0 | 0.77 | 0.85 | 0.81 | 27 | | 0.0 | 1.00 | 0.89 | 0.94 | 27 | | 0.0 | 1.00 | 0.93 | 0.96 | 27 |
| | 1.0 | 0.87 | 0.79 | 0.83 | 34 | | 1.0 | 0.92 | 1.00 | 0.96 | 34 | | 1.0 | 0.94 | 1.00 | 0.97 | 34 |
| | accuracy | | | 0.82 | 61 | | accuracy | | | 0.95 | 61 | | accuracy | | | 0.97 | 61 |
| | macro avg | 0.82 | 0.82 | 0.82 | 61 | | macro avg | 0.96 | 0.94 | 0.95 | 61 | | macro avg | 0.97 | 0.96 | 0.97 | 61 |
| | weighted avg | 0.82 | 0.82 | 0.82 | 61 | | weighted avg | 0.95 | 0.95 | 0.95 | 61 | | weighted avg | 0.97 | 0.97 | 0.97 | 61 |

Figura 66: Primeros resultados.

Fuente: Elaboración propia

De la figura 66 obtendremos los resultados de los algoritmos y sus costos Accuracy, Precision y Recall. Se programa la matriz de confusión y la curva de ROC para evaluar los resultados de los modelos.

En la siguiente tabla 23, se explican los resultados de los TP, TN, FP y FN.

Tabla 23: Entradas y salidas obtenidas por los algoritmos.

| Algoritmos Clasificando | | | | |
|-----------------------------|----------------|---------|-----------|--------|
| Desconexión Programada = | | | | 1 |
| Desconexión Curso Forzoso = | | | | 0 |
| TP | True Positive | Entrada | Clasifica | Salida |
| | | 1 | | 1 |
| FP | False Positive | Entrada | Clasifica | Salida |
| | | 0 | | 1 |
| TN | True Negative | Entrada | Clasifica | Salida |
| | | 0 | | 0 |
| FN | False Negative | Entrada | Clasifica | Salida |
| | | 1 | | 0 |

El resultado de la siguiente tabla 24 corresponde al Accuracy (cantidad de predicciones correctas).

Tabla 24: Primeros resultados del Accuracy

| Algoritmos Clasificando | |
|-------------------------|----------|
| Algoritmo | Accuracy |
| KNN | 82% |
| NB | 95% |
| RF | 97% |

Los resultados para la predicción de desconexiones programadas se muestran en la siguiente tabla 25, en donde encontramos un sobreajuste sobre el recall o sensibilidad llegando a un 100% en los positivos actuales.

Tabla 25: Primeros resultados de los algoritmos KNN, NB y RF, en la clasificación de desconexiones programadas

| Algoritmos Clasificando | | |
|--------------------------|-----------|--------|
| Desconexión Programada = | | 1 |
| Algoritmo | Precision | Recall |
| KNN | 87% | 79% |
| NB | 92% | 100% |
| RF | 94% | 100% |

Para el caso de la clasificación de desconexiones de curso forzoso nos encontramos en la tabla 26 un sobreajuste en la precisión, sobre todo en los resultados de predicción de positivos.

Tabla 26: Primeros resultados de los algoritmos KNN, NB y RF en la clasificación de desconexiones de curso forzoso

| Algoritmos Clasificando | | |
|-----------------------------|-----------|--------|
| Desconexión curso forzoso = | | 0 |
| Algoritmo | Precision | Recall |
| KNN | 77% | 85% |
| NB | 100% | 89% |
| RF | 100% | 93% |

La tabla 27 ordena la matriz (2x2) de confusión con los TP, TN, FP y FN.

Tabla 27: Matriz de confusión 2x2. TP, TN, FP y FN

| | Condición Actual es +V | Condición Actual es -V |
|----------------------------|------------------------|------------------------|
| Condición Predictiva es +V | TP | FP |
| Condición Predictiva es -V | FN | TN |

Podemos observar en la Figura 67 que en los algoritmos KNN y RF hay sobreajuste en los TN.

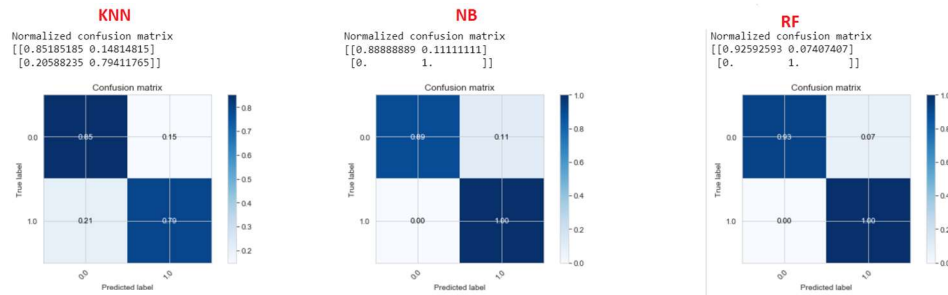


Figura 67: Primeros resultados.

Fuente: Elaboración propia

En la Figura 68, referente a la curva de ROC, el área sobre la diagonal del 50% indica un mejor resultado de los algoritmos. Pero, si los resultados de la curva se acercan demasiado al borde del 100%, nos indica que existe un sobreajuste. Gráficamente se aprecia este comportamiento de sobreajuste en el modelo RF.

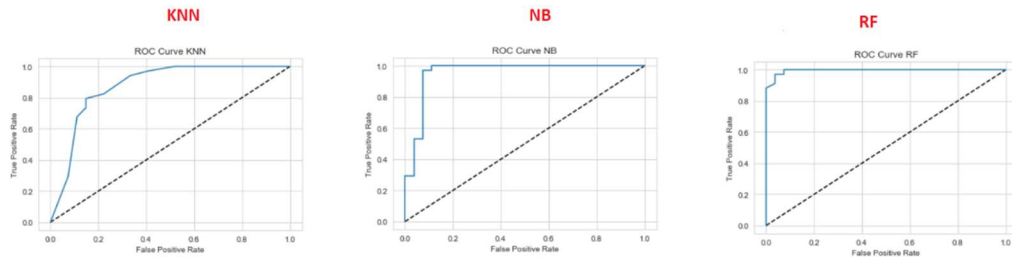


Figura 68: Curca de ROC de los Primeros resultado.

Fuente: Elaboración propia

Primera Iteración

Primera Iteración, ajustando los hiperparametros con Grid Search. Se considera una configuración en el Grid de cv = 5, verbose = 3 y n_jobs = -1.

Grid Search

```
Fitting 5 folds for each of 8 candidates, totalling 40 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.
[Parallel(n_jobs=-1)]: Done 16 tasks | elapsed: 13.1min
[Parallel(n_jobs=-1)]: Done 40 out of 40 | elapsed: 42.2min finished
{'C': 0.1, 'gamma': 'scale', 'kernel': 'linear'}
[[23  4]
 [ 0 34]]
      precision    recall  f1-score   support

      0.0         1.00     0.85     0.92         27
      1.0         0.89     1.00     0.94         34

 accuracy          0.93         61
 macro avg         0.95     0.93     0.93         61
 weighted avg      0.94     0.93     0.93         61

Normalized confusion matrix
[[0.85185185 0.14814815]
 [0.         1.         ]]
```

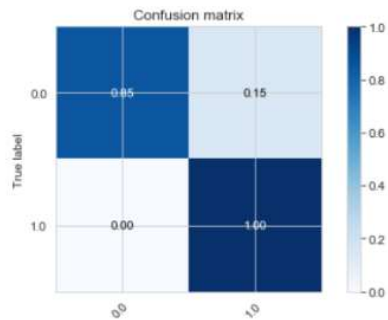


Figura 69: Grid Search de la primera iteración.

Fuente: Elaboración propia

Se ajustan los hiperparametros con Grid Search de la Figura 69 anterior, y se programa una validación cruzada (cv) de 5 folds, obteniendo un accuracy de 93%.

Con esta corrección del hiperajuste Grid Search, se corren nuevamente los tres algoritmos, teniendo los resultados de la Figura 70.

| KNN | | | | | NB | | | | | RF | | | | | | | |
|---------------------|--------------|--------|----------|---------|---------------------|-----------|--------------|----------|---------|---------------------|-----------|--------|--------------|---------|------|------|----|
| [[23 4] [7 27]] | | | | | [[24 3] [0 34]] | | | | | [[24 3] [0 34]] | | | | | | | |
| | precision | recall | f1-score | support | | precision | recall | f1-score | support | | precision | recall | f1-score | support | | | |
| | 0.0 | 0.77 | 0.85 | 0.81 | 27 | | 0.0 | 1.00 | 0.89 | 0.94 | 27 | | 0.0 | 1.00 | 0.89 | 0.94 | 27 |
| | 1.0 | 0.87 | 0.79 | 0.83 | 34 | | 1.0 | 0.92 | 1.00 | 0.96 | 34 | | 1.0 | 0.92 | 1.00 | 0.96 | 34 |
| | accuracy | | | 0.82 | 61 | | accuracy | | | 0.95 | 61 | | accuracy | | | 0.95 | 61 |
| | macro avg | 0.82 | 0.82 | 0.82 | 61 | | macro avg | 0.96 | 0.94 | 0.95 | 61 | | macro avg | 0.96 | 0.94 | 0.95 | 61 |
| | weighted avg | 0.82 | 0.82 | 0.82 | 61 | | weighted avg | 0.95 | 0.95 | 0.95 | 61 | | weighted avg | 0.95 | 0.95 | 0.95 | 61 |

Figura 70: Resultados de los algoritmos en la primera iteración.

Fuente: Elaboración propia

En la siguiente tabla 28, se rescatan los resultados del Accuracy de esta primera iteración

Tabla 28: Accuracy de la primera iteración

| Algoritmos Clasificando | |
|-------------------------|----------|
| Algoritmo | Accuracy |
| KNN | 82% |
| NB | 95% |
| RF | 95% |

Se aprecia que gracias a los nuevos ajustes del hiperparametro Grid Search, el algoritmo RF baja de 97% a 95%.

En la tabla 29 se colocan los resultados de las desconexiones programadas con los costos de Precision y recall.

Tabla 29: Precision y Recall de la primera iteración para desconexión programada

| Algoritmos Clasificando | | |
|--------------------------|-----------|--------|
| Desconexión Programada = | | 1 |
| Algoritmo | Precision | Recall |
| KNN | 87% | 79% |
| NB | 92% | 100% |
| RF | 92% | 100% |

El algoritmo RF, baja de 94 a 92. Disminuyen los FP en Precision.

En la tabla 30 se muestran los resultados de las desconexiones de curso forzoso de esta primera iteración.

Tabla 30: Precision y Recall de la primera iteración para desconexión curso forzoso

| Algoritmos Clasificando | | |
|-----------------------------|-----------|--------|
| Desconexión curso forzoso = | 0 | |
| Algoritmo | Precision | Recall |
| KNN | 77% | 85% |
| NB | 100% | 89% |
| RF | 100% | 89% |

El algoritmo RF, baja a 93 a 89. Aquí se ve la disminución de los FN en Recall.

De la matriz de confusión observamos en la Figura 71, que disminuye solo el RF con los TP de 93% a 88% y aumentan los FP de 7 a 11%. Por lo cual no es un buen resultado.

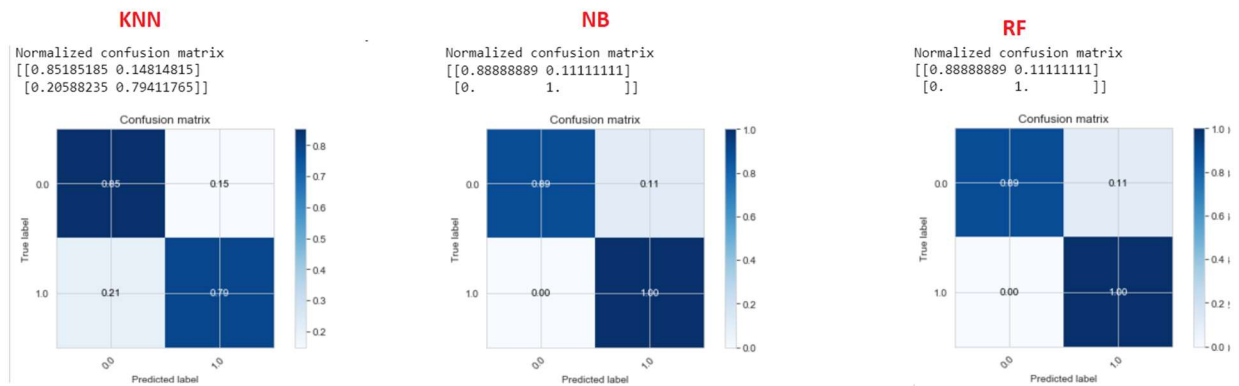


Figura 71: Matriz de confusión con los resultados de los algoritmos en la primera iteración.

Fuente: Elaboración propia

Para esta iteración, las curvas de ROC de la Figura 72, grafica que es solo un poco mejor también el algoritmo RF, porque sigue con sobreajuste.

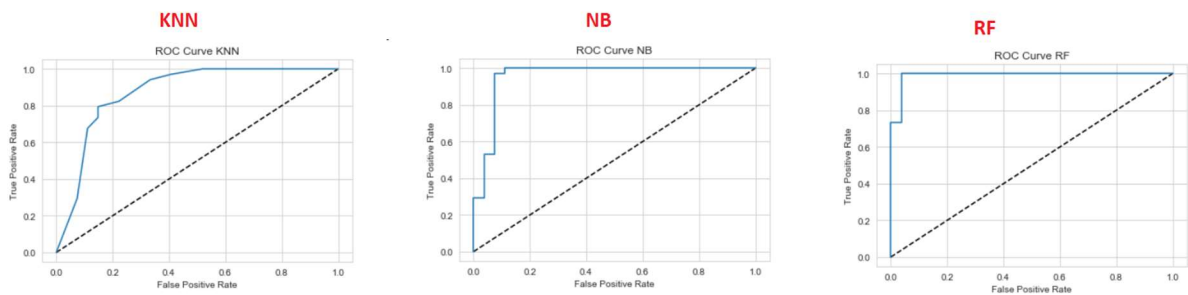


Figura 72: Curva de ROC de los algoritmos en la primera iteración.

Fuente: Elaboración propia

Segunda Iteración

En la segunda iteración, ajustando los hiperparametros con Random Search, se obtiene el accuracy de la Figura 73, que es 85% menor que el 93% del Grid Search anterior.

El código se configura de la siguiente manera: `random = RandomizedSearchCV(SVC(), space, refit = True, cv = 5, verbose=3, n_jobs=-1, n_iter=5, scoring='accuracy')`

Random Search

```
Fitting 5 folds for each of 5 candidates, totalling 25 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.
[Parallel(n_jobs=-1)]: Done 19 out of 25 | elapsed: 38.5s remaining: 12.1s
[Parallel(n_jobs=-1)]: Done 25 out of 25 | elapsed: 5.5min finished
{'C': 5.67403097300958, 'gamma': 0.020397879316757614, 'kernel': 'linear'}
      precision    recall  f1-score   support

    0.0         0.88         0.78         0.82         27
    1.0         0.84         0.91         0.87         34

 accuracy                   0.85         61
 macro avg          0.86         0.84         0.85         61
 weighted avg       0.85         0.85         0.85         61
```

Figura 73: Resultados de los ajustes de hiperparametros con Random Search.

Fuente: Elaboración propia

En esta etapa, se ajustan los hiperparametros con Random Search y una validación cruzada de 5 folds. La cantidad de iteraciones es importante, porque entre más iteraciones, son más ajustes, por ende, más tiempo toma en correr el algoritmo.

Nuevamente se corren los tres algoritmos, KNN, NB y RF. Los resultados se ilustran a continuación en la Figura 74.

| KNN | | | | | NB | | | | | RF | | | | |
|---------------------|--------|----------|---------|----|---------------------|--------|----------|---------|----|---------------------|--------|----------|---------|--|
| [[23 4] [7 27]] | | | | | [[24 3] [0 34]] | | | | | [[23 4] [0 34]] | | | | |
| precision | recall | f1-score | support | | precision | recall | f1-score | support | | precision | recall | f1-score | support | |
| 0.0 | 0.77 | 0.85 | 0.81 | 27 | 0.0 | 1.00 | 0.89 | 0.94 | 27 | 0.0 | 1.00 | 0.85 | 0.92 | |
| 1.0 | 0.87 | 0.79 | 0.83 | 34 | 1.0 | 0.92 | 1.00 | 0.96 | 34 | 1.0 | 0.89 | 1.00 | 0.94 | |
| accuracy | | | 0.82 | 61 | accuracy | | | 0.95 | 61 | accuracy | | | 0.93 | |
| macro avg | 0.82 | 0.82 | 0.82 | 61 | macro avg | 0.96 | 0.94 | 0.95 | 61 | macro avg | 0.95 | 0.93 | 0.93 | |
| weighted avg | 0.82 | 0.82 | 0.82 | 61 | weighted avg | 0.95 | 0.95 | 0.95 | 61 | weighted avg | 0.94 | 0.93 | 0.93 | |

Figura 74: Resultados de los algoritmos en la segunda iteración

Fuente: Elaboración propia

Para analizar los resultados por separado se colocan en nuevas tablas.

Tabla 31: Accuracy segunda iteración

| Algoritmos Clasificando | |
|-------------------------|----------|
| Algoritmo | Accuracy |
| KNN | 82% |
| NB | 95% |
| RF | 93% |

La tabla 31, nos muestra que el algoritmo RF baja otra vez de 95 a 93, mejorando el sobreajuste. Y en la tabla 32, para la desconexión programada los costos en Precision baja en RF de 92 a 89. Aquí se aprecia una mejora en los FP.

Tabla 32: Precision y Recall segunda iteración para desconexiones programadas

| Algoritmos Clasificando | | |
|--------------------------|-----------|--------|
| Desconexión Programada = | | 1 |
| Algoritmo | Precision | Recall |
| KNN | 87% | 79% |
| NB | 92% | 100% |
| RF | 89% | 100% |

En la siguiente tabla 33, para la clasificación de desconexiones de curso forzoso, los costos en Recall bajan en RF de 89 a 85. Se ve una mejora en los FN.

Tabla 33: Precision y Recall segunda iteración para desconexiones de curso forzoso

| Algoritmos Clasificando | | |
|-----------------------------|-----------|--------|
| Desconexión curso forzoso = | | 0 |
| Algoritmo | Precision | Recall |
| KNN | 77% | 85% |
| NB | 100% | 89% |
| RF | 100% | 85% |

La matriz de confusión de la Figura 75, en esta iteración disminuye solo el RF con los TP de 88% a 85% y aumentan los FP de 11 a 15%. Esto no es un buen resultado.

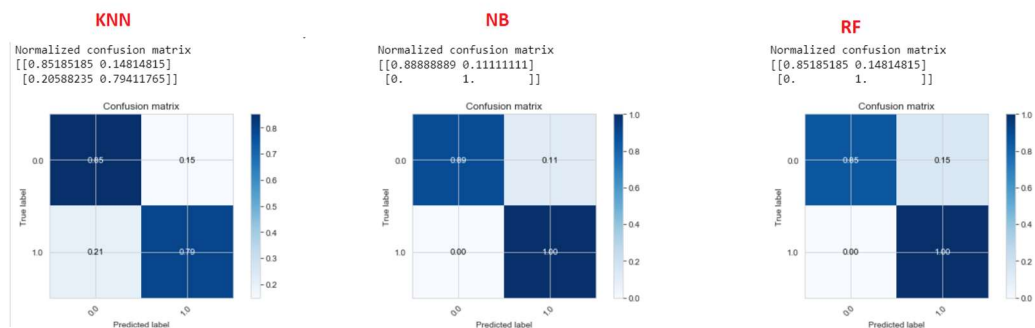


Figura 75: Matriz de confusión con los resultados de los algoritmos en la segunda iteración.

Fuente: Elaboración propia

La curva de ROC en la Figura 76, KNN Y NB no cambian, pero mejora el sobreajuste antes obtenido en RF, en comparación con los primeros resultados.

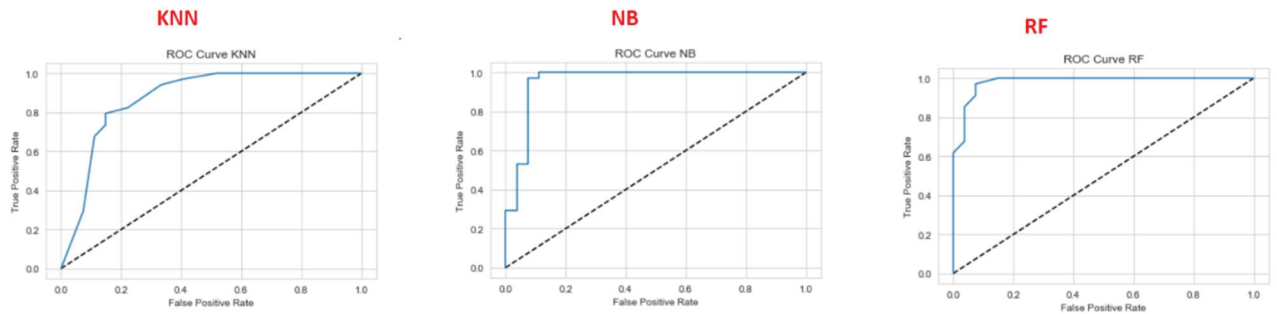


Figura 76: Curva de ROC de los algoritmos en la segunda iteración.

Fuente: Elaboración propia

Tercera Iteración

Figura 77 con la Tercera Iteración que cambia la selección de atributo con método embebido. Cambia el Support de 27 y 34 a 37 y 39.

Selección de atributos con Método Embedded

```
In [151]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0, stratify=y)
```

Figura 77: Nueva selección de atributos en una tercera iteración.

Fuente: Elaboración propia

Con este cambio se obtienen mejores resultados en los algoritmos, evitando sobreajustes. La Figura 78 muestra que todos los algoritmos de clasificación sufren un cambio en sus costos.

| KNN | | | | | NB | | | | | RF | | | | |
|---------------------|--------|----------|---------|----|---------------------|--------|----------|---------|----|---------------------|--------|----------|---------|----|
| [[31 6] [12 27]] | | | | | [[32 5] [3 36]] | | | | | [[35 2] [4 35]] | | | | |
| precision | recall | f1-score | support | | precision | recall | f1-score | support | | precision | recall | f1-score | support | |
| 0.0 | 0.72 | 0.84 | 0.77 | 37 | 0.0 | 0.91 | 0.86 | 0.89 | 37 | 0.0 | 0.90 | 0.95 | 0.92 | 37 |
| 1.0 | 0.82 | 0.69 | 0.75 | 39 | 1.0 | 0.88 | 0.92 | 0.90 | 39 | 1.0 | 0.95 | 0.90 | 0.92 | 39 |
| accuracy | | | 0.76 | 76 | accuracy | | | 0.89 | 76 | accuracy | | | 0.92 | 76 |
| macro avg | 0.77 | 0.77 | 0.76 | 76 | macro avg | 0.90 | 0.89 | 0.89 | 76 | macro avg | 0.92 | 0.92 | 0.92 | 76 |
| weighted avg | 0.77 | 0.76 | 0.76 | 76 | weighted avg | 0.90 | 0.89 | 0.89 | 76 | weighted avg | 0.92 | 0.92 | 0.92 | 76 |

Figura 78: Resultados de los algoritmos en la tercera iteración

Fuente: Elaboración propia

Tabla 34: Accuracy tercera iteración

| Algoritmos Clasificando | | |
|-------------------------|----------|-----------------|
| Algoritmo | Accuracy | |
| KNN | 76% | baja de 82 a 76 |
| NB | 89% | baja de 95 a 89 |
| RF | 92% | baja de 93 a 92 |

Como muestra la tabla 34 el accuracy baja de 93 a 92 en esta iteración. Y en las siguientes tablas 35 y 36, en los tres algoritmos de clasificación los costos de Precision y Recall sufren modificaciones.

Tabla 35: Precision y recall en la tercera iteración para desconexión programada

| Algoritmos Clasificando | | | | |
|--------------------------|-----------|--------|-----------------|------------------|
| Desconexión Programada = | | 1 | | |
| Algoritmo | Precision | Recall | Precision | Recall |
| KNN | 82% | 69% | baja de 87 a 82 | baja de 79 a 69 |
| NB | 88% | 92% | baja de 92 a 88 | baja de 100 a 92 |
| RF | 95% | 90% | sube de 89 a 95 | baja de 100 a 90 |

Tabla 36: Precision y recall en la tercera iteración para desconexión de curso forzoso

| Algoritmos Clasificando | | | | |
|-----------------------------|-----------|--------|------------------|-----------------|
| Desconexión curso forzoso = | | 0 | | |
| Algoritmo | Precision | Recall | Precision | Recall |
| KNN | 72% | 84% | baja de 77 a 72 | baja de 85 a 84 |
| NB | 91% | 86% | baja de 100 a 91 | baja de 89 a 86 |
| RF | 90% | 95% | baja de 100 a 90 | sube de 85 a 95 |

En las matrices de confusión de la Figura 79, todos tienen un cambio favorable. Se resumen los resultados en la tabla comparativa número 34.

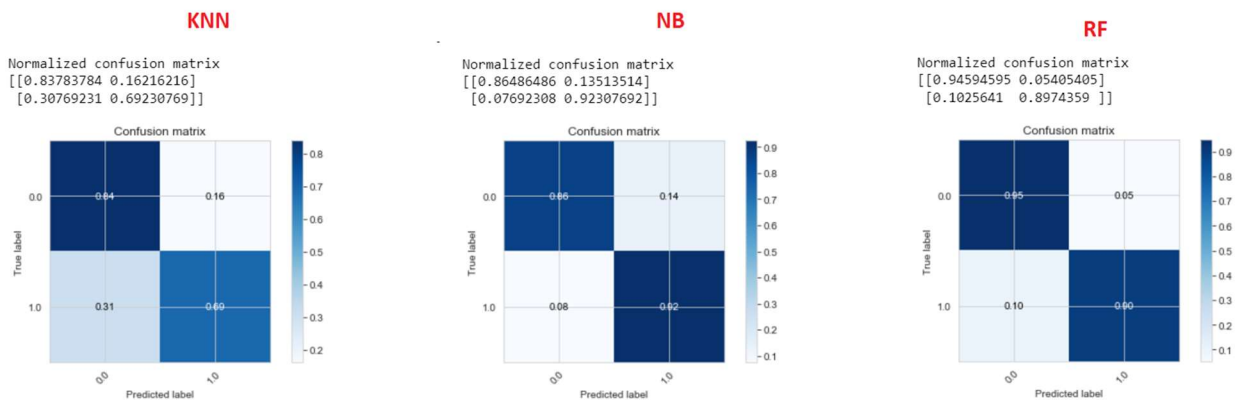


Figura 79: Matriz de confusión con los resultados de los algoritmos en la tercera iteración.

Fuente: Elaboración propia

Podemos ver que en esta tabla 37 los TN mejoran y FN disminuyen en los tres algoritmos. En RF los FP obtienen el mejor resultado hasta el momento.

Tabla 37: Comparación matriz de confusión entre la segunda y tercera iteración

| Comparación de Matriz de confusión | | | | |
|------------------------------------|----|----|----|-----|
| KNN | TP | FP | FN | TN |
| Segunda Iteración | 85 | 15 | 21 | 79 |
| Tercera Iteración | 84 | 16 | 31 | 69 |
| NB | TP | FP | TN | FN |
| Segunda Iteración | 89 | 11 | 0 | 100 |
| Tercera Iteración | 86 | 14 | 8 | 92 |
| RF | TP | FP | TN | FN |
| Segunda Iteración | 85 | 15 | 0 | 100 |
| Tercera Iteración | 95 | 5 | 10 | 90 |

Obteniendo de lo anterior, las gráficas de la Figura 80, en donde las curvas de ROC se encuentran sin sobreajustes. Sobre todo, mejoran los algoritmos NB y RF.

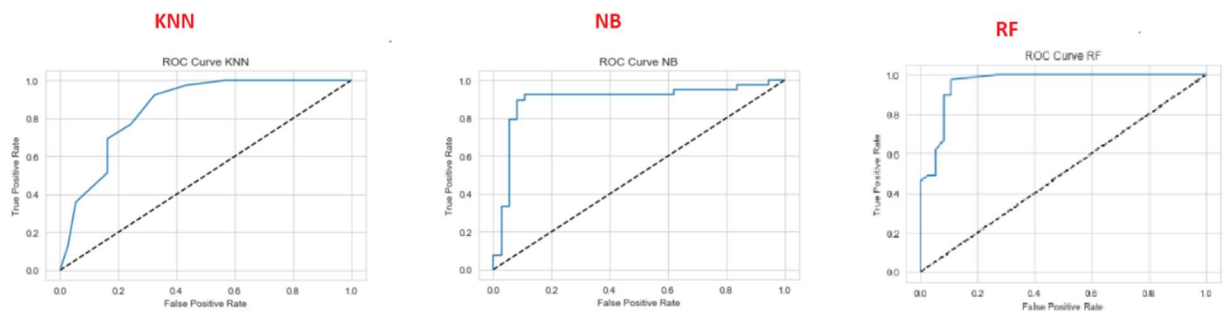


Figura 80: Curva de ROC de los algoritmos en la tercera iteración.

Fuente: Elaboración propia

Cuarta Iteración

En esta cuarta iteración se muestra en la Figura 81, que el accuracy baja de 93% a 82% en comparación con la tercera iteración.

Grid Search

```
Fitting 5 folds for each of 8 candidates, totalling 40 fits

[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.
[Parallel(n_jobs=-1)]: Done 16 tasks      | elapsed: 4.5min
[Parallel(n_jobs=-1)]: Done 40 out of 40 | elapsed: 14.8min finished

{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
      precision    recall  f1-score   support

      0.0         0.73     0.97     0.84         37
      1.0         0.96     0.67     0.79         39

 accuracy
macro avg      0.85     0.82     0.81         76
weighted avg   0.85     0.82     0.81         76
```

Figura 81: Resultados de los ajustes de hiperparametros con Grid Search en la cuarta iteración.

Fuente: Elaboración propia

Al correr los algoritmos obtenemos los resultados que se muestran en la Figura 82, en donde solo se aprecia que el algoritmo RF sufre de cambios en Precision y Recall.

| KNN | | | | | NB | | | | | RF | | | | | | | |
|---------------------|------|-----------|--------|----------|--------------|---------------------|------|-----------|--------|--------------|---------|---------------------|------|-----------|--------|----------|---------|
| [[31 6] [12 27]] | | precision | recall | f1-score | support | [[32 5] [3 36]] | | precision | recall | f1-score | support | [[34 3] [3 36]] | | precision | recall | f1-score | support |
| 0.0 | 0.72 | 0.84 | 0.77 | 37 | 0.0 | 0.91 | 0.86 | 0.89 | 37 | 0.0 | 0.92 | 0.92 | 0.92 | 37 | | | |
| 1.0 | 0.82 | 0.69 | 0.75 | 39 | 1.0 | 0.88 | 0.92 | 0.90 | 39 | 1.0 | 0.92 | 0.92 | 0.92 | 39 | | | |
| accuracy | | | 0.76 | 76 | accuracy | | | 0.89 | 76 | accuracy | | | 0.92 | 76 | | | |
| macro avg | 0.77 | 0.77 | 0.76 | 76 | macro avg | 0.90 | 0.89 | 0.89 | 76 | macro avg | 0.92 | 0.92 | 0.92 | 76 | | | |
| weighted avg | 0.77 | 0.76 | 0.76 | 76 | weighted avg | 0.90 | 0.89 | 0.89 | 76 | weighted avg | 0.92 | 0.92 | 0.92 | 76 | | | |

Figura 82: Resultados de los algoritmos en la cuarta iteración

Fuente: Elaboración propia

En la siguiente Tabla 38 se ve en la fila de RF, un cambio en los resultados de los costos relacionados al entrenamiento de Desconexión Programada = 1. Baja de 95 a 92 en Precision y sube de 90 a 92 en Recall o Sensibilidad.

Tabla 38: Precision y recall en la cuarta iteración para desconexión programada

| Algoritmos Clasificando | | | | |
|--------------------------|-----------|--------|-----------------|-----------------|
| Desconexión Programada = | | | 1 | |
| Algoritmo | Precision | Recall | Precision | Recall |
| KNN | 82% | 69% | | |
| NB | 88% | 92% | | |
| RF | 92% | 92% | baja de 95 a 92 | sube de 90 a 92 |

Ahora para la Tabla 39, también cambian los resultados de los costos del entrenamiento para las Desconexiones de curso forzoso = 0. Sube de 90 a 92 en Precision y baja de 95 a 92 en Recall.

Tabla 39: Precision y recall en la cuarta iteración para desconexión de curso forzoso

| Algoritmos Clasificando | | | | |
|-----------------------------|-----------|--------|-----------------|-----------------|
| Desconexión curso forzoso = | | | 0 | |
| Algoritmo | Precision | Recall | Precision | Recall |
| KNN | 72% | 84% | | |
| NB | 91% | 86% | | |
| RF | 92% | 92% | sube de 90 a 92 | baja de 95 a 92 |

De la Figura 83, en la Matriz de confusión, solo vemos cambios en RF.

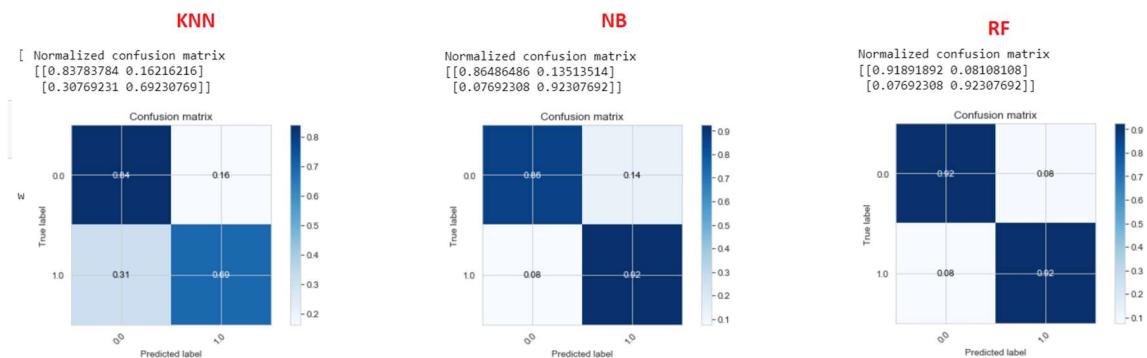


Figura 83: Matriz de confusión con los resultados de los algoritmos en la cuarta iteración.

Fuente: Elaboración propia

En la Tabla 40 se marca en rojo la diferencia de esta cuarta iteración.

Tabla 40: Comparación matriz de confusión entre la segunda, tercera y cuarta iteración

| Comparación de Matriz de confusión | | | | |
|------------------------------------|-----------|----------|----------|-----------|
| KNN | TP | FP | FN | TN |
| Segunda Iteración | 85 | 15 | 21 | 79 |
| Tercera Iteración | 84 | 16 | 31 | 69 |
| Cuarta Iteración | 84 | 16 | 31 | 69 |
| NB | TP | FP | FN | TN |
| Segunda Iteración | 89 | 11 | 0 | 100 |
| Tercera Iteración | 86 | 14 | 8 | 92 |
| Cuarta Iteración | 86 | 14 | 8 | 92 |
| RF | TP | FP | FN | TN |
| Segunda Iteración | 85 | 15 | 0 | 100 |
| Tercera Iteración | 95 | 5 | 10 | 90 |
| Cuarta Iteración | 92 | 8 | 8 | 92 |

De la Figura 84, la curva de ROC en KNN y NB se grafican sin cambios, pero RF se ve sobreajustada en comparación con la tercera iteración, la cual hasta el momento es el mejor resultado.

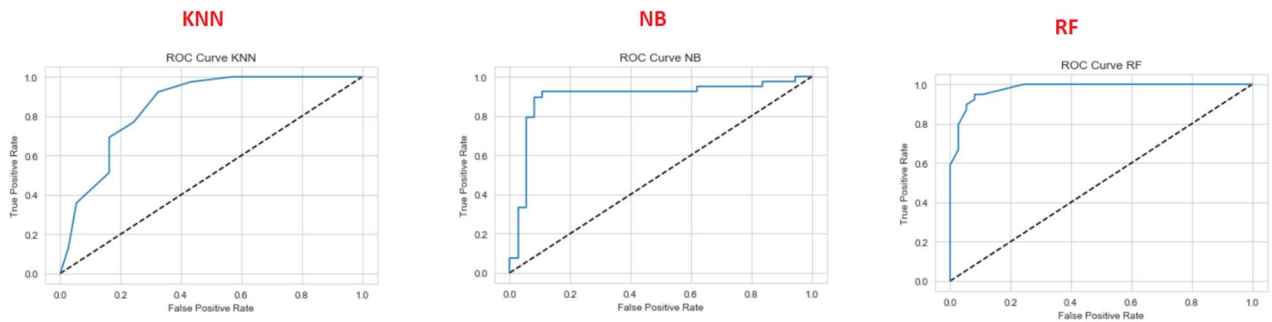


Figura 84: Curva de ROC de los algoritmos en la cuarta iteración

Fuente: Elaboración propia

Quinta Iteración

En esta quinta iteración Figura 85, con el hiperparametro Random Search se mejora el accuracy del anterior Grid Search de 82 a 91.

Random Search

Fitting 5 folds for each of 5 candidates, totalling 25 fits

```
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.
[Parallel(n_jobs=-1)]: Done 19 out of 25 | elapsed: 1.0min remaining: 19.8s
[Parallel(n_jobs=-1)]: Done 25 out of 25 | elapsed: 4.4min finished
```

```
{'C': 0.09789957767367158, 'gamma': 0.01323648994644468, 'kernel': 'linear'}
      precision    recall  f1-score   support

      0.0         0.89         0.92         0.91         37
      1.0         0.92         0.90         0.91         39

 accuracy                   0.91         76
 macro avg                   0.91         76
 weighted avg                0.91         76
```

Figura 85: Resultados de los ajustes de hiperparametros con Random Search en la quinta iteración.

Fuente: Elaboración propia

En la Figura 86 se pueden apreciar los nuevos resultados de los algoritmos en esta quinta iteración.

| KNN | | | | | NB | | | | | RF | | | | |
|---------------------|-----------|--------|----------|---------|---------------------|-----------|--------|----------|---------|---------------------|-----------|--------|----------|---------|
| [[31 6] [12 27]] | precision | recall | f1-score | support | [[32 5] [3 36]] | precision | recall | f1-score | support | [[33 4] [2 37]] | precision | recall | f1-score | support |
| 0.0 | 0.72 | 0.84 | 0.77 | 37 | 0.0 | 0.91 | 0.86 | 0.89 | 37 | 0.0 | 0.94 | 0.89 | 0.92 | 37 |
| 1.0 | 0.82 | 0.69 | 0.75 | 39 | 1.0 | 0.88 | 0.92 | 0.90 | 39 | 1.0 | 0.90 | 0.95 | 0.92 | 39 |
| accuracy | | | 0.76 | 76 | accuracy | | | 0.89 | 76 | accuracy | | | 0.92 | 76 |
| macro avg | 0.77 | 0.77 | 0.76 | 76 | macro avg | 0.90 | 0.89 | 0.89 | 76 | macro avg | 0.92 | 0.92 | 0.92 | 76 |
| weighted avg | 0.77 | 0.76 | 0.76 | 76 | weighted avg | 0.90 | 0.89 | 0.89 | 76 | weighted avg | 0.92 | 0.92 | 0.92 | 76 |

Figura 86: Resultados de los algoritmos en la quinta iteración

Fuente: Elaboración propia

De lo anterior, para entender los costos obtenidos, analizamos la Tabla 41 y baja de 92 a 90 en Precision y sube de 92 a 95 en Recall para la clasificación de Desconexión Programada.

Tabla 41: Precision y recall en la quinta iteración para desconexión programada

| Algoritmos Clasificando | | | | |
|--------------------------|-----------|--------|-----------------|-----------------|
| Desconexión Programada = | | | 1 | |
| Algoritmo | Precision | Recall | Precision | Recall |
| KNN | 82% | 69% | | |
| NB | 88% | 92% | | |
| RF | 90% | 95% | baja de 92 a 90 | sube de 92 a 95 |

En la tabla 42 sube de 92 a 94 en Precision y baja de 92 a 89 para el entrenamiento de Desconexiones de curso forzoso.

Tabla 42: Precision y recall en la quinta iteración para desconexión de curso forzoso

| Algoritmos Clasificando | | | | |
|-----------------------------|-----------|--------|-----------------|-----------------|
| Desconexión curso forzoso = | | | 0 | |
| Algoritmo | Precision | Recall | Precision | Recall |
| KNN | 72% | 84% | | |
| NB | 91% | 86% | | |
| RF | 94% | 89% | sube de 92 a 94 | baja de 92 a 89 |

Para esta nueva iteración solo cambia el RF que se grafica en la Figura 87 con las Matrices de confusión de los tres algoritmos.

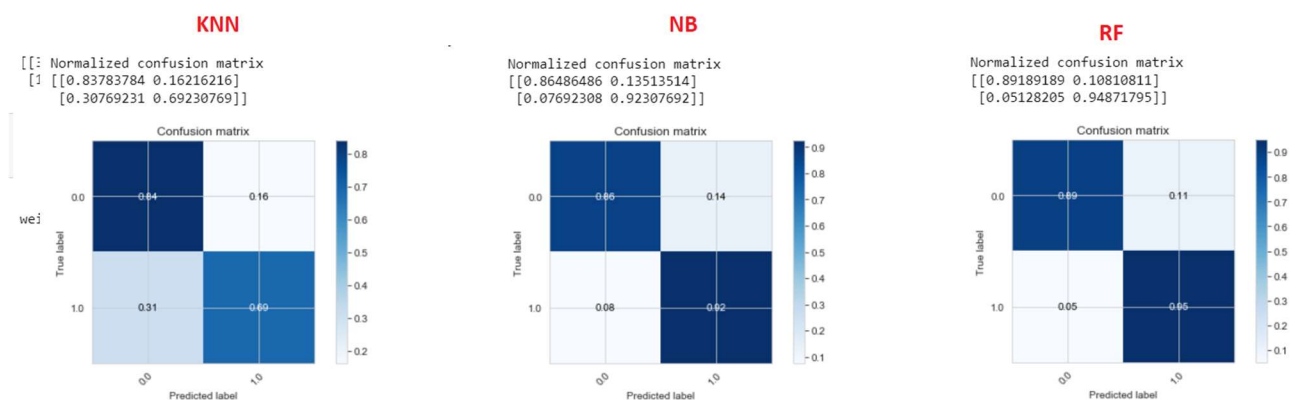


Figura 87: Matriz de confusión con los resultados de los algoritmos en la quinta iteración.

Fuente: Elaboración propia

En la Tabla 43 se marca nuevamente en rojo el cambio de RF en esta quinta iteración.

Tabla 43: Comparación matriz de confusión entre la segunda, tercera, cuarta y quinta iteración

| Comparación de Matriz de confusión | | | | |
|------------------------------------|-----------|-----------|----------|-----------|
| KNN | TP | FP | FN | TN |
| Segunda Iteración | 85 | 15 | 21 | 79 |
| Tercera Iteración | 84 | 16 | 31 | 69 |
| Cuarta Iteración | 84 | 16 | 31 | 69 |
| Quinta iteración | 84 | 16 | 31 | 69 |
| NB | TP | FP | FN | TN |
| Segunda Iteración | 89 | 11 | 0 | 100 |
| Tercera Iteración | 86 | 14 | 8 | 92 |
| Cuarta Iteración | 86 | 14 | 8 | 92 |
| Quinta iteración | 86 | 14 | 8 | 92 |
| RF | TP | FP | FN | TN |
| Segunda Iteración | 85 | 15 | 0 | 100 |
| Tercera Iteración | 95 | 5 | 10 | 90 |
| Cuarta Iteración | 92 | 8 | 8 | 92 |
| Quinta iteración | 89 | 11 | 5 | 95 |

La curva de ROC de la Figura 88, RF se aleja de 1 disminuyendo el sobreajuste de la cuarta iteración, pero gráficamente se puede interpretar que tampoco es mejor que la tercera iteración.

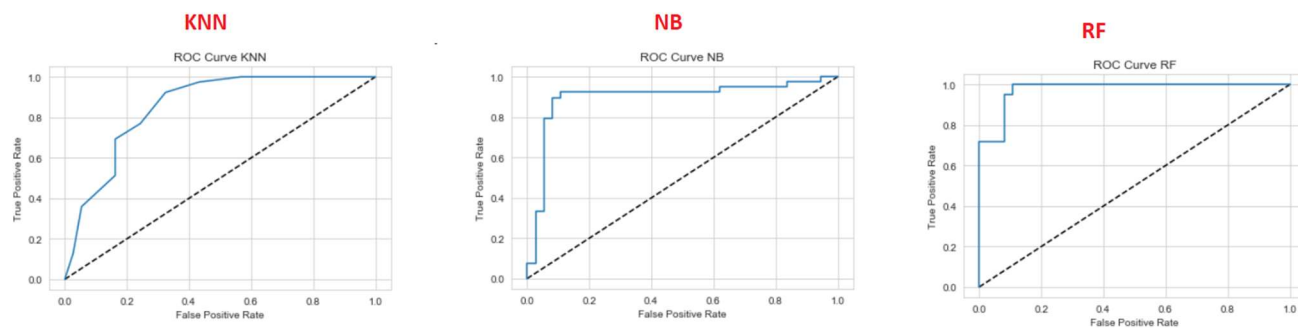


Figura 88: Curva de ROC de los algoritmos en la quinta iteración

Fuente: Elaboración propia