

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

**DESARROLLO MODELO DE MANTENIMIENTO PREDICTIVO PARA
EQUIPOS MINEROS EN DISTRIBUIDORA CUMMINS CHILE S.A.**

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

ALBERTO MATÍAS NICOLÁS SANTIBÁÑEZ ESPINOZA

PROFESOR GUÍA:
EZEQUIEL MUÑOZ KRSULOVIC

MIEMBROS DE LA COMISIÓN:
EDUARDO CONTRERAS VILLABLANCA
ÁNGEL JIMÉNEZ MOLINA
KAREN HERNÁNDEZ LEYTON

CONICYT-PCHA/MagísterNacional/2014 – 22140129

Powered@NLHPC: Esta investigación fue parcialmente apoyada por la infraestructura de supercómputo del NLHPC (ECM-02)
Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02)

SANTIAGO DE CHILE

2016

Resumen

Este proyecto busca la optimización del servicio de mantenimiento en una de las unidades de negocios más importantes de Distribuidora Cummins Chile: el *Mining Group*.

El *Mining Group* concentra aproximadamente el 50% de la venta y el 50% del margen de la empresa entregando servicios a la minería, es por lo anterior, que esta área es considerada estratégica para la búsqueda de oportunidades de mejora a través de sus procesos productivos, además de encontrarse en un periodo de recesión en cuanto a la carga de trabajo por la contracción actual del mercado, lo que genera oportunidades de mejora que deben ser aprovechados de la mejor forma posible.

Basado en las líneas de desarrollo que Distribuidora Cummins Chile S.A. ha definido en su estrategia corporativa, se ha decidido como objetivo la búsqueda de un modelo de predicción de fallas en la flota de motores Cummins que se encuentran en funcionamiento en las distintas faenas mineras a lo largo de Chile, esto permitirá optimizar los procesos productivos relacionados a los servicios de mantención de motores, cuyo efecto, ofrecerá beneficios de forma indirecta a las distintas faenas, siendo su principal consecuencia la búsqueda de disminución de costos de mantenimiento correctivo mediante la identificación de patrones de falla en la flota total de la empresa, lo anterior, desarrollado bajo la metodología de *Macroprocesos* definida por el profesor Oscar Barros.

Dedicatoria

A mí querida familia, mi padre Alberto Santibáñez, mi madre Marisol Espinoza, y mi hermana Carla Santibáñez, quienes siempre me han brindado su cariño, paciencia y apoyo incondicional.

A mi novia, Valeria Cabezas, quién estuvo en momentos críticos de este largo camino, entregándome palabras de aliento, así como críticas constructivas cuando fue necesario, me brindó su paciencia y cariño que fueron fundamentales para no decaer y salir adelante durante estos 3 años. Mi principal apoyo, mi novia, mi amiga y compañera de grandes batallas.

A mis queridos abuelos, que, aunque no se encuentran conmigo hace muchos años, tengo plena convicción de que me siguen acompañando y cuidando tal como siempre lo hicieron. Se les extraña mucho.

Agradecimientos

A mi familia y a mi novia por confiar plenamente en mí, por estar siempre conmigo, en las victorias como en las derrotas, siempre ayudándome a levantar y seguir adelante.

Al profesor Ezequiel Muñoz, por su gran ayuda para concretar este proyecto, por su disposición, por su tiempo y gran paciencia, que fueron fundamentales para lograr este ansiado objetivo.

A Cristian Julio y Patricio Wolf, que directa o indirectamente me ayudaron (en primera instancia) a definir de manera concreta y válida este proyecto de tesis, siendo parte importante del resultado final.

A Ana María Valenzuela y Laura Sáez, quienes siempre me brindaron su apoyo, aconsejándome y guiándome de la mejor forma posible, con su increíble paciencia y disposición durante todos estos años.

Al área del Mining Group de Distribuidora Cummins Chile S.A., en especial a Karen Hernández, por confiar y facilitar todo lo que fue necesario para desarrollar este proyecto. Su apoyo fue fundamental para la definición del proyecto, desarrollo de los pasos intermedios de implementación de esta prueba de concepto y análisis de los resultados finales.

A todas aquellas personas que permitieron el desarrollo de este proyecto.

A todos ellos, muchas gracias.

TABLA DE CONTENIDO

Resumen	2
1 Introducción.....	15
2 Entorno relevante asociado al proyecto.....	16
2.1 Contexto de la Industria	16
2.1.1 Mercado de metales en el mundo	16
2.1.2 Industria Minera en Chile.....	19
2.2 Contexto de la organización.....	22
2.2.1 Komatsu Cummins Chile Ltda.	22
2.2.2 Distribuidora Cummins Chile S.A.	24
3 Planteamiento del proyecto.....	30
3.1 Introducción.....	30
3.2 El Problema.....	30
3.3 Objetivos	32
3.3.1 Objetivo general	32
3.3.2 Objetivos específicos.....	32
3.4 Producto	32
3.5 Alcances.....	33
3.6 Decisiones de negocio	34
4 Marco teórico conceptual	37
4.1 Revisión de la literatura.....	37
4.1.1 Predicción en motores diésel	37
4.1.2 Proceso KDD, <i>data mining</i> y sus implicaciones	41
4.2 Metodología del proyecto.....	43

4.3	Mantenimiento.....	48
4.3.1	Definición de mantenimiento	48
4.3.2	Propósito del mantenimiento	48
4.3.3	Objetivo del mantenimiento	49
4.3.4	¿Qué ocurre cuando se interrumpe un proceso?.....	50
4.3.5	Gestión del mantenimiento	51
4.3.6	Tipos de mantenimiento	56
4.3.7	Ventajas y desventajas.....	60
4.4	<i>Knowledge Discovery in Databases (KDD)</i>	60
4.5	Regresión Logística	63
4.5.1	Objetivos Regresión logística	63
4.5.2	Modelo Logit	63
4.6	Árbol de decisión.....	65
4.6.1	Prueba de Chi-cuadrado (χ^2) (Monge & Pérez, 2002)	67
4.6.2	CHAID (Chi-squared Automatic Interaction Detection)	70
4.7	<i>Support Vector Machine</i>	73
4.8	Comparación de modelos	78
4.8.1	Métricas de evaluación.....	78
4.8.2	Función de costos	80
5	Planteamiento estratégico.....	83
5.1	Visión	83
5.2	Misión	83
5.3	Modelo delta.....	84
5.3.1	Efectividad operacional	85

5.4	Mapa estratégico <i>Mining Group</i>	86
5.5	Modelo de negocio	89
6	Arquitectura de macroprocesos	92
6.1	Modelamiento del Rediseño.....	92
6.2	Macro 1: Cadena de valor	93
6.2.1	Gestión de servicios de mantención y entrega equipos reparados	95
6.3	Lógica solución propuesta	99
6.3.1	Modelos de motores	99
6.3.2	Faenas mineras	100
6.3.3	Frecuencia de fallas	100
6.3.4	Metodología	101
6.3.5	Fuentes de datos	102
6.4	Situación Actual	103
6.5	Lógica implementación	106
6.5.1	Descripción lógicas requeridas.....	106
7	Arquitectura de Sistemas	115
7.1	Diagrama de paquetes	115
7.2	Casos de uso	116
7.3	Diagramas de sistemas y de clases.....	119
7.3.1	Diagrama de secuencia de sistema	119
7.3.2	Diagrama de secuencia de sistema extendido	128
7.4	Diagrama de clases	132
7.4.1	Base de datos.....	133
7.4.2	Diagrama de despliegue.....	133

8	Gestión del cambio	136
8.1	Contexto de la empresa	136
8.2	Creación de sentido y narrativas.....	137
8.3	Factores críticos de éxito	140
8.4	Aspectos a conservar.....	141
8.5	Gestión de poder.....	142
9	Prueba de concepto	144
9.1	Consideraciones particulares de la evaluación	144
9.1.1	Faenas objetivos	144
9.1.2	Faena a estudiar.....	144
9.1.3	Sistema a estudiar.....	146
9.1.4	Muestra y variables a analizar.....	147
9.1.5	Evaluación de resultados	148
9.2	Datos.....	149
9.2.1	Definición descarga de información	149
9.2.2	Definición Target	151
9.2.3	Creación Base de datos	151
9.2.4	Análisis exploratorio de la base de datos	153
9.2.5	Calidad de información agrupada:.....	154
9.3	Árbol de decisión (CHAID Exhaustivo)	154
9.4	Regresión Logística	156
9.5	<i>Support Vector Machines</i>	157
9.5.1	<i>Kernel</i> Lineal.....	157
9.5.2	<i>Kernel</i> Radial	157

9.5.3	<i>Kernel</i> Sigmoidal	158
9.6	Comparación de modelos	158
9.6.1	Análisis de resultados.....	158
9.7	Validación de modelo predictivo	161
9.7.1	Faena y sistema analizado.....	161
9.7.2	Evaluación de resultados y modelo matemático	163
10	Evaluación Económica	166
10.1	Inversiones.....	166
10.2	Beneficios	167
10.3	Costos.....	169
10.4	Gastos.....	169
10.5	Depreciación	171
10.6	Otras consideraciones	172
10.6.1	Horizonte de evaluación.....	172
10.6.2	Impuestos.....	172
10.6.3	Tasa de descuento.....	173
10.7	Construcción Flujo de Caja.....	173
10.8	Análisis de sensibilidad.....	181
11	Generalización de la experiencia	183
11.1	Introducción	183
11.2	Aplicación del <i>framework</i>	183
11.3	Construcción del <i>framework</i>	184
11.4	Aplicación generalización en otros dominios	186
12	Conclusiones	190

12.1	Planteamiento estratégico	190
12.2	Arquitectura de macroprocesos	191
12.3	Prueba de concepto.....	192
12.4	Evaluación económica	194
12.5	Conclusión general	195
13	Bibliografía.....	198
14	Anexos.....	205
14.1	BPMN Determinar motores para mantenimiento	206
14.2	BPMN Determinación de equipos Mantenimiento Predictivo	207
14.3	BPMN Controlar Mantenimiento	208
14.4	BPMN Determinar motores para mantenimiento (actual).....	211
14.5	Selección campos a considerar en base de datos	212
14.6	Análisis gráfico información (Frecuencia valores)	214
14.7	SPSS Statistics input: Árbol de decisión CHAID Exhaustivo	220
14.8	SPSS Statistics output: Árbol de decisión CHAID Exhaustivo.....	224
14.9	SPSS Statistics input: Regresión Logística	227
14.10	SPSS Statistics output: Regresión Logística	229
14.11	The R Project for Statistical Computing SVM Kernel Lineal	235
14.12	The R Project for Statistical Computing SVM Kernel Radial.....	237
14.13	The R Project for Statistical Computing SVM Kernel Sigmoide.....	239
14.14	Narrativa casos de uso: Interacción sistema e Ingeniero Planificación ...	241
14.15	Narrativa casos de uso: Sistema de control	243

INDICE DE GRÁFICOS

Gráfico 2.1: Consumo mundial cobre 2012-2013.....	16
Gráfico 2.2: Crecimiento Anual Producción Mundial Cobre 2.000-2.013.....	17
Gráfico 2.3: Incremento Producción Mundial Cobre.	18
Gráfico 2.4: Contexto Mundial Minería Chilena.....	20
Gráfico 2.5: Ventas históricas participantes de mercado minería	27
Gráfico 2.6: Participación de mercado, periodo 2013.	28
Gráfico 2.7: Participación de mercado histórico principales actores.....	29
Gráfico 3.1: Frecuencia fallas motores.....	31
Gráfico 6.1: Frecuencia fallas motores QSK60 – QSK45	101
Gráfico 9.1: Tiempo vs Frecuencia de falla por faena.....	145
Gráfico 9.2: Tiempo vs Frecuencia de falla por sistema	146

INDICE DE FIGURAS

Figura 2.1: Empresas Holding Komatsu Cummins Chile Ltda	23
Figura 4.1: Metodología Ingeniería Negocios	47
Figura 4.2: Evolución técnicas de mantenimiento.....	49
Figura 4.3: Aspectos a considerar en interrupción de un proceso	50
Figura 4.4: Relación Tipos de mantenimiento con RCM.....	54
Figura 4.5: Proceso KDD.	62
Figura 4.6: Gráfico de datos clasificados en 2 segmentos.....	74
Figura 4.7: Transformación no lineal.....	74
Figura 4.8: Hiperplanos candidatos para datos graficados	75
Figura 4.9: Hiperplano de máximo margen	76
Figura 4.10: Margen blando	76
Figura 5.1: Modelo Delta aplicado a Unidades de negocio.....	85

Figura 5.2: Mapa estratégico área minería	87
Figura 5.3: Modelo de negocios <i>Mining Group</i>	89
Figura 6.1: Arquitectura de macroprocesos <i>Mining Group</i>	92
Figura 6.2: Primer nivel macro 1..	94
Figura 6.3: Apertura Gestión de servicios mantención y entrega equipos.....	95
Figura 6.4: Planificación y control de mantenimiento a equipos mineros	96
Figura 6.5: Planificar Mantenimiento	97
Figura 6.6: Despliegue interfaz de usuario.....	108
Figura 7.1: Diagrama de paquetes	115
Figura 7.2: Interacción sistema e Ingeniero Planificación	116
Figura 7.3: Sistema de Control.....	118
Figura 7.4: Actualización y revisión de información	119
Figura 7.5: Selección tipo de análisis	120
Figura 7.6: Revisión y descarga de resultados.	120
Figura 7.7: Dirección web.....	121
Figura 7.8: Inicio Specto 2.0.....	122
Figura 7.9: Actualización base de datos	122
Figura 7.10: Actualización base de datos (resumen).	123
Figura 7.11: Selección algoritmo data Mining.	124
Figura 7.12: Resumen Resultados	125
Figura 7.13: Procesar más información	125
Figura 7.14: Estimación demanda equipos	126
Figura 7.15: Revisión y determinación de equipos críticos	126
Figura 7.16: Generación de alarmas.....	127
Figura 7.17: Actualización y revisión de información	128
Figura 7.18: Selección tipo de análisis.....	128
Figura 7.19: Revisión y descarga de resultados.	129
Figura 7.20: Estimación demanda equipos	129
Figura 7.21: Revisión y determinación equipos críticos	130

Figura 7.22: Generación de alarmas	130
Figura 7.23: Diagrama de clases	132
Figura 7.24: Relación Entity del sistema	133
Figura 7.25: Diagrama de despliegue	134
Figura 8.1: Actores relevantes implementación proyecto	137
Figura 9.1: Ejemplo planilla descarga información.....	147
Figura 9.2: Variables sistema Specto.....	148
Figura 11.1: Diagrama de clases de control.....	184
Figura 11.2: Diagrama Entity.....	185
Figura 14.1: Proceso Determinar motores para mantenimiento	206
Figura 14.2: Proceso determinación de equipos Mantenimiento Predictivo	207
Figura 14.3: Proceso Controlar mantenimiento.....	208
Figura 14.4: Proceso determinar motores para mantenimiento (actual)	211

INDICE DE TABLAS

Tabla 2.1: Ranking Top 20 Precios del Cobre	19
Tabla 4.1: Elementos de una matriz de confusión	78
Tabla 9.1: Planilla maestra información falla de equipos	150
Tabla 9.2: Balance muestra base entrenamiento 1.....	152
Tabla 9.3: Balance base de datos entrenamiento 2.....	154
Tabla 9.4: Tabla Variables independientes	155
Tabla 9.5: Matriz de confusión Árbol de decisión CHAID Exhaustivo	155
Tabla 9.6: Modelo Regresión Logística	156
Tabla 9.7: Matriz de confusión Regresión Logística	157
Tabla 9.8: Matriz de confusión Kernel Lineal	157
Tabla 9.9: Matriz de Confusión Kernel Radial.....	158
Tabla 9.10: Matriz de Confusión Kernel Sigmoidal	158
Tabla 9.11: Comparación Modelos de Predicción – Métricas	158
Tabla 9.12: Resultados función de costo	161

Tabla 10.1: Impuesto utilidades para empresas con fines de lucro	173
Tabla 10.2: Flujo de caja para proyecto MBE	180
Tabla 14.1: Selección de variables a considerar para análisis	213

INDICE DE ILUSTRACIONES

Ilustración 6.1: Seudocódigo Filtro_data	109
Ilustración 6.2: Seudocódigo función Recepción y almacenaje data	109
Ilustración 6.3: Seudocódigo Envío estado de actualización información.....	110
Ilustración 6.4: Seudocódigo Entrega información software estadístico	111
Ilustración 6.5: Seudocódigo Despliegue de resultados	112

1 Introducción

El crecimiento de la industria minera durante el auge del precio de los metales es algo innegable. Desde el 2004 al 2013, la producción mundial de cobre se elevó en un 24%, las grandes empresas se volvieron enormes tras la gran cantidad de proyectos que las mineras tenían en sus portafolios. La premura de las empresas por incorporar capacidad de producción las llevó a hacer estudios y tomar decisiones apresuradas, generando mayor demanda y precios abultados por equipos, servicios y habilidades, al tiempo que provocó la contratación de profesionales sin experiencia o mal capacitados para administrar proyectos y operaciones, tanto para empresas mineras como para proveedores de servicios. Estos factores hicieron que los costos de operación alcanzaran niveles históricos, en muchos casos excediendo por amplia diferencia las estimaciones originales, por lo que el rendimiento del capital empleado se contrajo en toda la industria de un 24% en 2006 a un 10% en 2012. El crecimiento desmedido de las empresas debido a la demanda existente, generó que se expandieran de forma muy compleja, apareciendo dentro de su estructura lo opuesto a estandarización, integración de los procesos y buenas prácticas que las empresas deben considerar para ser competitivas.

2 Entorno relevante asociado al proyecto

2.1 Contexto de la Industria

2.1.1 Mercado de metales en el mundo

La estabilización del crecimiento de China más el fortalecimiento de las actividades económicas en países desarrollados como Estados Unidos y países del centro de Europa permitió mantener una demanda robusta por *commodities*, incluyendo los minerales. Lo anterior generó, para el caso del cobre, que durante el 2013 existiese un superávit de oferta de este mineral, y por ende la baja de su precio cuya tendencia se espera que continúe en esa dirección. Pese a lo anterior, los precios siguen siendo atractivos para las faenas mineras en Chile, convirtiéndose en prioridad la identificación de oportunidades de mejora en los procesos productivos para continuar entregando valor agregado en los países donde se encuentran operando, buscando ventajas competitivas.

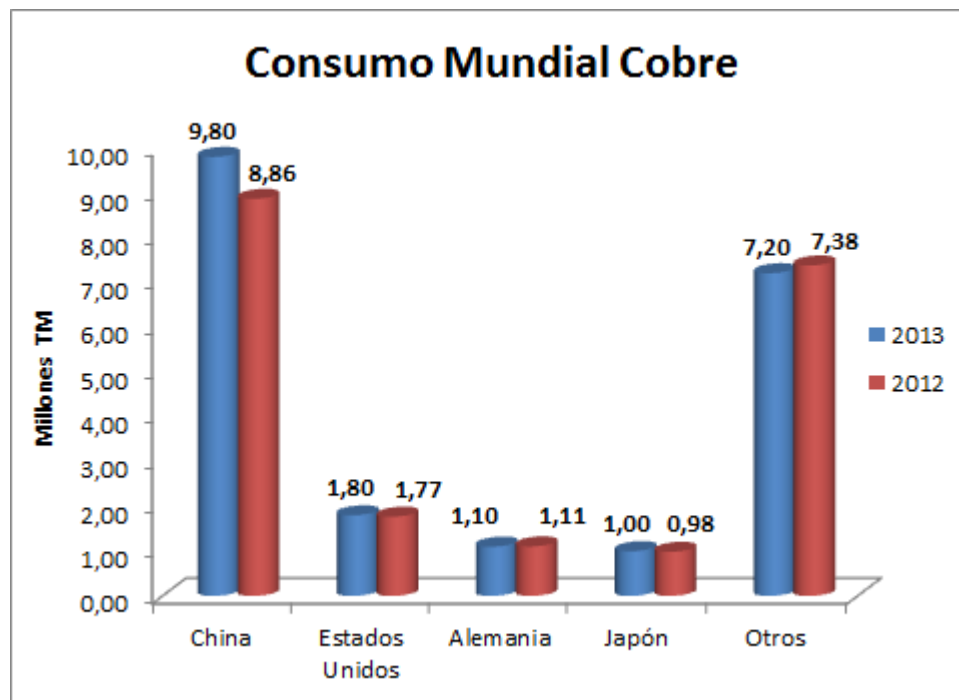


Gráfico 2.1: Consumo mundial cobre 2012-2013. Fuente: Consejo Minero, 2014.

Del Gráfico 2.1 se puede inferir que uno de los actores importantes para la determinación del precio del cobre es China, quién se convierte en un factor primordial para países cuya evolución económica depende de mercados *commodities*, como lo es Chile, sin dejar de mencionar que China es un exportador de productos terminados hacia países como Estados Unidos y países de la Euro-zona, cuya reactivación afectará al mercado de metales (cobre) de forma indirecta. Para metales como oro, plata y molibdeno, se ven afectadas principalmente por la economía de China, para lo cual no se esperan grandes cambios durante el 2013 – 2014 (Consejo Minero, 2014).

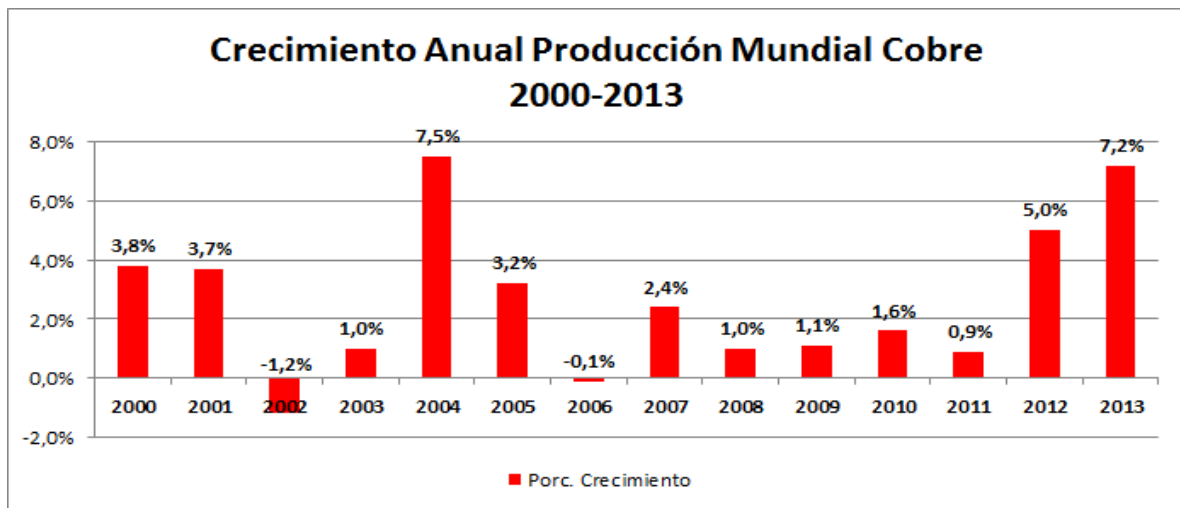


Gráfico 2.2: Crecimiento Anual Producción Mundial Cobre 2.000-2.013. Fuente: Consejo Minero, 2014.

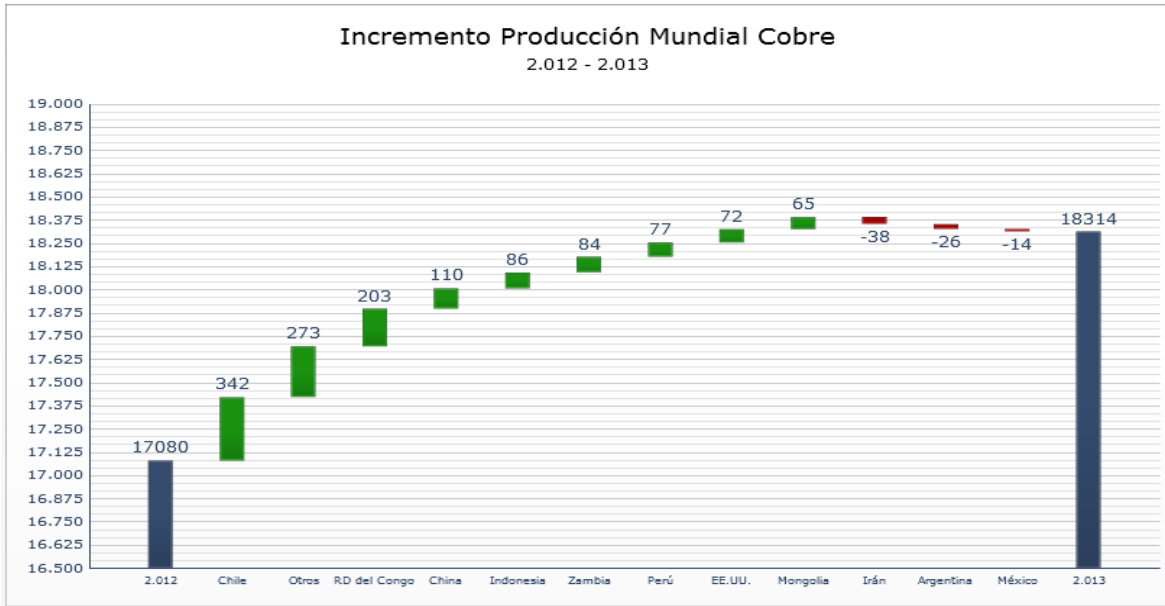


Gráfico 2.3: Incremento Producción Mundial Cobre. Fuente: Consejo Minero, 2014.

Además, los productores de cobre, en el Gráfico 2.2, durante el 2.013 la producción mundial tiene un crecimiento de un 7,2% con respecto al 2.012, siendo este el mayor crecimiento desde el año 2.004, donde Chile mantiene su liderazgo en aporte a la producción mundial, seguida de RD del Congo, China e Indonesia. En el caso del oro, este mercado se encuentra liderado por China, seguido de Australia y Estados Unidos, la producción de plata se obtiene en mayor parte como producto secundario de la extracción de otros metales como oro, cobre, plomo y zinc, y finalmente el molibdeno cuyo 50% de producción nace de la extracción de cobre, sin embargo, China y Estados Unidos lideran la extracción de este metal.

Para el precio del cobre, a continuación se muestra una tabla con el ranking de los 20 mejores precios en la historia:

Ranking	Año	USD de 2013xLibra
1	1966	424,2
2	2011	404,8
3	1969	379,7
4	2007	379,7
5	2006	375,8
6	2010	375,7
7	1965	367,8
8	2012	362,5
9	1973	362,4
10	1974	360,6
11	1970	353,9
12	2008	335,5
13	1968	334,7
14	2013	332,1
15	1967	311,1
16	1955	304,5
17	1964	282,6
18	1956	275,5
19	2009	273,6
20	1971	262,8

Tabla 2.1: Ranking Top 20 Precios del Cobre. Fuente: Consejo Minero.

El precio a nivel mundial del cobre en el 2013 bajó con respecto al 2011 de 404,8 Usd/Lb en comparación a 332,1 Usd/lb respectivamente, donde el precio continua siendo interesante para Chile ya que aún existe cierto rango donde puede buscar oportunidades de mejora y mantener su liderazgo mundial dentro de este mercado.

2.1.2 Industria Minera en Chile

Chile es un país donde la industria minera se basa principalmente en la extracción del cobre, sin embargo, existen otros metales en los cuales Chile tiene participación y que tienen bastante potencial, estos se muestran en el siguiente gráfico:

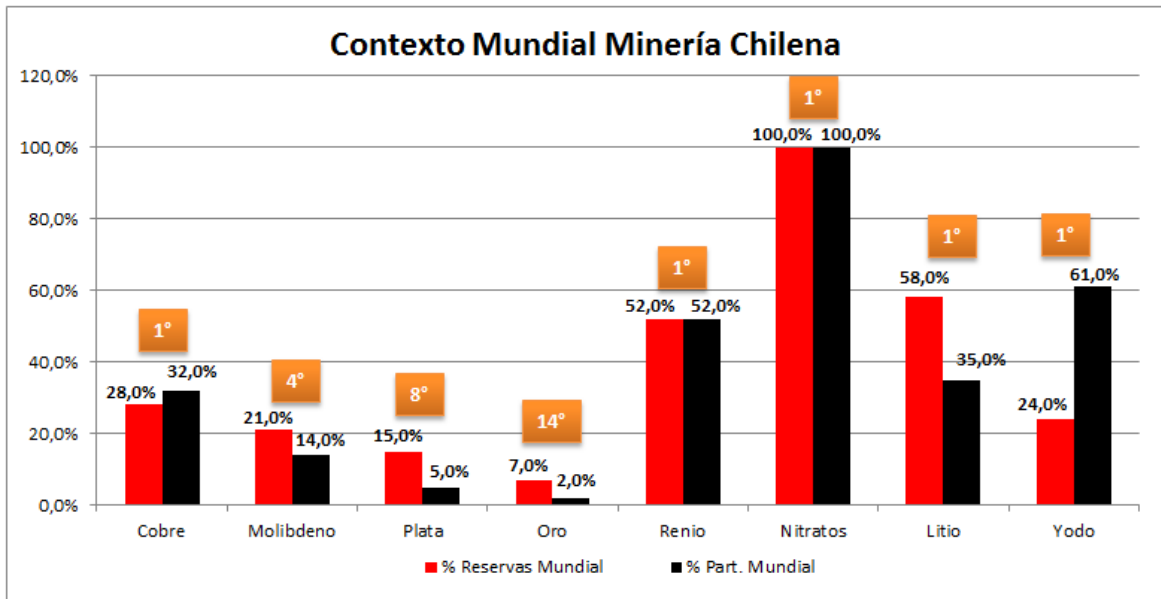


Gráfico 2.4: Contexto Mundial Minería Chilena. Fuente: Consejo Minero, 2014.

El número sobre cada metal indica la posición a nivel mundial en cuanto a producción, para Chile se puede ver que para el cobre, renio, nitratos, litio y yodo está posicionado en primer lugar, sin embargo, como ejemplo, en la producción de oro, Chile se encuentra en la posición 14° considerando que es la cuarta mayor reserva a nivel mundial, superando incluso a Perú y México.

Cuando se habla del principal metal explotado en Chile (que es el cobre), es importante hablar de las distintas faenas mineras. Durante el 2013, destaca el aumento de producción de cobre en Collahuasi con un 58% (+162 mil TM), esto se produjo por la superación de dificultades en el funcionamiento operacional durante años anteriores, por otro lado, le siguen Candelaria con un 37% (45 mil TM) y Escondida con un 11% (118 mil TM).

2.1.2.1 Indicadores de competitividad

Uno de los indicadores más importantes es el costo de producción de cobre en Chile y en el resto del mundo. Se establece que para hacer comparable estos montos es considerado el costo de operación, *cash cost*, depreciación, intereses y costos indirectos. Del análisis realizado por el Consejo Minero (2014), existen 2 tramos bien definidos donde el comportamiento cambia con respecto al resto del mundo desde 2003 al 2013; desde el 2003 al 2008, Chile mantuvo sus costos más bajos, en promedio un 10,7% por debajo de los demás países mineros, sin embargo, desde el 2009 en adelante los costos aumentaron en un 11,1% más altos, y es esto lo que actualmente tiene a Chile en un escenario de desventaja competitiva. Otros factores de la pérdida de competitividad es la baja en la ley del mineral, de 1,0 en 2004 a 0,7 en 2013, cuya tendencia en el tiempo indica que continuará decreciendo.

Otro indicador importante es el costo de inversión en proyectos (*Capex*) que en 2013 alcanzó USD 20.891 por tonelada de cobre fino, cuyo valor es el más alto de la historia. Actualmente las empresas están revisando e intentando revertir esta situación, sin embargo, queda mucho por hacer.

Existe una brecha en cuanto a productividad laboral desde el año 2006, el cual se sostuvo en el tiempo debido al aumento en la brecha de sueldos, que durante el 2013 intentó revertirse con el aumento de la producción de cobre en un 6,3% y una caída en el empleo directo de la minería de un 3,4%, con lo que la productividad laboral pudo aumentar, sin embargo, las remuneraciones siguieron en aumento, por lo cual esta brecha se encuentra lejos de ser cubierta. Por otro lado, la comparación de la mano de obra en Chile con respecto al resto del mundo (según estudio realizado por la consultora internacional McKinsey en el 2012) puede variar entre un 33% y un 188% más alto que en otros países. (Consejo Minero, 2015)

Por último, un factor muy relevante que considera aproximadamente el 20% de los costos de extracción son los costos de energía eléctrica, ya que según el mismo

estudio antes mencionado el costo de la energía en Chile oscila entre un 67% y 91% más alto que en el resto del mundo.

De lo antes descrito, Chile debe trabajar medidas en cuanto a la recuperación de ventaja competitiva, o en una primera instancia, “disminuir” estas desventajas que en el largo plazo podrían ser perjudiciales para los planes de desarrollo del país.

2.2 Contexto de la organización

Distribuidora Cummins Chile S.A. se encuentra dentro del holding Komatsu Cummins Chile Ltda., el cual será descrito en el siguiente punto.

2.2.1 Komatsu Cummins Chile Ltda.

El origen de Komatsu surge de la empresa Takeuchi Mining Industry (TMI), empresa japonesa fundada en 1894 por Meitaro Takeuchi. Takeuchi percibió que si el negocio fundamental de TMI era la minería, lo lógico era concentrarse en esa actividad, desligándose de la fabricación de máquinas para el movimiento de tierra que la misma TMI llevaba a cabo para apoyar sus operaciones y entregar a un tercero la responsabilidad de la fabricación de tales máquinas.

Con esa decisión, en Enero de 1917, nace Komatsu Iron Works. El 13 de mayo de 1921, separó Komatsu Iron Works de su compañía fundando la que hasta hoy sería Komatsu Ltda.

Bajo un concepto muy similar se desarrolla la historia de Cummins. La formación de Cummins Engine Corp. el 3 de febrero de 1919, reunió recursos poco comunes. William Glanton W.G. Irwin - un exitoso banquero e inversionista de Columbus aportó el capital inicial.

El nombre de la nueva compañía viene de Clessie Lyle Cummins, quien era un mecánico inventor autodidacta. Los Irwins lo contrataron en 1908 para manejar y mantener sus automóviles, y después lo iniciaron en el negocio como un mecánico de autos.

Durante la Primera Guerra Mundial, Clessie operó una tienda de maquinaria que creció gracias a los contratos del gobierno. En aquel entonces, él estaba convencido que una tecnología del motor inventada por Rudolph Diesel en 1890 - aunque aún no probada comercialmente - tenía una gran promesa para su economía de combustible y durabilidad. Para asegurar su entrada al negocio de motores, Cummins consiguió los derechos de fabricación de un licenciado holandés llamado Hvid.

De estas dos ideas empresariales con el tiempo se deriva la formación en el año 1999 de Komatsu Cummins Chile Ltda.

Desde esa fecha se han desarrollado múltiples negocios en las industrias Mineras, Construcción, Industrial y Forestal a través de sus distintas filiales.

El holding cuenta con las siguientes empresas que lo conforman:

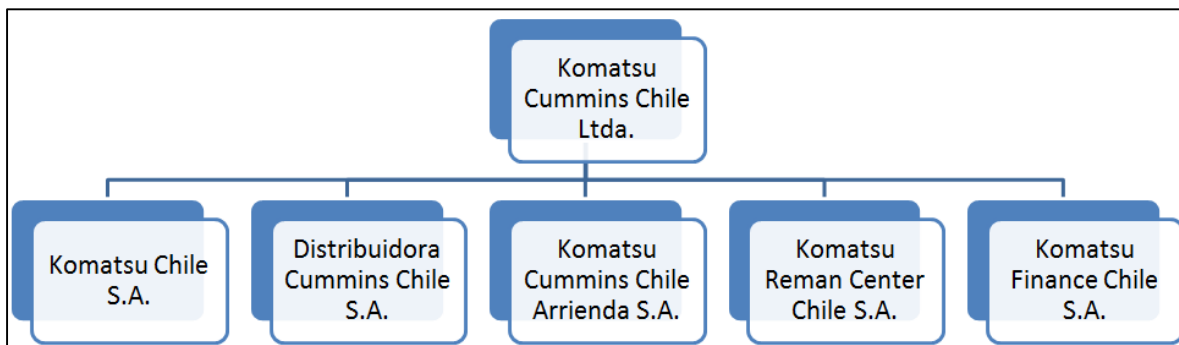


Figura 2.1: Empresas Holding Komatsu Cummins Chile Ltda. Fuente: Elaboración propia.

Komatsu Cummins Chile S.A. entrega servicios a las distintas empresas del holding, los cuales se resumen en:

- Logística
- Recursos humanos
- Finanzas y contabilidad
- Servicios Compartidos
 - Control de gestión corporativo

- Infraestructura
- Servicios Generales

Son estos 4 servicios principales los que el holding suministra a las distintas filiales trabajando como ente controlador de estos recursos de forma transversal.

A continuación se entregan antecedentes de Distribuidora Cummins Chile S.A., filial del holding donde será desarrollado este proyecto de memoria.

2.2.2 Distribuidora Cummins Chile S.A.

Líder en Chile en comercialización de motores diésel participa en variados mercados dentro del país como generación de energía, transporte terrestre, transporte marítimo, alta minería, entre otros. Distribuidora Cummins Chile S.A. es representante oficial de Cummins Inc. en Chile hace más de 60 años, cuya casa matriz se encuentra en Columbus, Indiana, Estados Unidos. Cummins Inc. fue fundada en el año 1919, logrando prestigio mundial por sus altos estándares de calidad y tecnología en el diseño de sus motores, cumpliendo con altas exigencias, destacando su espíritu innovador, constante mejora en sus procesos y preocupación con el medio ambiente.

En 1999 se forma el *joint venture* entre Komatsu (compañía de origen japonés y de enorme prestigio en la fabricación de equipos para minería, construcción y forestal) y Cummins Inc. formando así el holding Komatsu Cummins Chile Ltda., cuya alianza estratégica permitió una rápida y agresiva entrada en la industria de la alta minería en Chile, convirtiéndose en uno de los actores principales dentro de este mercado en el país.

En Chile, la empresa ha desarrollado variadas líneas de negocio, de las cuales destacan: venta de motores, venta de repuestos y filtros, soporte técnico, venta y arriendo de grupos electrógenos, venta de servicio técnico y venta de servicios a la minería.

A continuación se mencionarán de forma muy general las líneas de negocio que son desarrolladas en la empresa.

2.2.2.1 Unidades de negocio

Dentro de la estructura interna de la empresa, existen líneas de negocio que son las unidades fundamentales en las cuales Distribuidora Cummins Chile S.A. basa sus estrategias de desarrollo y expansión.

- **Engines** (motores): Unidad dedicada a la venta directa de motores a los clientes, esta unidad se encarga del envío de la orden de compra a Cummins Inc. (o donde exista stock), entrega parámetros de configuración a fábrica y coordinación de entrega del motor con el cliente.
- **Shops and Services** (talleres de servicio mantenimiento): Unidad dedicada a la entrega de servicios de mantenimiento a todos los equipos de Distribuidora Cummins Chile S.A. y algunas marcas de terceros. Esta función es realizada en las sucursales a lo largo de Chile como también es realizada en terreno.
- **Parts & Filters (repuestos y servicios)**: Unidad dedicada a la venta de repuestos y filtros para insumos de mantención en equipos Cummins como también para algunos modelos de terceros.
- **Mining Group (minería)**: Esta unidad se dedica a satisfacer las necesidades de un tipo especial de clientes, los cuales se encuentran en la industria de la minería. Debido a que es en esta unidad donde será desarrollado el proyecto de memoria, será desarrollada en el siguiente punto con más detalle.

2.2.2.2 Mining Group (minería)

El área de minería de Distribuidora Cummins Chile S.A. (desde ahora en adelante lo llamaremos *Mining Group*) es crítico debido a su importancia financiera dentro de la empresa, ya que aproximadamente el 50% de la venta se concentra en esta área – ocurriendo algo similar con el margen bruto.

Para contextualizar mejor esta unidad, serán descritas las actividades que son desarrolladas durante el servicio que es entregado.

El servicio que el cliente adquiere cuando contrata los servicios del *Mining Group* principalmente es disponibilidad de operación de los equipos (en este caso los motores diésel) que tienen operando en sus faenas. Para lo anterior existen varios modelos de tarificación, donde los servicios entregados se pueden resumir de la siguiente forma:

- Servicio de mantenimiento preventivo disponible las 24 hrs. a equipos mineros en faena.
- Servicio de mantenimiento correctivo disponible las 24 hrs. a equipos mineros en faena.
- Asesoramiento por personal experto a mecánicos en faena.
- Gestión, configuración e instalación de motores de reemplazo en faena.
- Entrega de insumos mantenimiento preventivo y correctivo.

Siendo estos los servicios que tienen mayor importancia dentro de la cadena de valor de ésta área.

Dentro del *Mining Group* se encuentra el área de soporte técnico, quienes se encargan de velar por toda la flota de equipos que sus clientes utilizan, administran todas las mantenciones de los equipos con los que tienen contrato de acuerdo a la información técnica que reciben desde los distintos sistemas de monitoreo, como también desde el cliente, entregando soporte a través de las personas que se encuentran en faena y

además desde Santiago a través del Subgerente de Soporte Técnico y/o del área de monitoreo.

A continuación se muestra el Gráfico 2.5, donde se puede ver la venta de motores desde el 2007 al 2013 de las empresas que se encuentran dentro del mercado minero:

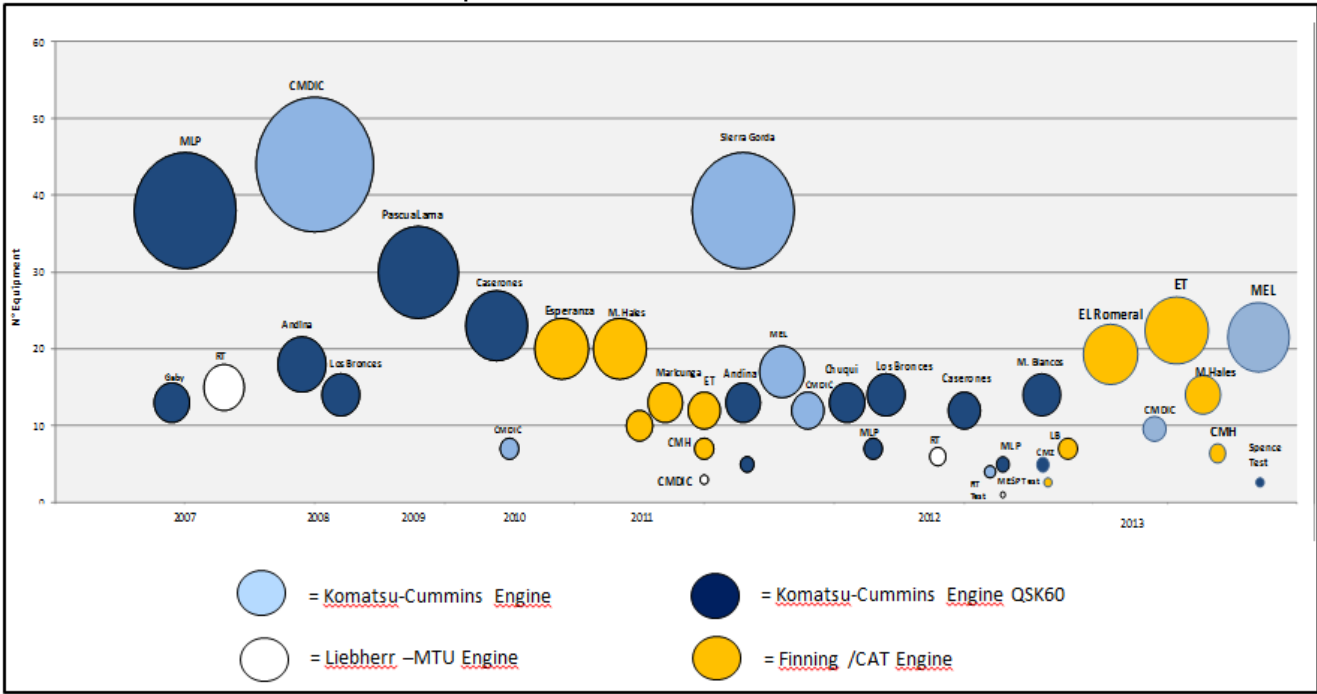


Gráfico 2.5: Ventas históricas participantes de mercado minería desde 2007 a 2013. Fuente: Elaboración propia.

Si bien, se menciona dentro del punto 2.2.2.1 que la unidad de negocio que vende motores es *Engines*, la población de motores dentro de las faenas mineras nos permite medir la participación de mercado debido a que a la fecha las faenas prefieren trabajar directamente con la marca oficial del motor en el país más que con terceros, manteniendo la garantía que otorga fábrica a sus motores.

En el gráfico se puede ver que la unión de Komatsu Chile S.A. y Distribuidora Cummins Chile S.A. es muy efectivo, logrando obtener mercado de manera agresiva y sostenida desde el 2007 hasta mediados de 2011, se refleja una disminución en la inversión por

parte de las empresas mineras entre el 2011 y 2013, siendo Finning quién asume una estrategia comercial agresiva logrando adjudicarse algunos proyectos durante el 2013. Lo anterior permite establecer la siguiente participación de mercado:

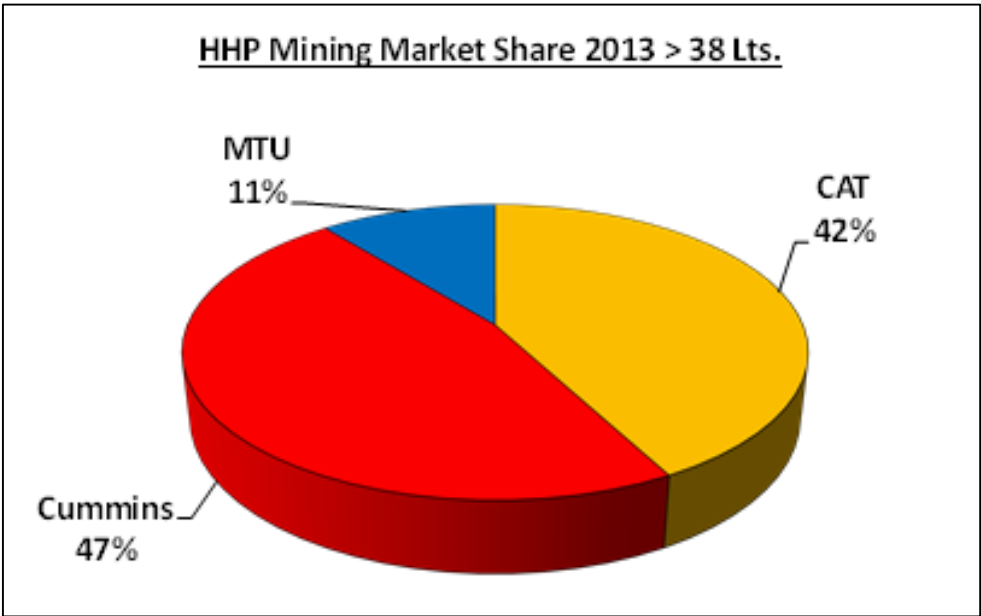


Gráfico 2.6: Participación de mercado, periodo 2013. Fuente: Elaboración propia.

A pesar de que la estrategia adoptada por el *joint venture* Komatsu Cummins Chile Ltda. fue potente durante el periodo 2007 – mediados 2013, este crecimiento sostenido en el tiempo fue detenido por la situación actual del precio del cobre, esto generó que faenas mineras postergaran sus proyectos de inversión, provocando un receso en la adquisición de motores y maquinaria pesada, en este escenario, la participación de mercado prácticamente tiene 3 actores principales, siendo liderado por Distribuidora Cummins Chile S.A. (47%) en comparación a su competidor directo Caterpillar (42%). Otra visión del avance de la participación de mercado se refleja en el siguiente gráfico:

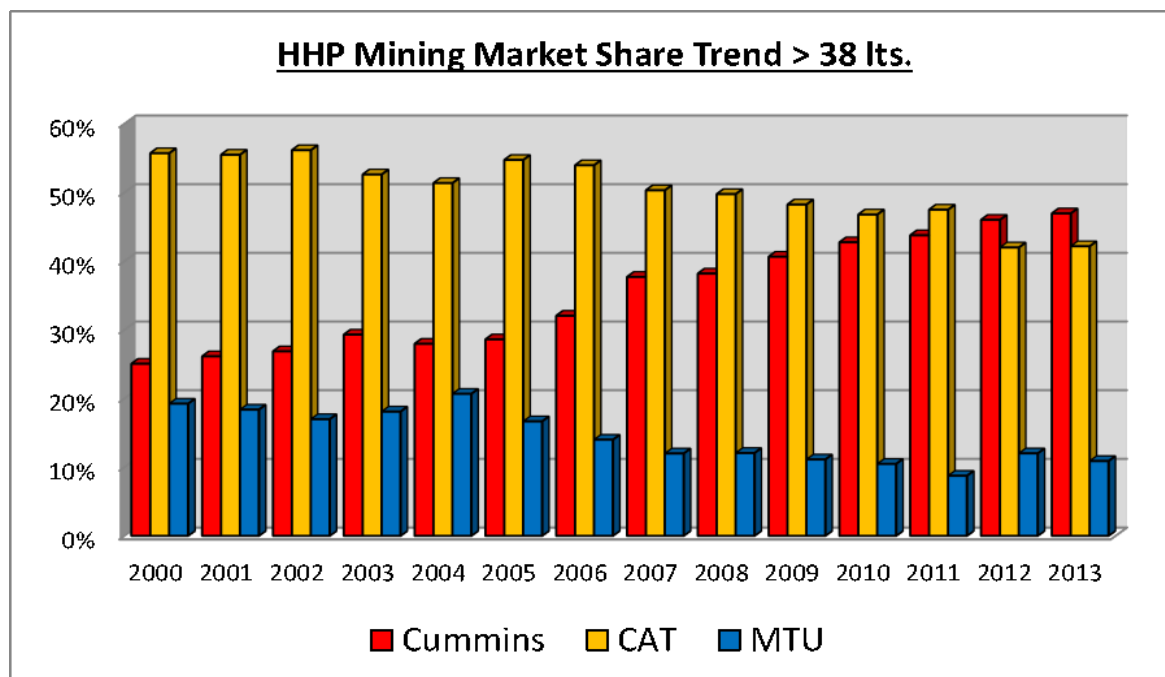


Gráfico 2.7: Participación de mercado histórico principales actores. Fuente: Elaboración propia.

En el Gráfico 2.7 se puede apreciar la potencia que generó el *joint venture* entre Distribuidora Cummins Chile S.A. y Komatsu Chile S.A. dentro de este mercado que inicialmente tenía sólo un actor importante que era Caterpillar.

Los resultados obtenidos entre el 2007 y el 2011 se deben principalmente a los lineamientos estratégicos de Komatsu Limited y Cummins Inc. los cuales apuntaban a abrir mercado en América del Sur, junto con las gestiones de coordinación realizada por el holding Komatsu Cummins Chile Ltda.

Desde este punto en adelante, el lector debe considerar que toda la información en este documento se referirá al *Mining Group*, que será la unidad de negocio donde será desarrollado el proyecto de memoria.

3 Planteamiento del proyecto

3.1 Introducción

De acuerdo a los intereses estratégicos del *Mining Group* y conversaciones previas con integrantes de esta área, el proyecto estará encargado principalmente en mejorar uno de sus procesos críticos, que involucra la parte más importante de su servicio de mantenimiento, que es la entrega de una herramienta para facilitar la detección y establecimiento de patrones de falla para motores Cummins dentro de la flota que administra directamente el *Mining Group*, o aquellos equipos cuyo cliente solicite, esto traerá como consecuencia que el valor agregado percibido por el cliente aumente, debido a la información que entregará la solución desarrollada para tomar mejores decisiones en cuanto a planificación de servicios de mantenimiento, esto será realizado mediante el uso de herramientas de *Data Mining* para generar de forma posterior patrones de fallas de acuerdo a información que entregan los distintos sensores que tienen los equipos, creando y/o modificando los procesos necesarios para su implementación, permitiendo entregar al *Mining Group* herramientas que facilitarán la gestión de nuevas características en su servicio, siendo el principal, la capacidad de predicción de fallas en motores Cummins.

3.2 El Problema

El servicio que entrega el *Mining Group* a sus clientes asegura horas de operación de los equipos en las faenas, donde el precio del servicio considera las mantenciones preventivas que se encuentran establecidas por pautas entregadas desde fábrica (manteniendo la garantía de los equipos), incluyendo un tipo de reparación que es drástico en cantidad de tiempo, insumos utilizados y costos involucrados, al cual llamaremos *Overhaul*.

El gran problema de este tipo de tarificación es que no considera costos ni tiempos de reparación provocados por fallas imprevistas (mantenimiento correctivo), por lo tanto, mientras menos fallas imprevistas ocurran, o dicho de otra forma, mientras más rápido

se pueda detectar una posible falla, el tiempo de reparación y los costos en insumos, dependiendo de lo que sea detectado puede disminuir los costos de forma considerable, e incluso, podría aumentar el tiempo entre una mantención preventiva y otra.

A continuación se muestra un gráfico con las fallas imprevistas más frecuentes en los motores modelo QSK60 y QSK45 (que son los motores que utilizados en mayor medida por las faenas):

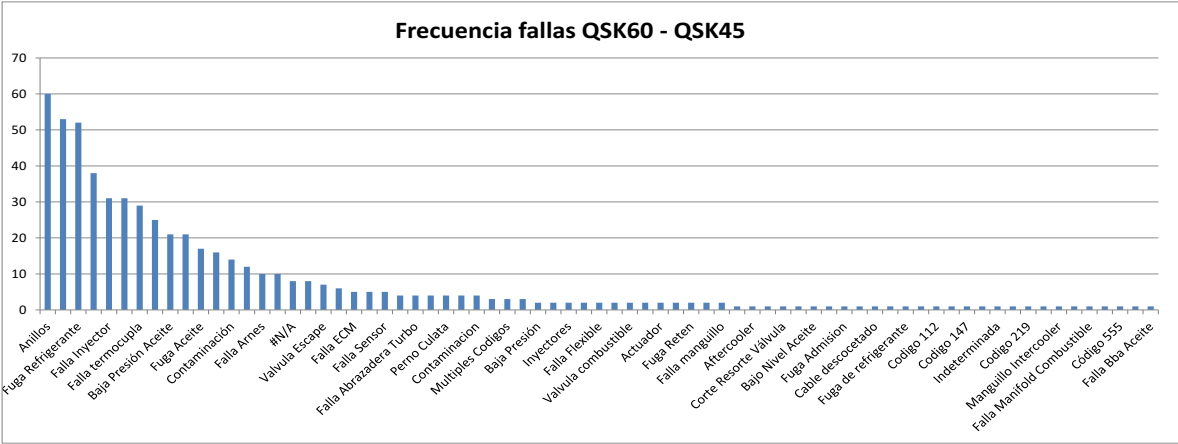


Gráfico 3.1: Frecuencia fallas motores QSK60 – QSK45 periodo Diciembre 2013 – Agosto 2014. Fuente: Elaboración propia.

Dentro de las grandes consecuencias de este proyecto de memoria, será establecer cuál de las fallas imprevistas mencionadas en el Gráfico 3.1 tiene mayor influencia en la realización de *Overhaul* imprevistos, por otro lado, la realización de los *Overhaul* en promedio tienen un costo aproximado de USD 365.000 por motor, por lo que el objetivo será minimizar este tipo de falla.

3.3 Objetivos

3.3.1 Objetivo general

Disminución de fallas imprevistas (*overhaul*) dentro de las faenas donde Distribuidora Cummins Chile S.A. se encuentra presente, estableciendo procesos de predicción de fallas que sean capaces de proveer información oportuna para el proceso de planificación de los servicios de mantenimiento, esto, basado en el rediseño y creación de procesos, además, del uso y establecimiento de herramientas tecnológicas necesarias, generando ahorros anuales de al menos USD 365.000 (disminuir en 1 *overhaul* al año).

Para alcanzar este objetivo se enumerarán los objetivos específicos en la siguiente sección.

3.3.2 Objetivos específicos

- Establecimiento de faenas objetivo para análisis de fallas.
- Definición de sistema de motor a analizar de acuerdo a clasificación del área de estadística del *Mining Group*.
- Selección de campos relevantes para el análisis de la información.
- Definición modelo matemático a utilizar asociado al problema en estudio.
- Diseño y desarrollo del modelo predictivo, el cual debe considerar las tolerancias de error que serán definidos con el Jefe de estadística del *Mining Group*.
- Diseño de funcionalidades del rediseño de procesos, el cual facilitará la ejecución y entrega de resultados al ingeniero de planificación.

3.4 Producto

Para cumplir con los objetivos descritos, serán añadidos los siguientes elementos dentro del proceso de planificación de mantenimiento del *Mining Group*, los cuales serán tratados con mayor profundidad en capítulos posteriores:

- Rediseño de proceso de planificación de mantenimiento. Es necesario rediseñar el proceso debido a que debe ser agregada la lógica automatizada de clasificación de motores que serán entregados como propuesta para recibir mantenimiento preventivo.
- Modelo predictivo de falla de motores, utilizando lógicas de clasificación mediante aprendizaje de modelos supervisados utilizando datos del sistema de monitoreo remoto existente, de esta forma se podrá optimizar el uso de recursos.
- Definición frecuencia de análisis para alertas proactivas.
- Definición muestra a analizar junto con lógica de limpieza de la *data*.

3.5 Alcances

- El desarrollo de la solución sólo involucrará al *Mining Group*, donde el usuario del sistema serán funcionarios del área de estadística.
- Será diseñada la interfaz gráfica del usuario, junto con la lógica que deberá seguir la solución desarrollada.
- El desarrollo de este proyecto se enfoca en el motor de los equipos mineros y sus sensores, no en sus periféricos.
- La estimación de demanda de mantenimiento predictivo estará basada en un modelo matemático que será diseñado e implementado a través de un software estadístico integrado con la solución web desarrollada.
- Los resultados del mantenimiento predictivo no serán traspasados en esta fase de forma directa al cliente, sino que serán utilizadas para gestiones internas.
- Los resultados obtenidos en este proyecto estarán dentro del marco de prueba de concepto.

3.6 Decisiones de negocio

De acuerdo a las características del proyecto, este contribuirá a la toma de decisiones en los siguientes aspectos:

- **Definición y detección de patrones de falla.**

Al contar con procesos de predicción en producción, permitirá detectar equipos que serán posibles candidatos de fallas, con esta especificación, el área de monitoreo podrá tomar esta muestra y analizar con otras herramientas de *Data Mining* (segmentación y/o clusterización) para generar información de la flota. La herramienta que será entregada en este proyecto permitirá que a futuro puedan crear su base de datos de fallas y poder clasificar de forma eficiente las fallas en el futuro.

- **Modificación plan de producción.**

Cuando el usuario ejecuta el modelo predictivo o el sistema lo ejecuta por una programación previa, lo que está buscando es que el sistema arroje candidatos de falla para evitar una reparación mayor, en el caso de que sea aceptada esta propuesta, será necesario avisar al área encargada de gestionar las distintas mantenciones, de tal forma que considere estos equipos en la rutina de mantenimiento.

- **Priorización de faenas.**

Distribuidora Cummins Chile cuenta con operaciones en la mayoría de las faenas a lo largo del país, sin embargo, existen faenas donde los equipos mineros tienen un mejor comportamiento con respecto a otras, y esto se debe a las condiciones ambientales y al entorno de trabajo donde se encuentran los equipos. Para efectos de este trabajo, las variables a considerar para la priorización de las faenas importantes están definidas por la cantidad de horas en mantención, donde los

factores que serán determinante son las horas de reparación imprevista (RI), reparación programada (RP) y las clasificaciones que sean consideradas continuaciones de estos trabajos, continuación reparación imprevista (CRI) y continuación reparación programada (CRP).

Consideraciones Particulares

- **Faenas:**

Para efectos de desarrollo de modelos de patrones de falla, se considerará faena Collahuasi como la faena donde existe mayor frecuencia de fallas.

- **Equipos:**

Las faenas poseen modelos de motor QSK45 y QSK60, de los cuales se analizará la cantidad de existente en la flota de cantidad de equipos y de frecuencia de falla para definir el modelo a utilizar, buscando patrones de comportamiento similares entre ellos maximizando el impacto del análisis y los resultados.

- **Sistema de monitoreo remoto:**

Existe un sistema de monitoreo remoto llamado Specto, el cual envía información de los sensores del motor cada 5 segundos hacia un *data warehouse*, se encuentra operativo en casi un 70% de las faenas donde Cummins tiene flota. Actualmente este sistema ha tenido problemas en cuanto a la gestión de su base de datos, sin embargo, se llegó a un acuerdo de trabajo.

- **Bitácora de mantenciones:**

Existe un proyecto en paralelo dentro del área el cual busca mantener información de todos los equipos de la flota, por lo que en el futuro se busca integrar esta información con el modelo predictivo y viceversa.

4 Marco teórico conceptual

Serán detallados los fundamentos que soportan este trabajo de tesis, estos permitirán entregar al lector una visión más amplia de las herramientas y supuestos utilizados.

Las principales áreas de desarrollo dentro de esta tesis son: metodología de análisis que permite el desarrollo de este proyecto, teoría del contexto donde se encuentra inmerso el proyecto y que es muy importante de mencionar, que es el Mantenimiento y sus distintas ramas, junto con lo que engloba la solución que se busca desarrollar con esta tesis que utiliza metodología *Knowledge Discovery in Databases* (KDD), abarcando principalmente lo que se refiere a *Data Mining* para la detección de patrones de falla en los motores de la flota de Distribuidora Cummins.

4.1 Revisión de la literatura

Para la solución que se pretende desarrollar, no existe una metodología concluyente para abarcar este tipo de problemas, por lo que esta revisión ha sido bastante guiada para entregar un entendimiento claro al lector.

La investigación estará enfocada en aquella literatura que busca predecir el comportamiento de un motor diésel, junto con literatura relacionada a algoritmos de clasificación binaria para cumplir con el objetivo de este proyecto y un *framework* que unifique todo el proceso de desarrollo.

4.1.1 Predicción en motores diésel

Existen muchos objetivos planteados, ya sea para entender el comportamiento de un motor en un ambiente específico, como también para pronosticar o clasificar ciertos comportamientos que cubren el objetivo de cada proyecto.

Son utilizadas técnicas tales como simulación del comportamiento del motor, predicción de ciertos valores de sensor mediante redes neuronales o algún algoritmo sofisticado que busca probar su efectividad, predominando el uso de redes neuronales.

Existen estudios de todo tipo relacionado a motores diésel, los cuales son utilizados en distintas aplicaciones (minería, marítima) donde, por ejemplo, una de las principales preocupaciones es la liberación de material particulado (PM).

En este proyecto, el foco es la detección temprana de fallas, donde se muestran varias técnicas que son utilizadas para este propósito (Sorbie, 2014):

- Detección de velocidad angular con un modelo de simulación, donde una de las limitaciones es que esta técnica no entrega información de la causa raíz, por lo mismo se puede volver muy costosa (Zhixiong, 2012).
- Modelo termodinámico, manteniendo el principio de simulación, con este enfoque es posible determinar causa raíz aplicando ingeniería inversa en el modelo al momento de la falla, siempre y cuando la falla se encuentre dentro del sistema simulado (Lamaris & Hountalas, 2010).
- Mantenimiento basada en condición es otra opción, monitoreando los sensores de los motores, aplicando un método de aprendizaje como red neuronal (Basurkoa & Uriondo, 2015).
- Kimmich, Schwarte & Isermann (2005) enfocan su trabajo en los síntomas del motor para definir comportamiento normal del “no normal”, indicando que existen 3 formas de detectar fallas:
 - Definición de tolerancias (rangos) para los parámetros del motor, estos deben ser monitoreados constantemente.

- El sistema detecta comportamiento no normal y propone alternativas de acción ante el comportamiento anormal.
- Vigilancia de los estados normales del motor permite definir comportamiento normal con el objetivo de encontrar fallas futuras.

Existen muchos trabajos donde lo anterior es ratificado, Hasan Aydogan (2015), Deng, Maass & Stobart (2011), Kokkulunk, Akdogan & Ayhan (2013) y Garg, Diwan & Saxena (2014) en sus trabajos concuerdan y evidencian que redes neuronales son una excelente alternativa para predecir valores que son temas de estudio en motores diésel.

Otros trabajos exponen el uso de *Support Vector Machine*, esto es trabajado por Srivastava & Bhambhu (2010), donde explica el uso de esta técnica para clasificación de información, dando un contexto de uso, ya sea en uso de reconocimiento de dígitos, tonos, clasificación de imágenes y detección de objetos, indicando que puede demostrar superioridad ante otras máquinas de aprendizaje, por otro lado, Domm (2003) menciona que *Support Vector Machine* es una buena alternativa para realizar clasificación binaria, donde el potencial de esta técnica es la capacidad de buscar un *kernel* capaz de ajustarse a la información que se tiene para el análisis del modelo, para esto, Maji, Berg & Malik (2008) en su trabajo realizan una comparación de los distintos *kernel* evaluando sus resultados.

El factor común relacionado a todos los trabajos revisados, es la predicción que está orientada a variables numéricas, por el contrario, lo que se busca en este proyecto es cubrir un problema de clasificación y de entrega de información para detectar la causa raíz de las fallas.

Una metodología que se acerca mucho a la detección de causa raíz es la simulación mediante el modelamiento de los distintos sistemas que existen en los motores.

Imamori, Hiraoka, Endo & Oda (2011) demostraron mejoras en el sistema de combustión en motores diésel en Industrias Mitsubishi Heavy Ltd., mediante el uso de sistemas de simulación, por otro lado, Tan, Deng & Lu (2004) logran generar un modelo de simulación que les permite predecir la emisión de material particulado, logrando extender este modelo en 3 tipos distintos de motores diésel.

Para lograr lo anterior es muy importante un conocimiento técnico del funcionamiento del motor diésel de modo tal que refleje la real operación del equipo, considerando supuestos realistas que eviten el sesgo que podría producirse en el modelo de simulación.

Cuando se habla de problemas de clasificación, y más aún, cuando la clasificación es binaria, la literatura coincide en muchas publicaciones que los algoritmos de Regresión Logística, *Support Vector Machine* y Árboles de decisión deben ser considerados. Otros estudios indican que redes neuronales es una buena opción de clasificación, sin embargo, de acuerdo a los trabajos anteriores, redes neuronales es utilizado fundamentalmente para pronosticar variables numéricas.

Corredor, Escobar, Portnoy & Vélez (2012) utilizan un modelo de regresión lineal para predecir el consumo específico de combustible en motores diésel de camiones, los cuales se encuentran con carga parcial.

Por otro lado, la literatura indica que los árboles de decisión son fáciles de aplicar, entregan información y ayuda a la definición de causa raíz de manera mucho más gráfica y sencilla para explicar a las personas del negocio que tiene estos propósitos, en especial a las áreas de interés.

Al buscar las mejores técnicas de clasificación binaria, las técnicas mayormente mencionadas son *Support Vector Machine* (donde el *kernel* debe ser determinado), regresión logística y árboles de decisión, lo cual es consecuente con lo mencionado anteriormente, salvo el uso de *Support Vector Machine* ya que si bien están

mencionados en predicciones mucho más complejas, también son utilizadas para el tipo de clasificación binaria que este trabajo está buscando.

4.1.2 Proceso KDD, *data mining* y sus implicaciones

Existe mucha información acerca de obtención de respuestas a problemas donde las características son similares:

- Problema existente, aunque no completamente definido
- Almacenamiento de *data* en fuentes no óptimas para su análisis.
- Gran cantidad información no relacionada o en sistemas *legacy*.
- Generación de información directamente desde el proceso productivo

En la mayor parte de la literatura se destaca el uso de los algoritmos matemáticos capaces de obtener información desde bases de datos, sin embargo, se debe entender que la selección del mejor algoritmo (ya sea el que toma menos tiempo de proceso, el que entrega la mejor predicción, etc.) es el paso final de un largo proceso de iteraciones y revisiones donde participan muchas personas con distintos niveles de conocimiento.

Según Fayyad, Piatetsky-Shapiro, & Smyth (1996), métodos tradicionales de análisis e interpretación eran realizados por personas (que actualmente son utilizados en muchas empresas), quienes deben invertir mucho tiempo en análisis, traducándose en un alto costo monetario para quién esté dispuesto a realizarlo, pero por sobre todo, la interpretación de resultados se vuelve subjetiva, dependiendo prácticamente de cada analista y su *know how*, considerando que este tipo de análisis es realizado utilizando tendencias en el tiempo como patrones de comportamiento que son ajustados a medida que es necesario. Sin embargo, esta forma de análisis no puede prosperar cuando las empresas y los sistemas generan tanta información por segundo que las personas no son capaces de analizar esta información (de forma tradicional),

cuando el problema actual de las distintas empresas es la forma correcta de almacenamiento para esta información.

En esta revisión, surgen de inmediato dos conceptos claves que han sido tema de varios *papers*, *Data Mining* y *Knowledge Discovery in Databases process*. Valcárcel (2004) define *Data Mining* como el proceso de extraer conocimiento desde bases de datos, permitiendo descubrir patrones, tendencias, etc. Este proceso intenta solucionar el problema con respecto a la gran cantidad de información existente y la incapacidad de las personas para realizar un análisis efectivo sobre la gran cantidad de información con la ayuda de la capacidad de procesamiento existente de los computadores.

Pese al gran potencial que existe en los procesos de *Data Mining*, los resultados pueden estar sesgados si no es incorporado dentro de un *framework* que considere aspectos relevantes como las fuentes de información, calidad de los datos, y una excelente definición del problema (tema fundamental para dar inicio a este proceso), junto con el conocimiento experto de las personas que participan en el proyecto.

Fayyad, Piatetski-Shapiro & Smyth (1996) presentan un trabajo donde el tema principal es la unificación de las técnicas de *Data Mining* con *Knowledge Discovery in Databases*, el cual fue presentado en un *workshop* en 1989.

Esta unificación es importante debido al concepto “*Garbage in, garbage out*”, donde Li, Wang & Lup (2000) en su trabajo “A Knowledge-Based Intelligent Data Cleaner” exponen de forma muy clara las implicaciones de ingresar información “sucia” o con “ruido”, donde se enfocan en la limpieza inicial de los datos.

Distribuidora Cummins Chile S.A., y en particular el área del *Mining Group* no es excepción y cumple con todas las características expuestas junto con sus problemas. Posee grandes cantidades de información en repositorios que no son los idóneos para realizar análisis necesarios, donde la metodología actual de análisis que se pretende utilizar corresponde sólo a análisis gráfico, univariable con criterios preestablecidos sólo por conocimiento experto que no consideran las variables críticas de los motores

como conjunto para análisis, ya que como es mencionado en la literatura, es imposible de realizar para las personas (Gibert & Cortés, 1998).

De acuerdo a la literatura con respecto a las necesidades de procesamiento, almacenamiento, calidad de información, y análisis de los resultados, es que el proceso KDD satisface todas las necesidades que la implementación de un proyecto como este requiere para que los resultados sean útiles y coherentes.

4.2 Metodología del proyecto

El desarrollo de este proyecto será abordado de acuerdo a lo impartido durante el *Master Business Engineering* desde el primer semestre del 2014, cuya metodología se encuentra detallada en el libro “Ingeniería de Negocios, diseño integrado de negocios, procesos y aplicaciones TI”, escrito por el profesor Oscar Barros utilizando su versión más reciente correspondiente al 2014.

Esta metodología está centrada en la estandarización de procesos de negocios que todas las empresas (no importando el rubro), y de acuerdo a la experiencia del autor y evidencia existente en la literatura respecto de la ingeniería de procesos, debe contar para lograr el grado de ordenamiento y mejoramiento en sus procesos internos y/o externos, evitando generar procesos ineficientes que puedan surgir por la urgencia del momento o inexperiencia en el análisis de los distintos procesos de la empresa.

Uno de los aspectos más importantes de la ingeniería de procesos en una organización es la existencia de varias capas de diseño, las que deben ser abarcadas de manera ordenada para lograr un entendimiento y desarrollo que se ajuste a las reales necesidades que se busca cubrir:

- **Arquitectura empresarial:** Se define como la relación entre la estrategia y estructura organizacional, con el modelo de negocios y arquitectura de procesos.

- Diseño de procesos: importante fase donde se puede crear, modificar y/o eliminar procesos para cumplir con los objetivos de la organización.
- Diseño y construcción de aplicaciones TI.

Para lograr un análisis acabado de lo ya mencionado, Barros en su metodología hace hincapié en la importancia de realizar los siguientes análisis de forma secuencial:

1. Planteamiento estratégico:

El análisis estratégico es el punto inevitable de partida para este tipo de trabajo, en ella se buscan los lineamientos que justificarán los procesos existentes, o en caso de ser necesario, generar rediseños. Se debe utilizar la mayor capacidad de análisis para obtener información referente a lo que la alta gerencia busca y pretende conseguir, cumpliendo con ello el objetivo de los distintos procesos existentes en la empresa. Para realizar lo anterior, importante son los trabajos desarrollados por Michael Porter (1996) y Arnoldo Hax (2010).

2. Definición modelo de negocios:

El negocio define sus pasos a seguir para cumplir con sus objetivos estratégicos, definiendo el cómo se entregará el servicio, cuáles serán los recursos a utilizar, como se obtiene el beneficio, y quienes son los clientes objetivo. Lo anterior es respaldado por el trabajo de Johnson, Christensen & Kagermann (2008).

3. Diseño arquitectura de negocios:

El desarrollo de la metodología propuesta por Oscar Barros, mediante el análisis de macroprocesos, los cuales son una especificación (abstracción) de todas (o la mayor parte) de los procesos que una organización debe contener para un funcionamiento óptimo de sus procesos y procedimientos, establece flujos de información, aprovechando la notación IDEF0. Barros establece la diferencia de 4 macroprocesos:

- Macroproceso 1: Considera la mayor cantidad de procesos que se encuentran dentro de lo que consideramos como la cadena de valor de una empresa.
- Macroproceso 2: Contiene los procesos que se relacionan a la creación de nuevas capacidades dentro de la organización de manera continua.
- Macroproceso 3: Agrupa los procesos que permiten dar pie a las actividades estratégicas de la organización, definiendo y controlando que la estrategia sea divulgada y ejecutada por toda la organización.
- Macroproceso 4: Agrupa los procesos que dan soporte operacional a los demás Macroprocesos, con procesos relacionados a Recursos Financieros, Recursos humanos, Infraestructura y Materiales.

4. Diseño detallado de los procesos:

Cuando se han establecido cuales son los macroprocesos que serán intervenidos, se debe comenzar con el diseño de los procesos en un nivel más detallado para que los cambios tengan efecto de forma coordinada. Para esto, se utiliza el diseño y/o rediseño de procesos utilizando notación BPMN 2.0 con

IDEF0, permitiendo indicar las tareas que van asociadas a cada proceso. La principal característica de este punto es que en los diagramas BPMN se colocan de manera explícita los distintos responsables junto con sus tareas.

5. Diseño de la aplicación de apoyo:

De acuerdo a las tareas que son expuestas en los diagramas BPMN, y en consecuencia con la orientación del MBE, es que se debe establecer la documentación necesaria para el diseño del apoyo TI. Esto se realiza identificando en los diagramas BPMN la interacción entre el sistema y el o los usuarios. Lo anterior implica documentar de forma explícita como se realiza el proceso o la interacción entre lo que hace el sistema y el usuario, para lo cual se utiliza la notación UML (*Unified Model Language*), junto con la metodología de programación orientada a objetos.

6. Construcción e implementación de la solución:

Con toda la información recogida en los puntos anteriores, es posible desarrollar un prototipo de lo planteado, y en el mejor de los casos se puede implementar en la empresa donde se desarrolla el proyecto para apoyar al negocio en la toma de decisiones.

A modo de resumen se muestra la siguiente figura:

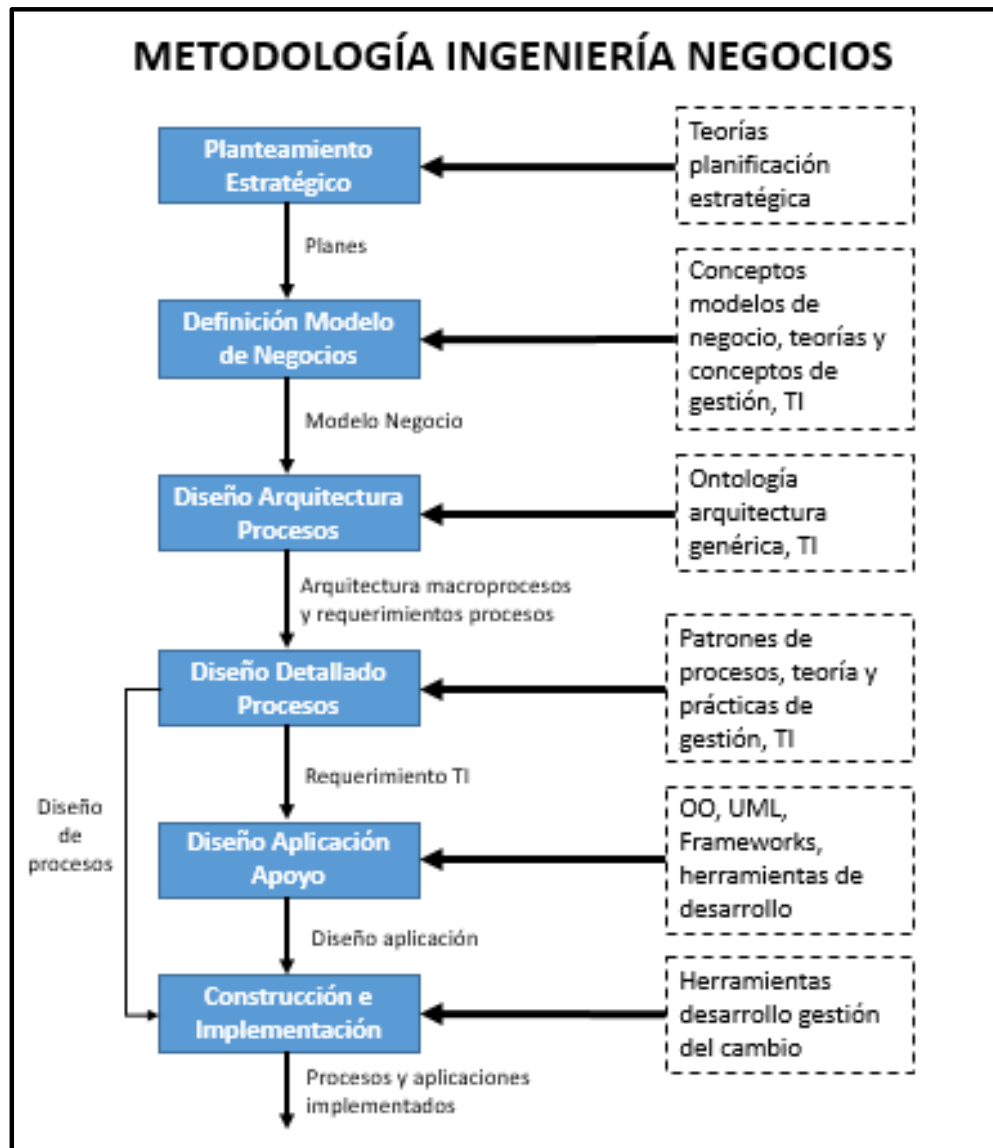


Figura 4.1: Metodología Ingeniería Negocios. Fuente: Ingeniería de Negocios, diseño integrado de servicios, sus procesos y apoyo TI. Oscar Barros 2014.

4.3 Mantenimiento

4.3.1 Definición de mantenimiento

Mantenimiento se define como:

- “Actividades necesarias para mantener los equipos e instalaciones en una condición particular o volverlos a dicha condición” (Prando, 1996).
- “El conjunto de técnicas destinadas a conservar equipos e instalaciones en servicio durante el mayor tiempo posible (buscando la más alta disponibilidad) y con el máximo rendimiento” (García Garrido, 2003).

Las definiciones son similares, cambian sus enfoques. Prando en el año 1996 define mantenimiento como la actividad sólo de mantener en funcionamiento la maquinaria, en cambio García Garrido en 2003 indica que esta actividad debe ser capaz de obtener el máximo rendimiento de una máquina o equipo.

4.3.2 Propósito del mantenimiento

El mantenimiento es el medio que tiene toda empresa para conservar operable con el debido grado de eficiencia y eficacia su activo fijo. Engloba el conjunto de actividades necesarias para:

- Mantener una instalación o equipo en funcionamiento.
- Restablecer el funcionamiento del equipo en condiciones predeterminadas.

Estas actividades han evolucionado a través de los años, agrupándose en técnicas como se muestra en la siguiente figura:

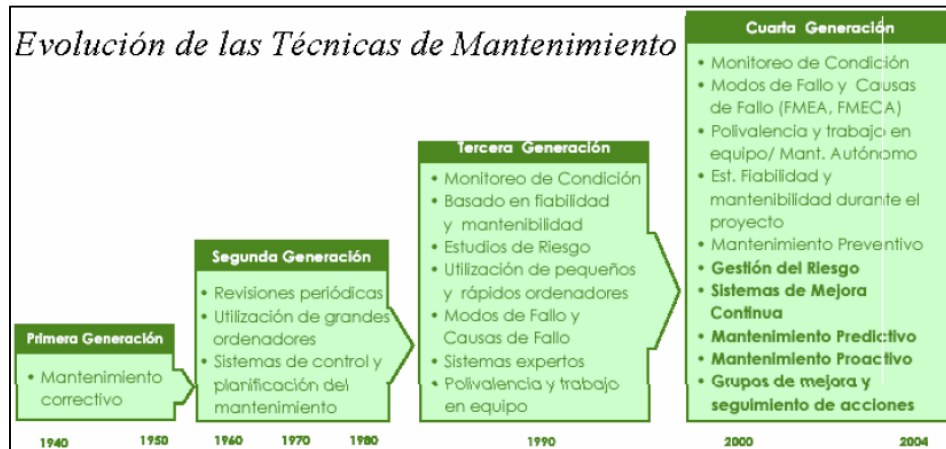


Figura 4.2: Evolución técnicas de mantenimiento. Fuente: Lincovil y Gutierrez, 2006

Se puede ver cómo han evolucionado estas técnicas en comparación a las distintas definiciones existentes, desde el mantenimiento correctivo a actividades mucho más sofisticadas como mantenimiento predictivo o proactivo, cuya selección dependerá de las necesidades que la organización en un momento dado.

El mantenimiento incide, por lo tanto, en la cantidad y calidad de la producción. En efecto, la cantidad de producción a un nivel de calidad dado está determinada por la capacidad instalada de producción y por su disponibilidad, entendiéndose por tal al cociente del tiempo efectivo de producción entre la suma de este y el tiempo de parada por mantenimiento.

4.3.3 Objetivo del mantenimiento

Asegurar la disponibilidad planeada al menor costo posible, dentro de las recomendaciones de garantía y uso de los fabricantes de los equipos e instalaciones y las normas de seguridad. Para ello actúa sobre:

- La continuidad de la operación de producción, es decir, la confiabilidad que se mide por el tiempo medio entre fallas consecutivas.
- El tiempo de paradas cuando estas se producen. El tiempo de paradas incluye el tiempo efectivo de la reparación (Mantenibilidad) que es función del diseño,

herramientas disponibles, destreza y capacitación del personal, y del tiempo de espera que es función de la organización (sistemas, rutinas, herramientas y talleres disponibles, documentación técnica, capacitación, entrenamiento y suministro de piezas y/o repuestos).

4.3.4 ¿Qué ocurre cuando se interrumpe un proceso?

El encargado de mantenimiento debe tener en cuenta las causas de interrupción, el tiempo de interrupción y las posibles consecuencias para tomar medidas eficientes y eficaces. En siguiente la figura se muestra un cuadro con respecto a este tema:

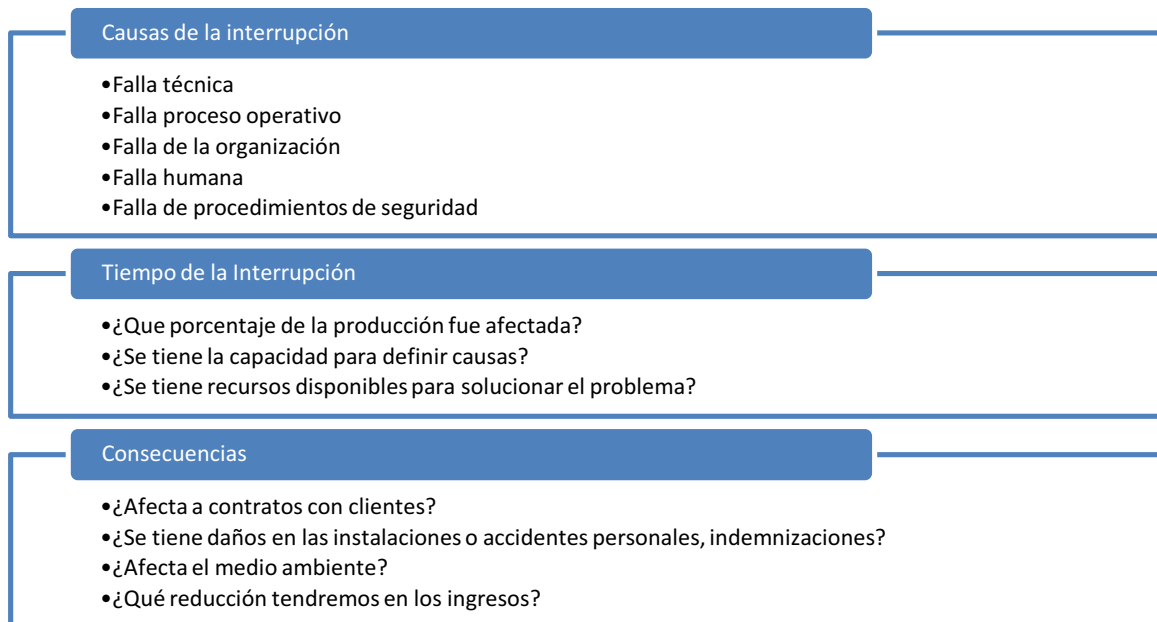


Figura 4.3: Aspectos a considerar en interrupción de un proceso. Fuente: Gestión de mantenimiento y su impacto en seguridad operativa y medio ambiente. Juan Tavorara. Bolivia. 2002

En el caso de que ocurra un suceso que interrumpe el proceso, se debe a que algún equipo presentó una falla. Se dice que algo falla cuando deja de brindar el servicio que debía entregar o cuando aparecen efectos indeseables, según las especificaciones de diseño con las que fue construido o instalado el bien en cuestión (Molina, 2001).

4.3.5 Gestión del mantenimiento

a. Producción v/s Mantenimiento

Desde el momento que las empresas se dieron cuenta que era necesario separar el departamento de mantenimiento con el de producción, el departamento de mantenimiento ha sido un subordinado de este último.

Esta forma de establecer la relación entre mantenimiento y producción tal vez sea válida en entornos donde no existe Gestión de Mantenimiento, donde Mantenimiento sólo se ocupa de las fallas que comunica producción. Esta situación cambia cuando el mantenimiento es gestionado (optimización de los recursos que se emplean). En este escenario, mantenimiento y producción son dos entes igualmente importantes para la producción. Para que la organización funcione, es necesario que todos los departamentos y áreas trabajen de forma coordinada, y la eficiencia de la organización estará dada por el departamento que peor funcione, de esta forma se puede decir que en las organizaciones donde el mantenimiento se gestiona, este departamento no es subordinado de producción, buscando la optimización en ambos departamentos de forma paralela (García Garrido, 2003).

b. ¿Por qué se debe gestionar el mantenimiento?

¿No es más fácil y barato ir a reparar los equipos cuando se averíen y evitar planes de mantenimiento que incrementan notablemente los costos de mano de obra indirecta? Algunos de los por qué son presentados a continuación:

1. Presión ejercida por la competencia a la disminución de los costos.

Necesidad de optimizar el uso de materiales y mano de obra. Es necesario identificar el modelo de organización que mejor se adapta a la planta, identificar la influencia de los equipos en los resultados de la empresa, consumo y stock de materiales empleados en mantenimiento, aumento de la disponibilidad de

los equipos de tal forma que los periodos de indisponibilidad de estos no interfieran con el plan de producción.

2. *Aparición de técnicas que ayudan a mejorar los resultados de la empresa.*

La implementación de este tipo de técnicas, que están dirigidas al mantenimiento, es necesario analizarlas de tal forma que sea posible desarrollarlas en la empresa (dependiendo de su situación actual) y la forma de implementar en el caso de ser factible. Algunas técnicas son TPM (*Total Productive Maintenance*, Mantenimiento Productivo Total), RCM (*Reliability Centered Maintenance*, Mantenimiento Centrado en Confiabilidad), sistemas GMAO (Gestión de Mantenimiento Asistido por Ordenador), diversas técnicas de Mantenimiento Predictivo (análisis vibracional, termografías, detección de fugas por ultrasonidos, análisis amperimétricos, etc.).

3. *Los distintos departamentos de una empresa necesitan estrategias.*

Es necesaria la implantación de directrices que les sirvan a los encargados de los departamentos para que sus metas sean guiadas a los objetivos de la organización.

4. *Aspectos como calidad, seguridad e interrelaciones, han tomado importancia en la gestión industrial.*

Es necesario gestionar estos aspectos para incluirlos en las formas de trabajo de los departamentos de mantenimiento.

Para efectos del proyecto, se describirá la filosofía que existe detrás del mantenimiento centrado en confiabilidad (RCM).

4.3.5.1 Mantenimiento Centrado en Confiabilidad (RCM, Reliability Centered Maintenance)

Dentro de las definiciones de mantenimiento, todos los autores coinciden como mínimo que la base del mantenimiento es “Asegurar que todo elemento físico continúe desempeñando las funciones deseadas” (Pérez & Moubray, 2004), y es en este concepto que RCM se basa para definir sus pasos a seguir, ya que un elemento físico no puede entregar mayor confiabilidad que la indicada por sus diseñadores, cada sistema opera en un ambiente único que puede ser caracterizado por distintas variables como ubicación, altitud, profundidad, atmósfera, presión, temperatura, humedad, salinidad, exposición a procesamiento de fluidos, aceleración, entre muchos otros.

Según Pérez & Moubray (2004), una definición amplia de RCM es “un proceso que se usa para determinar lo que se debe hacer para asegurar que un elemento físico continúe desempeñando las funciones deseadas en su contexto presente”.

Otra definición con mayor nivel de profundidad es la que entrega Mascarell (2007), el cual indica que RCM es “el proceso utilizado para determinar el mantenimiento más efectivo posible, el cual está basado en una serie de actividades sistemáticas generadas para evaluar u optimizar un programa de mantenimiento”.

Según J. Hung (2008), RCM se caracteriza por:

- Considerar la confiabilidad inherente o propia del equipo / instalación.
- Asegurar la continuidad del desempeño de su función.
- Mantener la calidad y capacidad productiva.
- Tiene en cuenta la condición operacional de donde y como se está usando

A modo de visualización se muestra donde se encuentra situado RCM en relación a los distintos tipos de mantenimiento existentes, los cuales serán desarrollados más adelante:

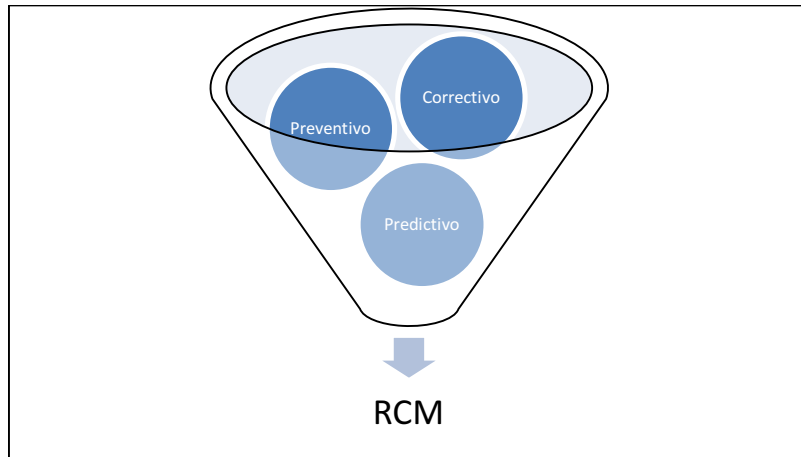


Figura 4.4: Relación Tipos de mantenimiento con RCM. Fuente: Mascarell, 2007.

El Mantenimiento Centrado en Confiabilidad (RCM) si bien es una técnica integral, ya que relaciona a los distintos tipos de mantenimientos existentes, requiere de mucha información previa para comenzar a realizar su función principal, que es identificar factores críticos, ojalá en flotas homogéneas, de tal forma de encontrar la mejor alternativa de mantenimiento al menor costo.

Mascarell (2007), Hung (2008) y Espinosa (2010), entre otros autores, coinciden en que RCM requiere de la información previa de cada elemento existente en la organización para poder implementar esta metodología de forma óptima, la información requerida debe contestar las siguientes preguntas (Espinosa, 2007):

- ¿Cuáles son las funciones y los parámetros de funcionamiento asociados al activo en su actual contexto operacional?
- ¿De qué manera falla en cumplir con sus funciones?
- ¿Cuál es la causa de cada falla funcional?
- ¿Qué sucede cuando ocurre la falla?
- ¿Qué puede hacerse para predecir/prevenir cada falla?
- ¿Qué debe hacerse si no se encuentra una tarea proactiva adecuada?

De este modo, la idea central de RCM (Hung, 2008) es que los esfuerzos de mantenimiento deben ser dirigidos a mantener la función que realiza un equipo más que enfocar el mantenimiento en el equipo mismo, y esto se traduce a la función desempeñada desde el punto de vista productivo.

De acuerdo con Mascarell (2007), la estrategia que ejecuta RCM es establecer la mejor estrategia de mantenimiento basado en la mejor decisión a tomar, y desde el punto de vista del negocio, la mejor decisión a tomar es aquella que reduce la probabilidad de falla al máximo posible y que además permite aumentar la rentabilidad.

El análisis basado en RCM se realiza mediante el detalle del Análisis detallado de fallas y sus consecuencias (*Failure Mode and Effect Analysis*, FMEA), el cual debe considerar como mínimo la probabilidad de fallo y los cálculos de confiabilidad del sistema como parámetros de entrada para discriminar, con lo anterior, el análisis permitirá determinar las tareas de mantenimiento que identifican las formas de fracaso y sus consecuencias.

- Beneficios de implementación de RCM (Pérez & Moubray, 2004):
 - Mayor seguridad y protección del entorno debido a la búsqueda de las consecuencias de las fallas y por ende su identificación y mitigación.
 - Mejoramiento de rendimientos operativos de los equipos, debido a la optimización de recursos (gestión de personal, insumos y repuestos).
 - Control de los costos de mantenimiento, ya que se encontrarán estandarizados desde el momento que se clasifica toda la información requerida.
 - Aumento de la vida útil de los equipos de la empresa.
 - Generación de *data* histórica de los activos en la empresa, lo que permite mayor nivel de análisis.
 - Mejora el trabajo en equipo (producción y mantenimiento).

RCM se presenta como una de las metodologías que mejor se ajusta para las necesidades de este proyecto, ya que permite balancear al menos 2 tipos de mantenimiento que actualmente existen y que se describen a continuación, que son mantenimiento correctivo y mantenimiento preventivo, y como se desarrollará más adelante dentro de esta tesis, mantenimiento predictivo.

4.3.6 Tipos de mantenimiento

Según García, existen 2 grandes tipos de mantenimiento, mantenimiento correctivo y mantenimiento preventivo (Olarte, Botero & Cañón en su trabajo desarrollado en el 2010 coinciden con García), donde este último se subdivide en predictivo, *hard time* o cero horas y mantenimiento en uso. A continuación se define cada uno.

a. Mantenimiento correctivo (después de la falla):

Es el conjunto de tareas destinadas a corregir los defectos que se van presentando en los distintos equipos y que son comunicados al departamento de mantenimiento por los usuarios de los mismos.

b. Mantenimiento preventivo (antes de la falla):

Tiene por misión mantener un nivel de servicio determinado en los equipos, programando las correcciones de sus puntos vulnerables en el momento más oportuno. Este tipo de mantenimiento recibe distintos nombres dependiendo de las actividades que sean realizadas. A continuación se describen algunos tipos.

- **Mantenimiento predictivo:**

Es el que persigue conocer e informar permanentemente del estado y operatividad de las instalaciones mediante el conocimiento de determinadas variables, representativas de tal estado y operatividad. Para aplicar este mantenimiento es necesario identificar variables físicas

(temperatura, vibración, consumo de energía, etc.) cuya variación sea indicativa de problemas que puedan estar apareciendo en el equipo. Este tipo de mantenimiento es el más tecnológico debido a que requiere medios técnicos avanzados y fuertes conocimientos matemáticos, físicos y técnicos.

- Mantenimiento cero horas (*overhaul*):

Es el conjunto de tareas cuyo objetivo es revisar los equipos a intervalos programados, antes de que aparezca ningún fallo, cuando la confiabilidad del equipo ha disminuido apreciablemente, de manera que resulta arriesgado hacer previsiones sobre su capacidad productiva. Dicha revisión consiste en dejar el equipo a cero horas de funcionamiento, es decir, como si el equipo estuviese nuevo. En estas revisiones se sustituyen o se reparan todos los elementos de desgaste. Se pretende asegurar, con gran probabilidad, un tiempo de buen funcionamiento, fijado de antemano.

- Mantenimiento en uso:

Es el mantenimiento básico de un equipo realizado por los usuarios del mismo. Consiste en una serie de tareas elementales (toma de datos, inspecciones visuales, limpieza, lubricación, reapriete de tornillos) para las que no es necesario una gran formación, sino tan solo un entrenamiento breve. Este tipo de mantenimiento es la base de TPM (*Total Productive Maintenance*).

Es necesario señalar que estos tipos de mantenimiento son aplicables pero no de forma absoluta uno u otro, y esto se debe a que los equipos, dependiendo de sus características, necesitarán la aplicación de más de un tipo de mantenimiento. La selección de la mezcla idónea de estos tipos de mantenimiento estará dada por razones que se encuentran estrictamente ligadas al costo de las pérdidas de

producción en el equipo en estudio, el costo de la reparación, el impacto ambiental, la seguridad y la calidad del producto o servicio, entre muchas otras razones estratégicas.

4.3.6.1 Mantenimiento Predictivo

Cuando se habla de mantenimiento predictivo, se refiere al seguimiento continuo de variables de un equipo, los cuales son considerados críticos dentro de su sistema de funcionamiento y por lo tanto entregan información valiosa para la operación del mismo, son considerados sus signos vitales y por lo tanto entregan información de estado actual (Piñol, Ortega & Romeral, 2004). Según Piñol, esta estrategia es la más complicada de implementar debido a la fuerte inversión inicial en equipos de medida y formación de técnicos que puedan interpretar y convivir con este tipo de tecnología, sin embargo, es el mantenimiento que entrega mejores resultados, aumentando la vida útil del equipo garantizando máxima calidad y seguridad del sistema productivo, además, permite inferir las causas de la falla, generando medidas de corrección para el entorno donde se encuentra inmerso el equipo (exterior) y corregir elementos alrededor de la falla (en caso de que sea necesario).

Uno de los problemas (o restricción) que tiene esta metodología, es que se requiere la medición y monitorización de la máquina en funcionamiento con técnicas no invasivas para que los resultados sean válidos, ante esto, Piñol, Ortega & Romeral (2004), basado en su trabajo para motores de inducción sugiere 2 métodos de diagnóstico:

- a. *Diagnóstico por análisis de vibraciones*, donde la teoría indica que un motor rotativo no debiese tener vibraciones en condiciones ideales, y en caso de tenerlas se debe a imperfecciones de construcción de máquinas, de los procesos donde se encuentran insertos o por averías, el inconveniente de este diagnóstico es que implica incorporar acelerómetros en el equipo que se desee monitorear.

- b. *Motor Current Signature Analysis* (MCSA), el cual se centra en el análisis espectral de las corrientes del estator. Con esta técnica se pueden diagnosticar fallas mecánicas y eléctricas, aunque se utiliza principalmente para detectar fallas en las barras del rotor.

Por otro lado, Hidalgo (2013) en su documento relacionado a los impactos de la tecnología el mantenimiento, indica otro nombre al mantenimiento predictivo, al cual llama mantenimiento basado en condición (MBC), cuyas ventajas están relacionadas a la disminución de la tasa de salida forzada de equipos, aumento de confiabilidad y disponibilidad, aumento de la vida útil de los equipos y disminución de inventario de repuestos, por otro lado las desventajas se traducen al alto costo por el equipamiento de herramientas de diagnóstico, el costo del servicio contratado, y el personal especializado.

Es importante recalcar que este tipo de mantenimiento permite detectar la degradación de componentes críticos en tiempo real logrando que los encargados puedan tomar decisiones de manera oportuna.

Hidalgo indica que los beneficios asociados al mantenimiento predictivo son:

- Ahorro por uso óptimo de recursos.
- Aumento de la disponibilidad del equipo.
- Disminución en los costos por inventario.

El autor menciona que las técnicas más modernas (al 2013) en MBC o Mantenimiento predictivo son la termografía infrarroja, análisis de vibraciones mecánicas (que coincide con Piñol), análisis de aceite, análisis de características eléctricas de motores y alineación de ejes.

La mayor parte de la literatura relacionada a mantenimiento predictivo o *condition-based maintenance* (CBM) evidencia la necesidad de medición de un elemento no invasivo, en tiempo real con el equipo funcionando, donde la gran coincidencia entre

los autores mencionados en este documento y la mayoría de los trabajos realizados es el uso del análisis de aceite y vibraciones. Sin embargo, existe a lo menos un *paper* (Parlak, Islamoglu, Yasar, & Egrisogut, 2005) quienes demuestran con un error menor al 2% que se puede monitorear y predecir cierta variable de un motor *diesel* sin la necesidad de monitoreo externo, considerando sólo las variables monitoreadas por el motor de forma interna (basado en sensores), indicando que es una buena alternativa el uso de estas variables junto con el uso de redes neuronales.

4.3.7 Ventajas y desventajas

Muchas son las ventajas al aplicar estrategias de mantenimiento de forma eficiente y correcta, en forma general es garantizar la producción y mantener los equipos operables aumentando su vida útil.

Existen modelos de mantenimiento que ayudan a una inspección constante para tomar decisiones basadas en criterios de ingeniería y desempeño de los elementos que conforman la producción.

Se puede decir (por decir algo) que una de las desventajas que implica una Gestión del Mantenimiento (o cualquier tipo de gestión), es la necesidad de recolección de información para realizar los análisis necesarios. Sin embargo, en los escenarios actuales que varían de forma muy constante, es importante contar con información necesaria para identificar debilidades y mejorarlas, o identificar oportunidades y explotárlas, lo anterior, generará ventajas competitivas en la organización que en el largo plazo se deberán convertir en claves de éxito para la empresa.

4.4 Knowledge Discovery in Databases (KDD)

Para el desarrollo de la solución propuesta ante la problemática de este proyecto, es necesario contar con una metodología válida y estructurada para el análisis de la información, esta metodología es la propuesta como *framework* por Fayyad en 1996, llamado *Knowledge Discovery in Databases* (KDD).

Fayyad, dentro del desarrollo de su trabajo, indica que este proceso no debe ser confundido con lo que el resto de la gente llama *Data Mining*. Fayyad enfatiza la diferencia entre el proceso KDD y *Data Mining* indicando que el proceso KDD es un *framework* donde *Data Mining* es un subproceso dentro del mismo, por lo cual, ambos procesos no pueden ser homologados.

La definición de Fayyad para el proceso KDD es “proceso no trivial de identificación de información válida, novedosa con uso potencial que permitan identificar patrones entendibles y/o interpretables”, lo anterior quiere decir que este proceso permite detectar información, generando relaciones válidas entre las distintas variables que no están presentes de forma explícita en la base de datos, pero que si tienen relación de forma natural debido al contexto con el cual nació esa *data*, es muy importante el contexto dentro de este proceso ya que será parte del establecimiento de los supuestos para el desarrollo de este proceso.

Por otro lado, *Data Mining* se refiere al proceso donde se aplican algoritmos de análisis de datos y descubrimiento de información, generando patrones sobre la *data* analizada, patrones que bajo los supuestos iniciales genera patrones finitos.

El proceso KDD y el proceso de *Data Mining* se unifican al momento de interpretar estos patrones, ya que es en ese momento cuando KDD entrega su valor agregado.

El proceso KDD es interactivo e iterativo, lo que quiere decir que el proceso inicia, y de acuerdo a los resultados que se obtienen en cada paso, puede ser que se tenga que continuar iterando en un mismo paso, como adelantar o retroceder en las distintas fases.

Los pasos definidos en el proceso KDD (Fayyad, 1996) son:

1. Definición e identificación del objetivo buscado.
2. Creación de la base de datos objetivo.
3. Limpieza de la *data* y reproceso.

4. Reducción de la *data* dependiendo del objetivo.
5. Determinar el método de minería de datos de acuerdo al objetivo inicial.
6. Selección de métodos y parámetros de algoritmos de minería de datos.
7. Buscar patrones que sean de interés y que tengan sentido basado en la información histórica.
8. Interpretar la información explotada (minada).
9. Consolidar y documentar el nuevo conocimiento para traspasarlo a otra aplicación.

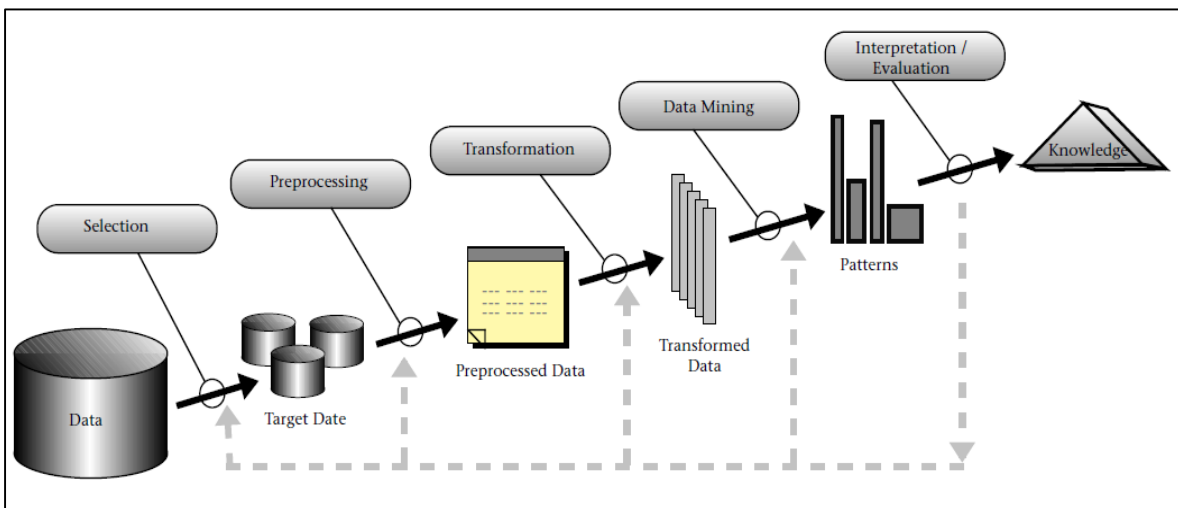


Figura 4.5: Proceso KDD. Fuente: Fayyad, 1996.

El desafío de este proceso es identificar la iteración que conducirá a los resultados esperados de tal forma de optimizar los recursos utilizados, para esto, se debe generar un equipo multidisciplinario capaz de interpretar los resultados obtenidos, indicar y entregar la mayor cantidad de información disponible para definir de mejor forma los objetivos dentro de las bases de datos, y por último, es necesario el conocimiento experto de la información analizada.

A continuación serán presentados los algoritmos que serán utilizados para el análisis de la información.

4.5 Regresión Logística

La regresión logística es una técnica estadística multivariante, que permite estimar la relación entre una variable dependiente (dicotómica) y un conjunto de variables métricas o no métricas (Salcedo, 2004).

La gran diferencia con una regresión lineal es que la variable dependiente no es continua, y que la ecuación del modelo no es lineal, sino que es exponencial, que por simple transformación logarítmica puede ser presentada como exponencial.

4.5.1 Objetivos Regresión logística

Según (Salcedo, 2004), los objetivos de la regresión logística son:

- Predicción de probabilidad de ocurrencia de un evento de interés.
- Definir variables predictores (independientes) que sean relevantes para el caso en estudio.
- Eliminar supuestos no aplicables en modelos de regresión lineal al momento de predecir variables dependientes dicotómicas.

Para efectos de este proyecto de título, el tipo de regresión a detallar se encuentra enfocada en variables dependientes binarias.

4.5.2 Modelo Logit

De acuerdo a Pérez, Kisys & Manzanedo (2002), sea la siguiente expresión algebraica para regresión múltiple:

$$Y = f(X_1, X_2, \dots, X_k, u) \quad (1)$$

Donde:

Y : Variable dependiente

X_k : Variable independiente

u : Constante de perturbación

Para el caso de regresión lineal simple, la expresión queda de la siguiente forma:

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u \quad (2)$$

Para el caso de la regresión lineal, el objetivo será calcular los valores $\beta_1, \beta_2, \dots, \beta_k$ que permitan a (2) el mejor ajuste posible de los datos, que en el caso de regresión lineal se logra mediante el método de Mínimos Cuadrados (MCO).

Para el caso del modelo Logit, la función logística (f) es:

$$f(z) = \frac{\exp(z)}{1 + \exp(z)}$$

Reemplazando con el mismo razonamiento que en regresión lineal tenemos que:

$$E[Y] = P(Y = 1) = \frac{\exp(\beta_1 + \beta_2 X_2 + \dots + \beta_k X_k)}{1 + \exp(\beta_1 + \beta_2 X_2 + \dots + \beta_k X_k)}$$

Para el cálculo de los estimadores se utiliza el método de Máxima Verosimilitud (MV).

Se debe tener en cuenta que por la transformación realizada de f , no se pueden interpretar directamente los estimadores β , esto, ya que son modelos no lineales, ante esto, para la interpretación de los resultados se debe enfocar en los signos de los estimadores, cuando el signo es positivo de β_2 indica que la variable X_2 aumenta el valor de $P(Y=1)$, cuando el signo es negativo, indica lo contrario.

Logit además utiliza 2 conceptos para profundizar en los estimadores:

- Odds: cociente de probabilidades:

$$Odds = \frac{P(Y = 1)}{1 - P(Y = 1)} = \exp(\beta_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

Tomando logaritmo natural para la expresión anterior se obtiene una expresión lineal:

$$Logit[P(Y = 1)] = \ln\left(\frac{P(Y = 1)}{1 - P(Y = 1)}\right) = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

- Odds-ratio: incremento unitario variables explicativas modelo
Aparece el concepto de odds-ratio como el cociente entre los dos odds asociados (antes del incremento y el del incremento correspondiente), suponiendo que existe un incremento unitario en la variable X_i tenemos:

$$Odds - ratio = \frac{Odds2}{Odds1} = \exp(\beta_i)$$

Se deduce que un β_i cercano a cero (o un Odds-ratio cercano a 1) significará que cambios en la variable explicativa X_i no tendrá efecto alguno sobre la variable independiente Y .

4.6 Árbol de decisión

Los árboles de decisión son algoritmos que permiten analizar información, tomar decisiones secuenciales basada en el uso de *data* histórica y el uso de probabilidades asociadas a las ocurrencias de interés (Tang, Zheng, Huang, Shu & Wang, 2005).

Las ventajas de un árbol de decisión son:

- Permite clasificar nuevos casos siempre y cuando no existan cambios sustanciales de las condiciones donde estos casos fueron generados en comparación a la base de entrenamiento.
- Permite segmentar en grupos que sean relevantes entre sí.

- Facilita la interpretación de los resultados.
- Permite disminuir la cantidad de variables independientes de la muestra.

El desarrollo de los árboles de decisión nace principalmente con fines exploratorios de la *data* (Bouza & Santiago, 2012), lo que termina convirtiéndose en una de las metodologías más populares gracias a su gran facilidad de interpretación de resultado para distintos niveles de usuario/cliente.

Existen muchos árboles de decisión, de los cuales se mencionan los siguientes:

- Software AID (*Automatic Interaction Detection*) (Sonquist-Morgan, 1964):
Cuyo objetivo inicial estaba basado en encontrar relaciones e ilustrar los resultados obtenidos de acuerdo al problema planteado.
- CART (*Classification and Regression Trees*) (Breiman-Friedman-Olshen-Stone, 1984):
Este método permite la creación recursiva del árbol de decisión.
- ID3 (*Iterative Dichotomiser 3*) (Quinlan, 1986):
Se utiliza la medida de entropía entre los campos para formar el árbol de decisión.
- C4.5 (Quinlan, 1993):
Algoritmo mejorado de ID3 por su autor.

El factor común de los métodos antes mencionados es que su clasificación es binaria (2 sub-nodos por nodo principal).

- CHAID (*Chi-squared Automatic Interaction Detection*) (Kass, 1980):
Introdujo un algoritmo recursivo de clasificación no binario, algoritmo en el cual nos centraremos en este proyecto.

4.6.1 Prueba de Chi-cuadrado (χ^2) (Monge & Pérez, 2002)

Dentro de los algoritmos de árbol de decisión CHAID, importante es esta distribución, utilizada como principal componente en la designación de los grupos que serán relevantes o no para la formación del árbol de decisión.

Esta prueba tiene 3 pasos fundamentales:

- Prueba de bondad de ajuste
- Prueba de homogeneidad
- Prueba de independencia

4.6.1.1 Prueba de bondad de ajuste

Es necesario determinar que los datos de una muestra aleatoria simple de tamaño n corresponden a cierta distribución teórica. Primer paso es descomponer el recorrido de la distribución teórica en un número finito de subconjuntos: A_1, A_2, \dots, A_K , luego clasificar las observaciones muestrales según el subconjunto al que pertenezcan y luego comparar las frecuencias observadas de cada A_i con las distribuciones de la frecuencia muestral teórica.

Suponiendo que existe un número k de clases donde se han registrado n observaciones ($n =$ tamaño muestral), las frecuencias observadas en cada clase serán denominadas como O_1, O_2, \dots, O_K (O_i es el número de valores de la clase A_i), donde se cumplirá que:

$$O_1 + O_2 + \dots + O_k = n$$

Las frecuencias esperadas (teóricas) las denotaremos por E_1, E_2, \dots, E_K , y se cumplirá que:

$$E_1 + E_2 + \dots + E_k = n$$

Ahora se tratará de decir si las frecuencias observadas están en concordancia con las frecuencias teóricas, donde el estadístico de contraste será:

$$\chi^2 = \sum_1^k \frac{(O_i - E_i)^2}{E_i}$$

Se puede deducir que mientras más pequeño sea este índice, mayor será la coherencia con los valores esperados. La hipótesis nula indica que “los valores observados son coherentes con los esperados”, por lo que para valores altos, esta hipótesis es rechazada, esto, cuando el estadístico es mayor que un determinado valor crítico.

El uso de este estadístico está sujeto a 2 restricciones:

- Todas las frecuencias serán igual o mayores a 5 y la muestra como mínimo debe tener 30 registros.
- Las observaciones son obtenidas de un muestreo aleatorio simple a partir de una población particionada en categorías.

4.6.1.2 Prueba de homogeneidad

Se requiere determinar si los datos correspondientes a 2 o más muestras aleatorias provienen de la misma población. El conjunto de posibles valores de las observaciones se divide en k conjuntos disjuntos: A_1, A_2, \dots, A_k , clasificando en ellos las observaciones de cada muestra. Si n_{ij} representa el número de observaciones de la muestra i , que pertenecen a A_j .

El estadístico de contraste es:

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^k \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Donde e_{ij} es la frecuencia esperada bajo el supuesto de homogeneidad, que puede representarse como $e_{ij} = n_i p_j$, el número de individuos en la muestra i por la probabilidad de que ocurra la característica j en la población.

Para la probabilidad tenemos que: $p_j = n_j/n$, por lo tanto $e_{ij} = n_i n_j/n$.

La hipótesis nula es: “los valores observados son coherentes con los esperados”, para valores pequeños de χ^2 más coherentes serán los datos, por el contrario, para valores grandes de χ^2 , y dependiendo de un valor crítico, los valores no serán coherentes.

El uso de este estadístico está sujeto a las mismas restricciones de la prueba de bondad de ajuste.

4.6.1.3 Prueba de independencia

Se busca determinar si dos variables de una población están relacionadas. Sea una población de n elementos, en la cual se han observado características X e Y , obteniéndose una muestra aleatoria simple bidimensional $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, el objetivo es conocer si estas características son independientes una de la otra. Para lo anterior, se dividirá el conjunto de posibles valores de X en k conjuntos disjuntos A_1, A_2, \dots, A_k , mientras que el conjunto de posibles valores de Y será descompuesto en r conjuntos disjuntos: B_1, B_2, \dots, B_r .

Al clasificar los elementos de la muestra aparecerán n_{ij} correspondiente a $k \times r$ clases, que, al igual que en el *Test* de homogeneidad, el estadístico de contraste será:

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^k \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Donde:

$$e_{ij} = n_i n_j / n$$

4.6.2 CHAID (Chi-squared Automatic Interaction Detection)

Esta técnica tiene como principal característica con respecto a los demás algoritmos de árbol de decisión que la muestra no se segmentará de modo binario, lo que permitirá generar más de dos segmentos por nodo.

A continuación se definen los pasos a seguir de este algoritmo:

4.6.2.1 Reducción categorías por variable independiente

Suponiendo que una variable independiente contiene c valores, se busca reducir a un número $k \leq c$ de forma tal que reduzca la complejidad de segmentación sin pérdida sustancial de valor. Para lo anterior, se debe analizar la naturaleza de las variables independientes:

- a. Variables nominales: cualquier valor de esta categoría puede ser agrupado con otro valor de sí mismo sin problemas.
- b. Variables ordinales: los valores pueden agruparse siempre y cuando estos mantengan su orden jerárquico, establecido previamente.
- c. Variables cuantitativas: para ser utilizadas en el procedimiento CHAID, deben ser recodificadas en valores discretos y ser tratadas como ordinales.

La formación de categorías homogéneas se basa en el estadístico chi-cuadrado, esto con los pasos siguientes:

- i. Se forman todos los pares posibles de categorías.

- ii. Para cada posible par se calcula chi-cuadrado, correspondiente al cruce con la variable dependiente, mezclándose aquellas categorías que posean el valor más bajo de chi-cuadrado.
- iii. Si se han fusionado un determinado par de categorías, se procede a realizar nuevas fusiones de los valores del pronosticador, considerando una categoría menos.
- iv. El proceso acaba cuando ya no se pueden realizar fusiones debido a valores de chi-cuadrado significativo.

4.6.2.2 Selección de los mejores pronosticadores

Para identificar los mejores pronosticadores, se debe realizar el cálculo de chi-cuadrado para cada una de las variables independientes y comparar su significación con respecto a las demás, sin embargo, se recomienda modificar el nivel de significación con el ajuste de Bonferroni.

El ajuste de Bonferroni indica que la significación total (p_T), para B pruebas de significación, debe ser menor o igual que la suma de cada una de las significaciones (p_i).

$$p_T \leq \sum_{i=1}^B p_i$$

El número posible de pruebas de significación (B) se puede calcular a partir de fórmulas combinatorias a partir del número de categorías iniciales de la variable (c) y del número de agrupación por categoría (k), por lo que para variables nominales:

$$B = \sum_{i=0}^{K-1} (-1)^i \frac{(k-i)^c}{i! (k-i)!}$$

Para variables ordinales:

$$B = \binom{c-1}{k-1}$$

Y para variables ordinales con parámetros perdidos:

$$B = \frac{k-1 + k(c-k)}{c-1} \binom{c-1}{k-1}$$

Se debe multiplicar la significación de chi-cuadrado por el resultado de B para evitar el riesgo de rechazo inadecuado de hipótesis por realizar múltiples ensayos.

Una vez realizada la primera segmentación, se debe realizar este procedimiento con cada uno de los grupos generadas por la primera segmentación.

4.6.2.3 Finalización proceso de segmentación

En caso de que no existan límites el proceso de segmentación, este podría generar grupos con muy pocos individuos en cada uno de los nodos terminales, es por ello que se definen los siguientes criterios:

a. Filtro de significación

Este filtro elimina las segmentaciones que no posean un mínimo de significación para el modelo, significación que es un parámetro de entrada, generalmente 95% ($\alpha = 0.05$).

b. Filtro de asociación

Es una función análoga a los filtros de significación de pronosticadores, se pueden aplicar a los siguientes coeficientes de significación: Phi, V de Cramer, Coeficiente de Contingencia, etc., en este caso, se detiene la segmentación no por su grado de significación, sino, porque su grado de asociación es bajo con respecto a un valor dado. Este filtro es eficiente a niveles más profundos del árbol de decisión.

c. Filtros de tamaño

Su principal función es que no se pueda seguir agregando segmentaciones con pocos individuos pertenecientes a un nodo.

d. Filtros por nivel

Permite asignar un número máximo de profundidad del árbol de decisión.

4.7 Support Vector Machine

Algoritmo desarrollado por Vapnik y sus colaboradores en AT&T, el cual ha tenido excelentes resultados en reconocimiento de escritura a mano, clasificación de imágenes, análisis de secuencias de ADN, entre otras, que puede ser utilizada para regresión como para clasificación (Srivastava & Bhambu, 2010). Se basa en transformar el espacio de entrada en otra dimensión superior donde el problema pueda ser resuelto con la incorporación de un hiperplano óptimo (de máximo margen).

De acuerdo a Burgueño (2013), *Support Vector Machine* de forma general, está compuesto de 4 partes básicas:

a. Hiperplano separador

Como se puede ver en la Figura 4.6, los datos se encuentran dispuestos de manera tal, que se puede dibujar una línea recta entre los datos de color azul y rojos, donde la función que define esa recta es el hiperplano separador.

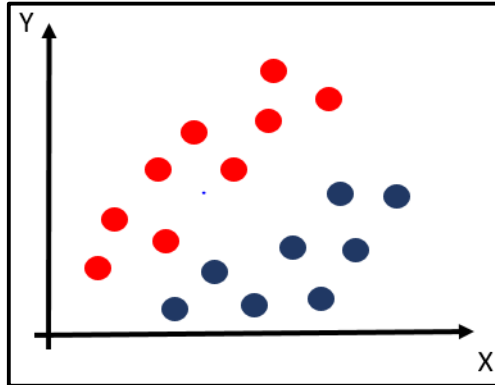


Figura 4.6: Gráfico de datos clasificados en 2 segmentos. Fuente: Elaboración propia.

En *support vector machine*, el objetivo es identificar el hiperplano separador óptimo en el espacio de características χ' .

Este espacio de características no es otra cosa que una transformación no lineal $\Phi(x)$ de altas dimensiones del espacio de entrada χ (espacio original de los datos).

$$\Phi: \chi \rightarrow \chi'$$

$$x \mapsto \mathbf{x} := \Phi(x)$$

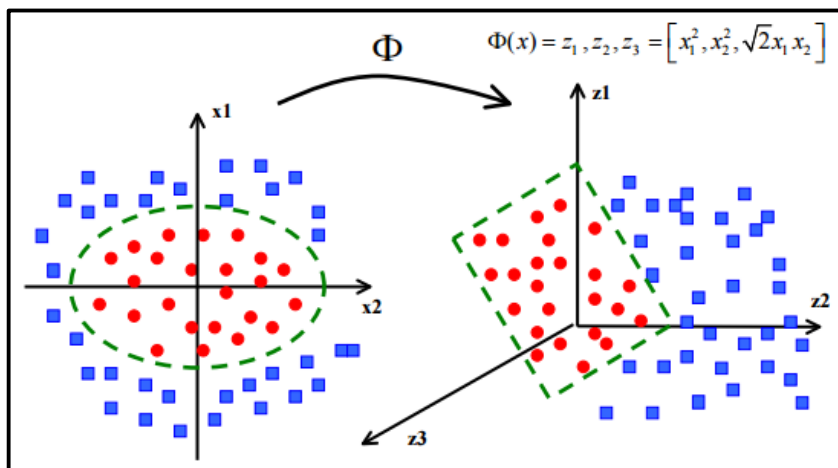


Figura 4.7: Transformación no lineal. Fuente: García Díaz, 2005.

Como se puede ver en la Figura 4.7, la función del hiperplano es generar una nueva disposición de los datos en una dimensión mayor, para establecer el hiperplano óptimo que permitirá segmentar los datos con el menor error posible.

b. Hiperplano de máximo margen

En la Figura 4.6 se puede ver claramente como los datos pueden ser divididos mediante una línea imaginaria, sin embargo, esta línea tiene infinitas opciones, como se ve en la Figura 4.8:

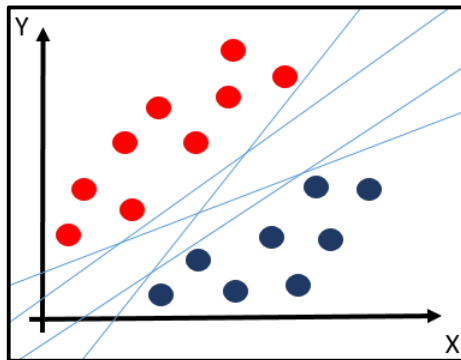


Figura 4.8: Hiperplanos candidatos para datos graficados. Fuente: Elaboración propia.

Support Vector Machines, busca el hiperplano óptimo utilizando el concepto de maximización del margen entre el hiperplano formulado y la distancia con los puntos de cada segmentación, como se muestra en la siguiente figura:

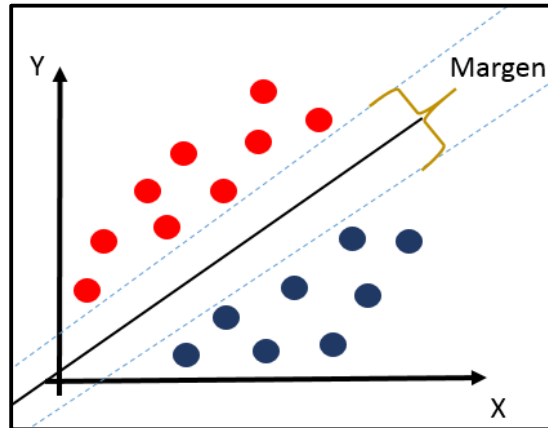


Figura 4.9: Hiperplano de máximo margen. Fuente: Elaboración propia.

c. El margen blando

Se refiere a la definición del hiperplano óptimo, en el sentido de que el proceso de entrenamiento permita que existan datos mal clasificados, lo que, debe ser considerado como *outliers*, sin embargo, el objetivo de todo modelo de clasificación es la búsqueda de minimización de este error, por lo que se incorpora una función de costo, convirtiéndose en un parámetro de entrada en el proceso de entrenamiento, gráficamente, el margen blando se puede visualizar como en la Figura 4.10:

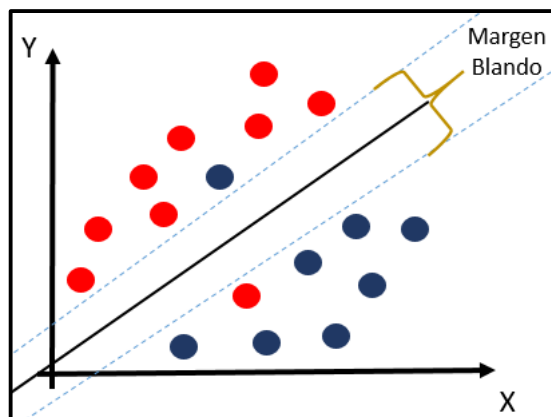


Figura 4.10: Margen blando. Fuente: Elaboración propia.

d. La función *kernel*

Como se puede ver en la Figura 4.8, la determinación del hiperplano óptimo es relativamente fácil ya que es lineal, en el caso de tener datos reales, es posible aplicar ecuación de recta en dos dimensiones y se acabó el problema, sin embargo, como se puede ver en la Figura 4.9, en el gráfico de la izquierda, si bien la distribución de los datos y sus clasificaciones a simple vista es sencilla de definir, matemáticamente es complejo el problema ya que no es un problema lineal. Para solucionar lo anterior, *Support vector machine* permite cambiar, en el caso que sea necesario, su enfoque de búsqueda y ajuste del hiperplano, esto a través del concepto *kernel*.

De acuerdo con Burgueño (2013), la función de *kernel* se define como:

$$K(x_i, y_j) = \phi^T(x_i)\phi(x_j)$$

Este es el principio con el que se logra el aumento de dimensionalidad que se muestra en la Figura 4.7, y por ende, pueden existir infinitas opciones de *kernel* a utilizar, sin embargo, se muestran aquellos que son los más utilizados y que tienen demostraciones de su efectividad:

- i. *Kernel* Radial: función base Radial (también conocido como *Kernel* Gaussiano).

$$K(x_i, y_j) = e^{-\gamma \|x_i - x_j\|^2}$$

- ii. *Kernel* Lineal: función lineal.

$$K(x_i, y_j) = x_i^T x_j$$

iii. *Kernel* Polinomial: Función polinomial de grado d

$$K(x_i, y_j) = (\gamma(x_i^T x_j + 1))^d$$

iv. *Kernel* Sigmoide: también utilizada por redes neuronales.

$$K(x_i, y_j) = \tanh(\gamma x_i^T x_j + d)$$

4.8 Comparación de modelos

4.8.1 Métricas de evaluación

Para problemas de clasificación binaria como el planteado en este proyecto, serán utilizadas técnicas de medida de evaluación de los modelos que generalmente son utilizadas (Barrientos & Ríos, 2013), para someter a comparación los distintos modelos estimados.

4.8.1.1 Matriz de confusión

Sea la siguiente matriz de confusión para cualquier modelo de clasificación binario:

		Predicción	
		Condición predictiva positiva	Condición predictiva negativa
Real	Condición positiva	Positivos reales (TP)	Negativos falsos (FN)
	Condición negativa	Positivos falsos (FP)	Negativos reales (TN)

Tabla 4.1: Elementos de una matriz de confusión. Fuente: Elaboración propia.

De acuerdo a la clasificación anterior, se describen sus componentes:

- Total: N° de casos a clasificar dentro de la muestra.
- Positivos reales (*True Positive*): Corresponde al número de casos clasificado como positivos y que efectivamente son positivos.

- Positivos falsos (*False Positive*): Corresponde al número de casos clasificado como positivos pero realmente son negativos.
- Negativos falsos (*False Negative*): Corresponde al número de casos clasificado como negativo pero realmente son positivos.
- Negativos reales (*True negative*): Corresponde al número de casos clasificado como negativo y efectivamente son negativos.

Con las definiciones anteriores se puede establecer que:

$$N_{total\ de\ casos} = TP + TN + FP + FN$$

$$N_{total\ reales\ positivos} = TP + FP$$

$$N_{total\ reales\ negativos} = TN + FN$$

$$N_{total\ casos\ clasificados\ correctamente} = TP + TN$$

$$N_{total\ casos\ clasificados\ incorrectamente} = FP + FN$$

Por último, las métricas más utilizadas para comparar modelos son las siguientes:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{FP + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F - measure = \frac{Precision * Recall}{Precision + Recall}$$

4.8.2 Función de costos

De acuerdo a la matriz de confusión indicada en el punto 4.8.1.1, el modelo puede contener errores que es necesario cuantificar de alguna forma. En este punto se detallarán funciones de costo para las 4 variables de la matriz de confusión, de tal forma que permita evaluar los resultados de modo cuantitativo. Estas funciones fueron creadas para efectos específicos de esta tesis en conjunto con personas del área de estadística del *Mining Group*.

4.8.2.1 True Positive (TP)

Para el caso de la disposición de esta matriz de confusión, los TP corresponden a cuando el sistema indica que el equipo no tiene fallas, y efectivamente el equipo no falla, por lo tanto, en ese caso no se genera la alerta de falla y el personal de mantenimiento no debe realizar ninguna tarea de mantenimiento, por lo que esta variable se mantiene fuera de la función de costo.

4.8.2.2 False Positive (FP)

Un FP indica que el modelo predictivo clasificó de manera errónea un caso como que no tenía problemas, pero en realidad, el equipo falló. En este caso se puede formular la siguiente función de costo:

$$f_{costo FP} (FP) = P_{ri} * FP * C_{ri} + (1 - P_{ri}) * FP * 2 * C_{HH prom} * HH_{prom}$$

Donde:

FP: Número de casos clasificados como *False Positive*.

P_{ri}: Probabilidad estimada de Reparación imprevista.

C_{ri} : Costo estimado de Reparación imprevista.

$C_{HH\ prom}$: Costo horas hombre promedio.

HH_{prom} : Cantidad de horas hombre promedio por reparación.

El resultado de esta función es expresado en dólares.

4.8.2.3 False Negative (FN)

Un FN indica que el modelo matemático clasificó un evento como que el equipo falla, sin embargo, en realidad no existía problema alguno, por lo tanto, la función de costo se reduce a la siguiente expresión:

$$f_{costo\ FN}(FN) = FN * 2 * C_{HH\ prom\ 2} * HH_{prom\ 2}$$

Donde:

FN : Número de casos clasificados como False Negative.

$C_{HH\ prom\ 2}$: Costo hora hombre por trabajo no necesario de realizar.

$HH_{prom\ 2}$: Cantidad de horas hombre promedio utilizada para diagnosticar la no ocurrencia del evento.

El resultado de esta función es expresado en dólares.

4.8.2.4 True Negative (TN)

Un TN indica que el modelo matemático clasificó un evento como que la falla en el motor ocurrirá, y efectivamente el equipo tiene problemas, en este caso, la función de costo es la siguiente:

$$f_{costo\ TN}(TN) = P_{ri} * TN * C_{ri} + (1 - P_{ri}) * TN * 2 * C_{HH\ prom} * HH_{prom}$$

TN : Número de casos clasificados como *True Negative*.

P_{ri} : Probabilidad estimada de Reparación imprevista.

C_{ri} : Costo estimado de Reparación imprevista.

$C_{HH\ prom}$: Costo horas hombre promedio.

HH_{prom} : Cantidad de horas hombre promedio por reparación.

El resultado de esta función es expresado en dólares.

5 Planteamiento estratégico

El *Mining Group* se encuentra en un periodo de bajas solicitudes por cotizaciones de proyectos, existiendo algunas de tipo marginal dentro de los proyectos ya existentes. Es por lo anterior, que se busca potenciar el área aprovechando el periodo de baja carga de trabajo, mediante la optimización de procesos, creación de nuevos servicios, buscando siempre la oportunidad de realizar mejora en los procesos de entrega del servicio a sus clientes.

La visión y misión del *Mining Group* se alinea con la de Distribuidora Cummins Chile S.A., siendo la que se describe a continuación:

5.1 Visión

- Mejorar la vida de las personas liberando la fuerza de Cummins.

5.2 Misión

- Exigir que todo lo que hagamos lleve a un medio ambiente más limpio, más saludable y más seguro.
- Motivar a la gente a actuar como dueños del negocio trabajando juntos.
- Aliarnos con nuestros clientes para asegurarnos que sean exitosos.
- Superar las expectativas del cliente siendo siempre los primeros en el mercado con los mejores productos.
- Crear riqueza para todas las partes interesadas.

De lo planteado en la visión y misión, deben enfocar todos los esfuerzos en mejorar la experiencia del cliente a través de los procesos productivos, ya que es esa la interacción que tiene la empresa con el cliente (con las distintas faenas dentro del país), y no se debe pensar que es sólo el resultado final el que influye sobre la opinión de los distintos clientes, teniendo claro que el proceso de mantenimiento en sus

distintos niveles es crítico para el desarrollo de la relación entre el *Mining Group* y el cliente final.

En el punto 5.3 se profundizará en el posicionamiento estratégico de Distribuidora Cummins Chile S.A. mediante el modelo Delta desarrollado por Arnoldo Hax.

5.3 Modelo delta

Esta herramienta permite clasificar las estrategias que adoptan las empresas para determinar el rumbo que quieren (o deben) seguir, propuesto por Arnoldo Hax, quién además de considerar lo que Porter determina como las características que una empresa debe tener para ser exitosa, incorpora la dimensión del cliente, el cual es muy relevante cuando la estrategia de la empresa que se pretende analizar tiene a su cliente como centro de todas sus operaciones, convirtiéndose en una herramienta de análisis integral.

Debido a la estructura de unidades de negocio que forma a Distribuidora Cummins Chile S.A. descrita en el punto 2.2.2.1, ésta debe asumir varias estrategias al mismo tiempo como empresa, pero se enfoca en cada una de las unidades de negocio, como por ejemplo, el área de *Engines* se enfoca en mejor producto, y esto se refleja en los esfuerzos de innovación en la tecnología utilizada en los procesos de fabricación de los mismos, junto con la investigación y capacitación que necesita recibir su personal, por otro lado, se encuentra la unidad de negocio que es *Shops and Services*, quién se enfoca en Redefinición relación con el cliente ya que tienen relación directa y constante. Para el *Mining Group*, que es la unidad de negocio donde será desarrollado este proyecto de memoria, de acuerdo a las necesidades que el actual ambiente económico del país está exigiendo para que empresas prestadoras de servicios a la minería puedan mantenerse en el tiempo, además de ajustarse a la estrategia corporativa, esta área se enfoca según como se muestra en la figura:

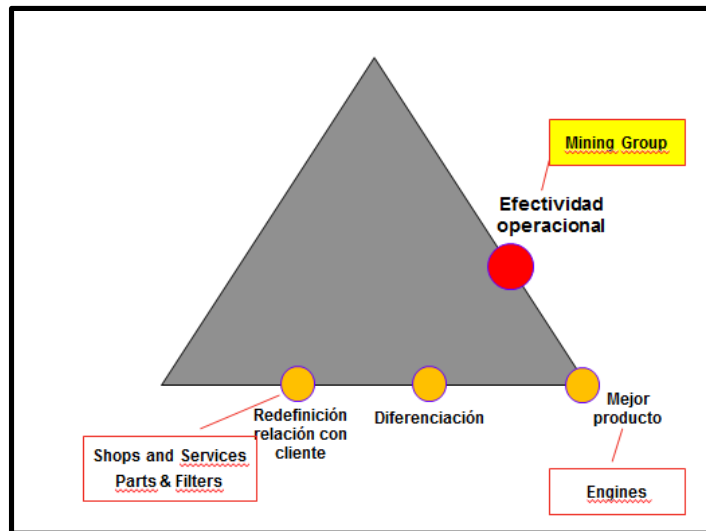


Figura 5.1: Modelo Delta aplicado a Unidades de negocio de Distribuidora Cummins Chile S.A. Fuente: Elaboración propia.

Basado en el modelo Delta, es que se establece que el *Mining Group* enfoca su estrategia en “Efectividad operacional”, la cual será descrita en el siguiente punto.

5.3.1 Efectividad operacional

El *Mining Group* nace de la necesidad que se manifiesta en el año 1949 en la zona norte para atender compañías mineras Anaconda, Bethlehem Steel y Kennecot Copper, los cuales provocaron un crecimiento desmedido de esta área, teniendo un auge sostenido en el tiempo, principalmente las áreas de operaciones (logística, abastecimiento, gestión y coordinación en faena entre cliente) y administrativas en faena (estados de pago, coordinación con casa matriz, entre otros). Lo anterior, generó un efecto en que si bien, el área logra ventas y margen bruto que son los de mayor importancia para la empresa, esta área posee problemas graves de eficiencia operacional, tales como:

- Duplicación de procesos.
- Redundancia de información.
- Aumento de costo operacional en faenas.

- Procesos productivos y administrativos no integrados de manera correcta a operación.
- Recurso humano no utilizado de manera óptima.
- Mejora y análisis de procesos productivos es casi nulo.
- Calidad de información no óptima entre procesos.

El punto 5.3.1 es la base en la cual *Mining Group* centra sus esfuerzos para mantener y mejorar su posición de líder actual en el mercado chileno, en la industria de la minería, siendo el desarrollo de este punto como objetivo para mejorar sus ventajas competitivas.

5.4 Mapa estratégico *Mining Group*

El *Mining Group* tiene 4 pilares fundamentales, los cuales desarrolla en cada una de sus actividades con el fin de lograr los objetivos estratégicos planteados, los cuales son:

- Clientes
- Procesos
- Recursos humanos
- Resultados financieros

Los resultados financieros están posicionados en último lugar ya que para este periodo en particular se considera (sólo en el *Mining Group*) un ingreso cautivo con incrementos marginales que se mantendrán por al menos 3 a 4 años desde el 2014, siendo las proyecciones de venta quienes indican que después de este receso vendrá un incremento de inversión por parte de las faenas mineras, y por ende, una sobrecarga sobre los procesos productivos actuales.

A continuación se muestra un esquema del mapa estratégico para el área de minería:

▪ Mapa estratégico Minería:

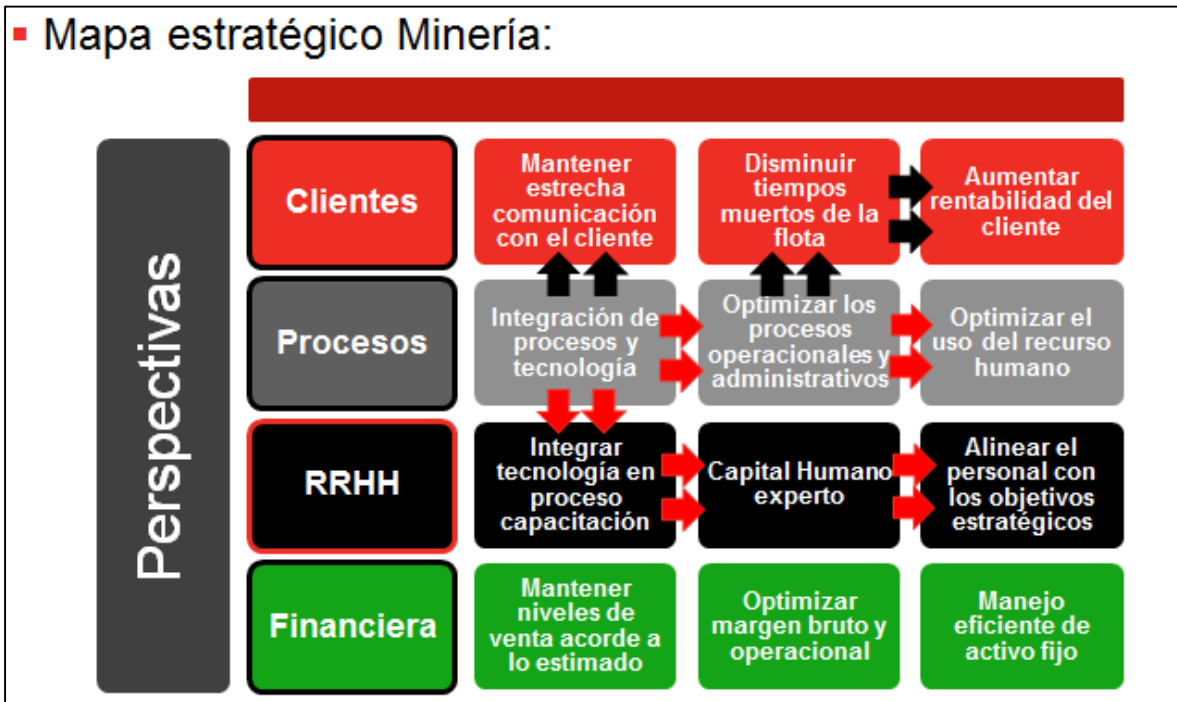


Figura 5.2: Mapa estratégico área minería Distribuidora Cummins Chile S.A. Fuente: Elaboración propia.

En la Figura 5.2 se muestra como al pilar de “Procesos”, se agrega un objetivo llamado “integración de procesos y tecnología”, el cual a la fecha no es considerada y como se ve en la figura, afecta a la mayor parte de los objetivos estratégicos. A pesar de que no se indica de manera explícita en Figura 5.2 que facetas del pilar “Financiera” afecta, es claro que esta acción generará disminución en costos asociados a la producción debido a que provocará un efecto directo en la eficiencia de operación.

Al considerar mejorar aspectos relacionados a la operación y tecnología, se busca obtener mayor información para gestionar de mejor forma los recursos, entregando beneficios tales como:

- Clientes:
 - Mayor capacidad de respuesta ante imprevistos, entregando soluciones óptimas que se adelanten a las necesidades del cliente.

- Mantener informado al cliente (sin que él solicite la información) implica que se sentirá respaldado por una empresa confiable.
 - Permite aumentar la disponibilidad, confiabilidad y rentabilidad de la flota para el cliente.
- Procesos:
 - Al incorporar nueva tecnología, estos se volverán más eficientes, entregando información relevante a usuarios, volviéndolos eficaces en su operación.
 - La entrega de mayor calidad de información permite optimizar el uso del recurso humano existente, y es aquí donde existe un punto crítico dentro de la operación en este tipo de industria.
- Recursos humanos:
 - El recurso humano debe ser gestionado de acuerdo a sus capacidades, junto con las necesidades del negocio, por ende, este debe ser clasificado de tal manera que permita tomar acciones cuando sean necesario, y este es un punto importante cuando es un recurso finito, muy difícil de conseguir, además, de requerir que sea especializado, por lo que una salida de personal es muy costoso para la empresa.

Los procesos productivos pueden entregar información valiosa en cuanto a que tipo de personal es requerido en el momento o qué tipo de persona será requerido de acuerdo a como se analice la información existente.

Si bien todos los puntos mencionados son importantes, es claro que durante el desarrollo de este trabajo el punto de desarrollo será el de procesos, por la urgencia y el impacto que causará en el cliente y en la compañía completa.

5.5 Modelo de negocio

En la Figura 5.3 se muestra el modelo de negocio del *Mining Group* en Distribuidora Cummins Chile S.A.:

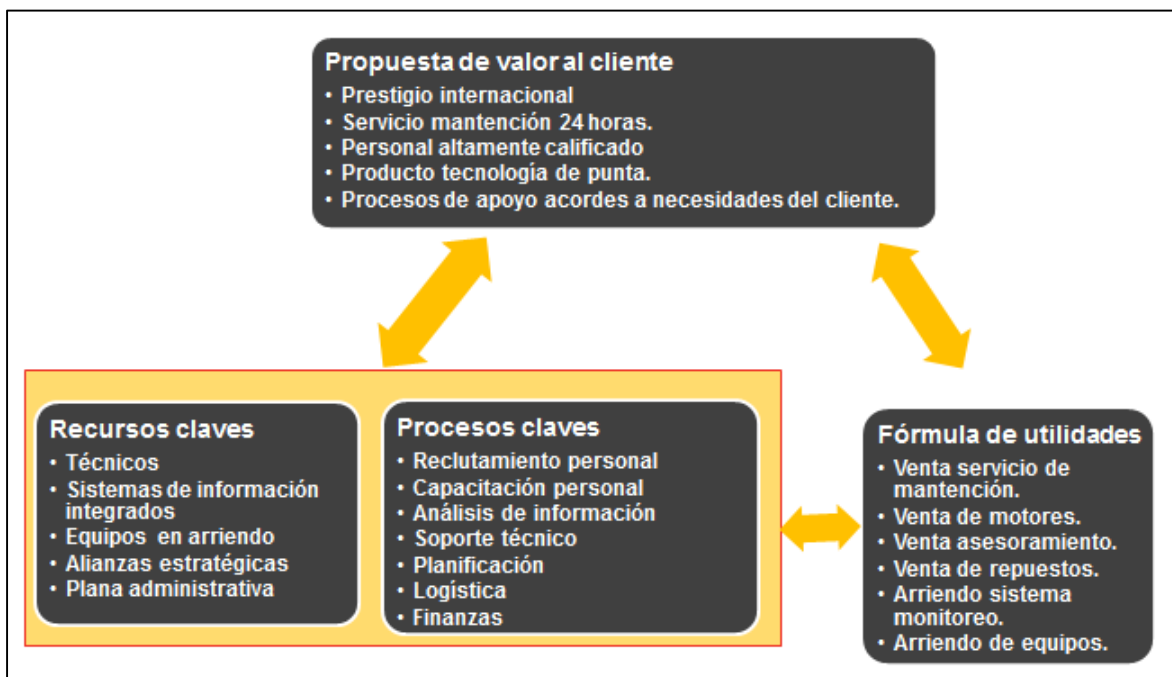


Figura 5.3: Modelo de negocios *Mining Group* Distribuidora Cummins Chile S.A. Fuente: Elaboración propia.

A continuación se detalla el modelo de negocios:

- Propuesta de valor al cliente:
 - Los clientes objetivos son todas aquellas faenas mineras que cuentan con operaciones a rajo abierto u operaciones subterráneas en los cuales exista maquinaria de extracción de mineral marca Komatsu o motores marca Cummins.
 - El principal problema de las faenas mineras es el aseguramiento de forma óptima de la extracción de la mayor cantidad de mineral las 24 horas del día, esto se realiza a través de maquinaria marca Komatsu

con motores Cummins, por lo que el trabajo a realizar para satisfacer esta necesidad es un servicio de mantenimiento integral.

- La oferta que realiza Distribuidora Cummins Chile S.A. a través del *Mining Group* son tipos de contratos que se adaptan a las necesidades del cliente, el cual ofrece desde contratos de asesoramiento técnico hasta gestión de motores de respaldo, todo esto en búsqueda de minimizar los tiempos perdidos de equipos de extracción de mineral y ajustándose a lo que el cliente necesite.
 - Ofrece servicio de 24 horas de asesoría técnica, servicios de mantención, siempre dependiendo de lo que el cliente requiera. El apoyo técnico es brindado desde la mano de obra en faena (técnicos y mecánicos) hasta altas jerarquías ubicadas en casa matriz dentro de la unidad.
- Recursos claves
 - Dentro de los recursos clave, el principal es la mano de obra en faena (técnicos y supervisores), quienes tienen directa relación con el cliente, son el rostro visible de la empresa. Existe recurso administrativo en faena y en casa matriz, quienes dan soporte directo a faena y se encargan de temas distintos al quehacer diario de la operación.
 - Son importante los sistemas de información integrados que actualmente dan soporte a la operación en faena, esto, ya que permite monitorear a distancia el estado de cada motor.
 - Por último, los equipos en arriendo (motores de respaldo), que son equipos dispuestos para el cliente, en caso de imprevistos, para que no se pierda tiempo de operación en reparaciones y obtenga continuidad operacional.

- Procesos claves:
 - Reclutamiento y capacitación de personal son clave ya que van de la mano con la misión indicada en el punto 5.4.
 - Análisis de información, soporte técnico, planificación y logística son procesos que están conectados entre sí y son críticos debido a su influencia en los tiempos promedio de respuesta al cliente cuando tiene requerimientos específicos.
 - Los procesos de planificación y gestión financiera son relevantes debido a los montos que se manejan en esta industria, es necesario una base financiera sólida que sea capaz de reaccionar al tiempo que el negocio de la minería lo requiera.

- Fórmula de utilidades
 - Venta de servicios de mantención por contrato con faena.
 - Venta de motores de alta potencia para minería.
 - Venta de asistencia técnica.
 - Venta de insumos para motores Cummins.
 - Arriendo sistema de monitoreo remoto.
 - Arriendo de equipos de respaldo.

Con la información entregada se espera que el lector pueda entender de mejor forma el medio ambiente donde será desarrollado este proyecto de título junto con la empresa.

6 Arquitectura de macroprocesos

El diseño de los procesos de mantenimiento predictivo es una particularización de los procesos de negocio que fueron elaborados por el profesor Oscar Barros, para lo cual, los diagramas que están en notación IDEF0 fueron desarrollados en iGrafx y los procesos se encuentran en BPMN 2.0, para lo cual se utilizó Bizagi Process Modeler.

6.1 Modelamiento del Rediseño

Para efectos de análisis desde una visión general de los macroprocesos, se muestra la Figura 6.1 correspondiente a la interrelación genérica dentro de una empresa:

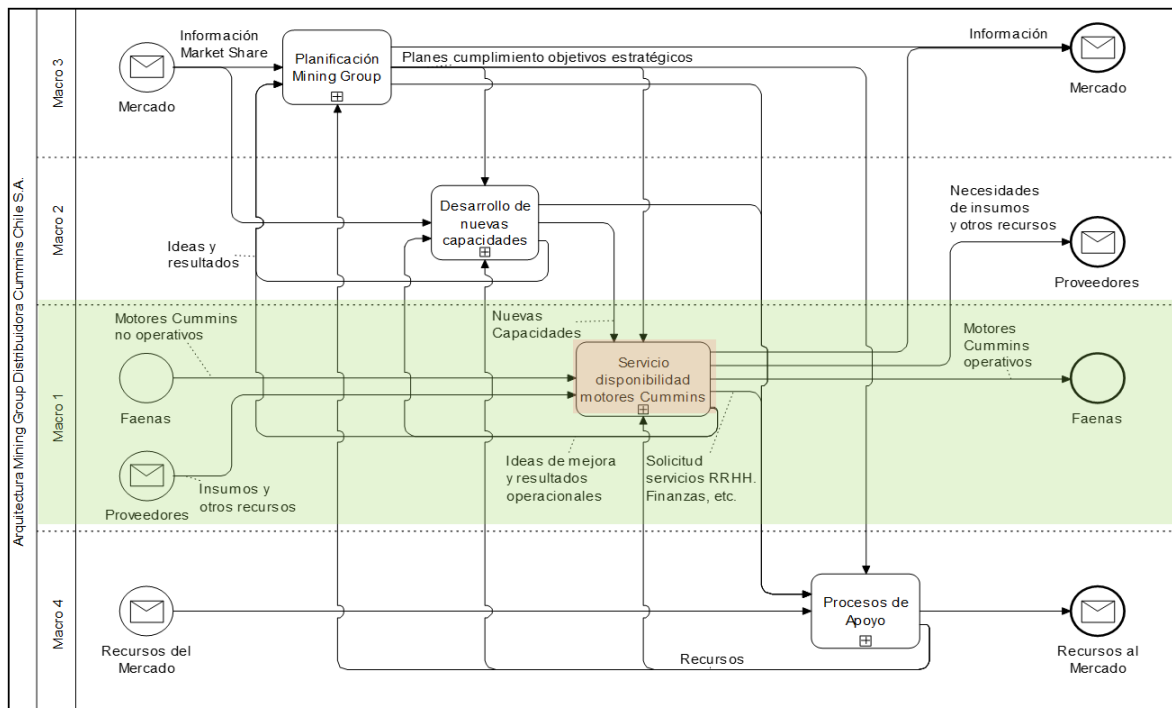


Figura 6.1: Arquitectura de macroprocesos Mining Group Distribuidora Cummins Chile. Fuente: Oscar Barros V. 2014.

De acuerdo al patrón seleccionado, se desprende que el desarrollo de este proyecto estará basado en Macro 1: Cadena de Valor, al que en el *Mining Group* llamaremos “Servicio disponibilidad motores Cummins”.

- Macro 1: Cadena de Valor.

El proceso que será diseñado modificará el comportamiento de las *Stream Value* que existen y cuya función es poner en funcionamiento el servicio principal de esta área que es el servicio de mantenimiento de motores Cummins que se encuentran en faena, específicamente en Macro 1, la llamaremos “Servicio disponibilidad motores Cummins”, se incorporarán nuevos procesos que permitirán apoyar a los procesos que ya están definidos, con esto, será posible modificar la forma de planificar y gestionar los servicios de mantenimiento, específicamente, el mantenimiento predictivo, permitiendo agregar el nivel de sofisticación que se busca para lograr la ventaja competitiva.

A continuación, se profundizará en Macro 1 para especificar de mejor manera cómo afectará el desarrollo de este proyecto en la estructura actual de la empresa.

6.2 Macro 1: Cadena de valor

En la siguiente figura se muestra Macro 1 con respecto a la cadena de valor del *Mining Group*:

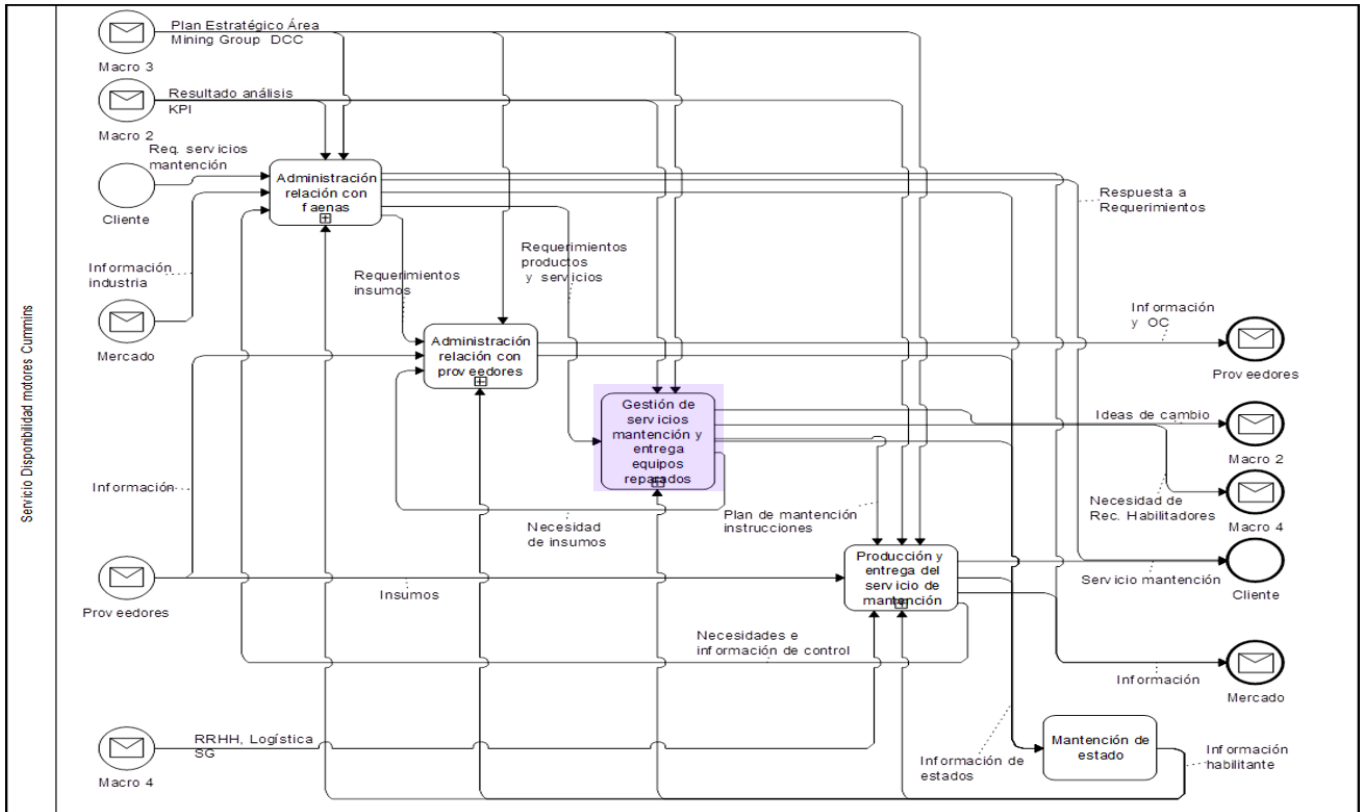


Figura 6.2: Primer nivel macro 1. Fuente: Oscar Barros 2014.

Los procesos que serán afectados por el rediseño se describen a continuación:

- **Gestión de servicios mantenimiento y entrega equipos reparados**

Se destaca con color morado debido a que en este proceso se profundizará para generar el cambio que persigue este proyecto, es en este proceso donde se realiza la planificación, programación y control necesario para la entrega de los servicios requeridos.

- **Producción y entrega del servicio de mantenimiento**

Este proceso se verá afectado en menor medida, ya que el servicio seguirá siendo entregado de la forma actual (mismo *value stream*), sin embargo, se hará con mayor calidad de información, esta información será procesada desde Gestión de servicios mantenimiento y entrega equipos reparados.

Para mostrar mayor detalle de cómo afectará las modificaciones en Macro 1, se muestran las aperturas más importantes:

6.2.1 Gestión de servicios de mantenimiento y entrega equipos reparados

Este proceso es uno de los más importantes, donde las mejoras tendrán efectos directamente relacionados a la operación, debido a su función principal que es planificación. A continuación se muestra la apertura de este proceso:

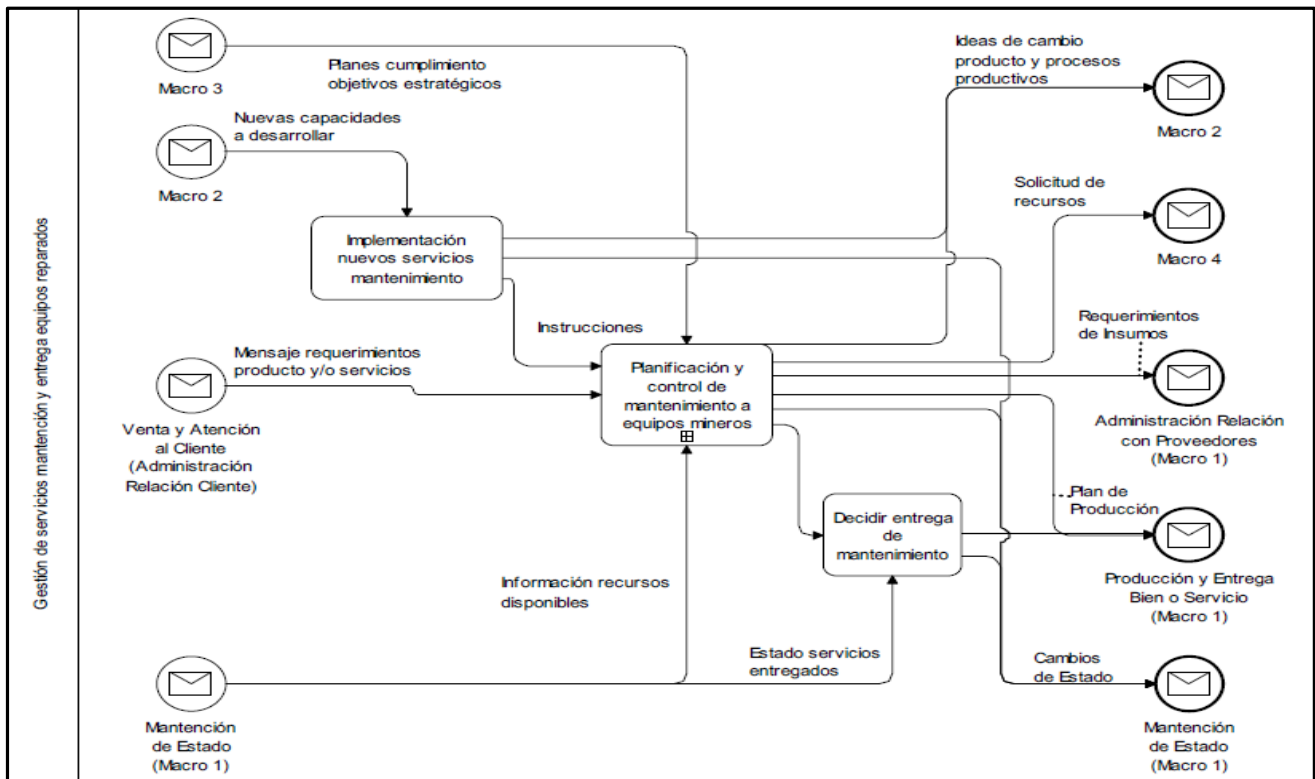


Figura 6.3: Apertura Gestión de servicios mantenimiento y entrega equipos reparados. Fuente: Oscar Barros 2014.

- **Planificación y control de mantenimiento a equipos mineros:** En este proceso se definen los recursos a utilizar para entregar el servicio de mantenimiento, por lo que es muy importante para indicar claramente donde el rediseño afectará dentro de la cadena de valor. A continuación se muestra la apertura de este proceso:

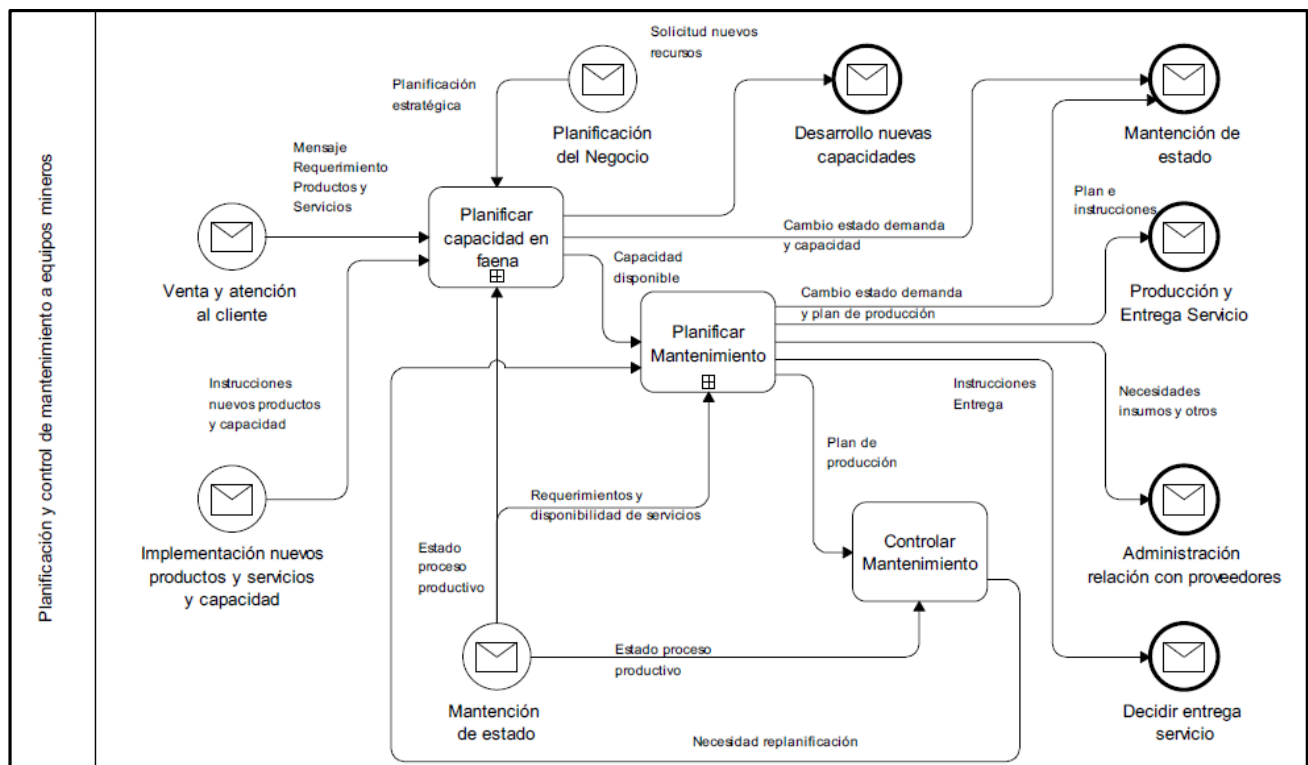


Figura 6.4: Planificación y control de mantenimiento a equipos mineros. Fuente: Oscar Barros 2014.

Se puede ver que el proceso que se será rediseñado es Planificar Mantenimiento, se deberá agregar el paso proceso de mantenimiento predictivo donde se ejecuta el proceso de estimación de mantenimiento preventivo y mantenimiento correctivo, el cual deberá entregar un listado de prioridades basados en la información de los equipos que deben ser revisados, además, para que el sistema entregue información de forma proactiva, se diseñará, basándose en el mantenimiento predictivo un proceso de control, de tal forma que el sistema pueda generar alertas de acuerdo a ciertos

parámetros, dando aviso a planificación de la demanda para que pueda calibrar el modelo matemático, por otro lado, es importante medir el rendimiento de las predicciones del modelo predictivo, por lo que se diseña el proceso que se encuentra dentro de Controlar Mantenimiento, cuyo proceso BPMN se encuentra en el Anexo 14.3.

Se muestra la apertura de Planificar Mantenimiento:

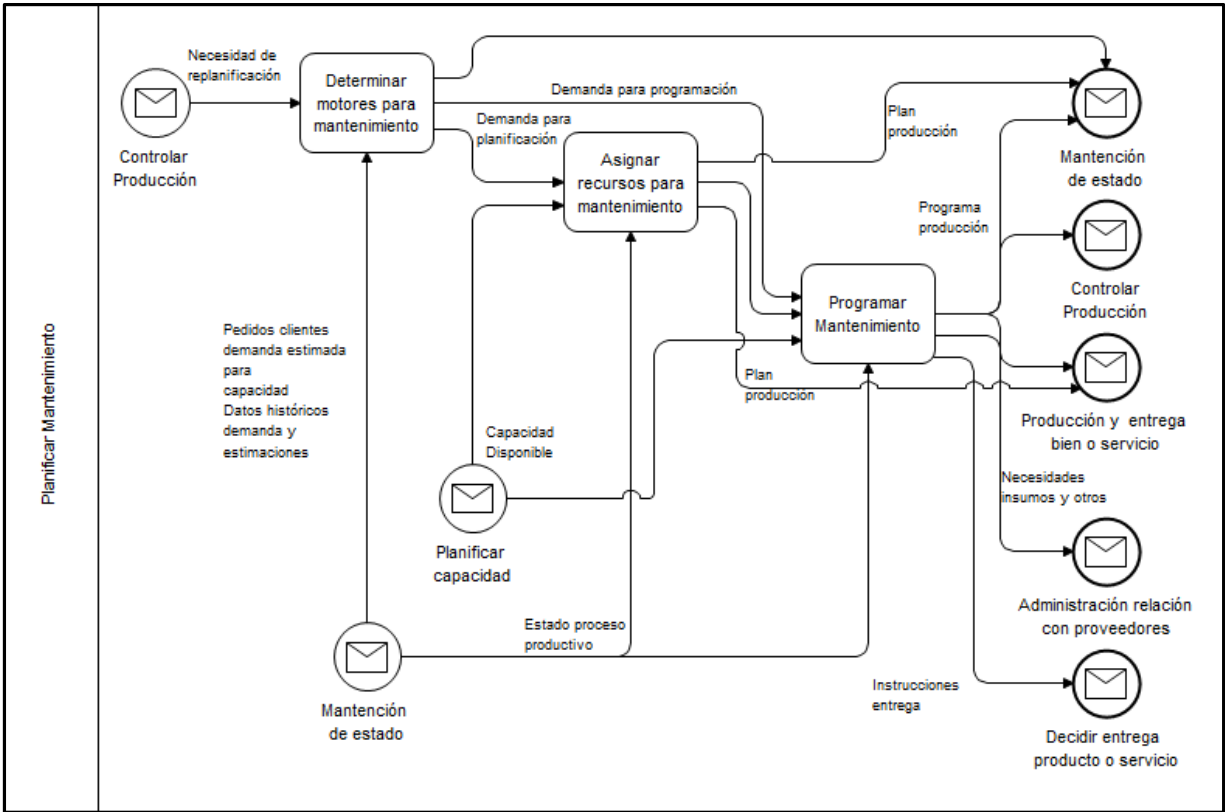


Figura 6.5: Planificar Mantenimiento. Fuente: Oscar Barros 2014.

Dentro de Planificar Mantenimiento, se encuentra el proceso que deberá ser rediseñado:

- Determinar motores para mantenimiento
 1. Este es el punto principal a desarrollar para efectos del presente proyecto, es donde se aplicarán todos los esfuerzos necesarios en búsqueda de la

información relevante para alcanzar el objetivo principal que es generar un modelo predictivo que pueda ayudar a detectar (o evitar) fallas imprevistas. En el anexo 14.1 se muestra la apertura de este proceso como actualmente se ejecuta (sin considerar el mantenimiento predictivo). El proceso inicia desde faena, donde los mecánicos retroalimentan al ingeniero de planificación, el cual de forma paralela está consolidando la información para mantenimiento correctivo y preventivo.

2. Lo anterior se realiza mediante la información que aportan los mecánicos a través de las bitácoras de equipo (para el caso de mantenimiento preventivo) y las alertas directas por el personal de mantenimiento para la planificación del mantenimiento correctivo.
3. Dentro de este proceso, se agregan dos procesos que permiten establecer los candidatos debido a la ejecución de mantenimiento predictivo y determinación del plan por mantenimiento predictivo.

Para el caso de la caja que se llama “Determinación de equipos Mantenimiento Predictivo”, en el anexo 14.2 se encuentra el BPMN correspondiente a este proceso. La importancia de este proceso, es la interacción del usuario y del sistema, el cual se comporta como un asesor de acuerdo al análisis de la información existente en los sistemas sobre los equipos, de cuáles equipos deben recibir mantención, permitiendo al usuario discriminar los equipos a intervenir de acuerdo a factores internos (sensores) y externos (faenas), además, es en este proceso donde se especifica la forma en que el sistema de forma automática entregará información sobre posibles candidatos para ser considerado en la programación de mantenimiento predictivo de acuerdo a probabilidad de falla y frecuencia que se establezca en el sistema.

6.3 Lógica solución propuesta

La solución propuesta es la implementación de un modelo predictivo basado en minería de datos, capaz de detectar fallas oportunamente dentro de las distintas variables que son monitoreados por el sistema remoto Specto en los motores que administra Distribuidora Cummins Chile.

El uso del modelo predictivo debe aplicarse cada día (dependiendo de la complejidad del algoritmo y del tamaño de la base de datos), permitiendo detectar posibles motores con riesgo de sufrir alguna falla para aplicar las acciones que sean definidas. Para lograr lo anterior, deben ser considerados varios factores que son mencionados a continuación.

6.3.1 Modelos de motores

Distribuidora Cummins Chile S.A. tiene varios modelos de motores dedicados a la alta minería, el cual se distribuye dependiendo de factores asociados a las necesidades del cliente final.

Algunos modelos de motor más importantes presentes en faenas son:

- QSK78
- QSK60
- QSK45
- QSK19
- K2000
- K50

El listado de motores se encuentra ordenado en forma descendente basado en su potencia (o consumo de combustible), siendo las fallas de los motores más potentes los que generan mayor costos para el área, y por ende, las que concentran el esfuerzo para el análisis de las causas de origen y su posible prevención.

6.3.2 Faenas mineras

Los motores de Distribuidora Cummins Chile se encuentran distribuidos en las distintas faenas mineras a lo largo de Chile, estas tienen características específicas dependiendo de su ubicación en el país. Estas variables influyen en la performance de rendimiento y comportamiento del motor, por lo que deben ser considerados al momento de generar el modelo matemático de predicción. Los aspectos a considerar son:

- Altura
- Mineral a extraer
- Presión atmosférica
- Temperatura
- Clima
- Terreno
- Tipo faena (rajo abierto – mina subterránea)

6.3.3 Frecuencia de fallas

Dependiendo del tipo de faena y motor, existe distinta frecuencia en los tipos de fallas, y por ende, también es una variable a considerar cuando se esté buscando el modelo matemático a utilizar. En el Gráfico 6.1 se puede ver la frecuencia de los distintos tipos de falla, para el periodo de diciembre 2013 hasta agosto 2014, para los motores QSK60 y QSK45:

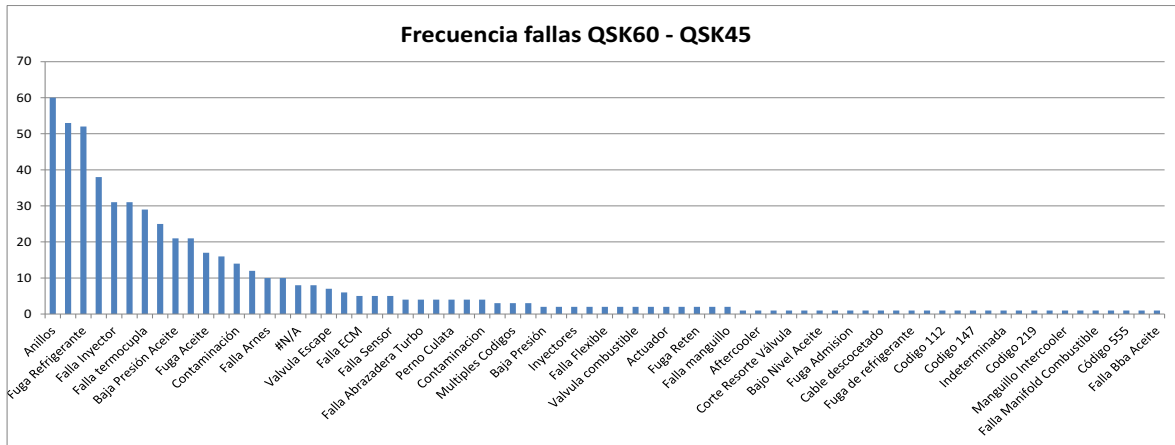


Gráfico 6.1: Frecuencia fallas motores QSK60 – QSK45 periodo Diciembre 2013 – Agosto 2014. Fuente: Elaboración propia.

Una de las fallas que tienen mayor cantidad frecuencia es lo que se denomina como “Anillos”, lo cual es válido si lo consideramos como un análisis agrupado, luego, para acotar el universo de información disponible, se deberá revisar el tipo de falla de acuerdo a su frecuencia y en qué tipo de motor y faena se está generando el problema como aproximación inicial de la solución.

6.3.4 Metodología

La metodología a utilizar para el análisis de información será a través del proceso KDD (*Knowledge Discovery in Databases*), el cual permitirá encontrar información relevante dentro del repositorio de datos.

Las fases a desarrollar son las siguientes:

- a. Selección de datos: fuente de datos y datos a utilizar.

Los datos serán extraídos de la base de datos que tiene actualmente el sistema Specto, sistema desarrollado por Distribuidora Cummins Chile para los motores que tienen en faena. Este sistema tiene su base de datos en PostgreSQL, cuyo servidor físico se encuentra administrado por servicios provistos por Amazon. Esta base de datos crece diariamente a un ritmo de 16 gigabytes, por lo que inicialmente se deberá acotar la *data* a utilizar. En un inicio la muestra de datos

a analizar comenzará con un análisis exploratorio de datos, principalmente orientados a:

- Priorizar faenas críticas, de acuerdo a la estrategia o contingencias del momento.
- Identificar frecuencia de fallas, también identificando fallas críticas.
- Costo de fallas, buscando obtener beneficios en disminución de este concepto.
- Cantidad de motores, buscando la oportunidad de disminuir costos basando en volumen.

b. Preprocesamiento: limpieza y datos del punto a.

Limpieza de los datos, buscando puntos atípicos, ruido, incongruencias, completando datos faltantes con ayuda de métodos de limpieza y con conocimiento experto de los datos.

c. Transformación, tratamiento de los datos obtenidos y generación de nuevas variables a partir de las ya existentes (en caso de que sea necesario).

d. *Data Mining*, extracción de patrones a través de distintos modelos.

e. Interpretación y evaluación, donde serán identificados los patrones obtenidos que sean de utilidad para los objetivos de este proyecto.

Para el análisis sugerido, será necesario la habilitación de un servidor de base de datos, que permita extraer y almacenar la información a analizar, junto con la habilitación del software que permita aplicar los métodos de *data mining*.

6.3.5 Fuentes de datos

La información será extraída del sistema Specto que es el sistema de monitoreo de Distribuidora Cummins Chile S.A., cuya base de datos se encuentra en PostgreSQL,

además, tenemos información que debe ser ingresado al *datamart* que será complementaria a los datos en “bruto”, lo cual nos permitirá desde herramientas de bases de datos gestionar de forma eficiente la gran cantidad de información que será procesada.

6.4 Situación Actual

El problema de Distribuidora Cummins Chile S.A. se encuentra en la forma en que son realizados los servicios de mantenimiento en la flota del *Mining Group*.

En el anexo 14.4 se muestra el BPMN Determinar motores para mantenimiento (actual) con el cual se detalla a continuación.

- a. Diariamente se inicia el proceso de discriminación entre mantenimiento preventivo y correctivo a todos los equipos de la flota en cada faena.

Cuando no existe desperfecto en el equipo:

- b. Ejecución mantenimiento rutina:

Pauta breve que es realizada por los mecánicos y/u operadores de cada equipo, ingresando información a la bitácora del equipo revisado.

- c. Ingreso información bitácora por faena:

La información obtenida se ingresa en un formulario basado en Access, que contiene la información de las bitácoras de todos los equipos de la faena. Esta información es ingresada por el ingeniero de planificación de la faena.

- d. Envío bitácora a casa matriz:

Se envía a través de carpetas compartidas en la nube (Google Drive, Dropbox, etc.) el archivo Access con el historial de la bitácora de la faena (y sus equipos).

- e. Consolidación bitácoras nacionales:
Se recibe la información de todas las bitácoras de las distintas faenas del país.
- f. Verificación parámetros de motor:
Es verificado que se encuentre la información correcta de todos los parámetros del motor que son considerados como *input* para decidir si serán considerados en mantenimiento preventivo (horómetro principalmente), lo anterior se realiza para todos los motores de las faenas del país.
- g. Revisión parámetros MP:
Se verifican los parámetros de cada motor para definir cuales equipos se deben ir a mantenimiento preventivo.
- h. Determinación Plan MP:
Son definidos los equipos que deben ser detenidos para realizar el mantenimiento preventivo.
- i. Consolidación plan de Mantenimiento:
Se unifica en un documento todos los equipos que deben ser retirados de la operación por faena, clasificando en mantenimiento correctivo y preventivo.
- j. Envío Plan a faena:
Se envía el plan consolidado a faena para que estos aprueben el trabajo que le corresponde a los motores que están bajo su responsabilidad.
- k. Revisión Plan Mantenimiento:
Las faenas revisan el plan de mantenimiento, si la faena aprueba el plan, se procede al envío de la información a Asignación de recurso para demanda, en

caso contrario, se vuelve a verificar la información para definir un nuevo plan de mantenimiento preventivo y correctivo.

Cuando existe desperfecto en el equipo:

l. Envía reporte MC:

Este reporte contiene información del problema detectado en el equipo, donde se especifica la criticidad de reparación, siendo enviado al ingeniero de planificación de casa matriz.

m. Registro motores con falla:

Se genera un listado de los equipos y las faenas que presentan fallas correctivas, siendo categorizadas y priorizadas según gravedad del problema, entre otros.

n. Determinación plan de equipos a reparar:

Son definidos los equipos que deben recibir servicio de mantenimiento correctivo, y los que quedan en lista de espera.

o. Determinación Plan MC:

Se envía el plan de mantenimiento correctivo para todas las faenas.

Como se puede ver, no existe ningún proceso que incorpore dentro de la planificación actual de la demanda de equipos para mantenimiento, el pronóstico de falla de manera predictiva, sólo se enfoca en mantenimiento preventivo y correctivo.

6.5 Lógica implementación

Será desarrollada la lógica necesaria tras el anexo 14.2, ya que actualmente no se encuentra disponible en la empresa, por lo que de ser necesario será utilizado pseudocódigo para desarrollar algunas lógicas.

La lógica tras este proceso será descrito aprovechando los diagramas de pista que BPMN 2.0 entrega, utilizando al responsable y el nombre de la tarea asociada.

6.5.1 Descripción lógicas requeridas

6.5.1.1 Determinación de equipos Mantenimiento Predictivo

Para el caso en que el sistema es activado por el usuario (pista: Ingeniero de planificación):

- Ingreso Página Web:
El sistema estará basado en un entorno Web, por lo que cada vez que el usuario requiera ingresar y ejecutar el sistema, este deberá ingresar a su navegador e ingresar a la página de inicio. El usuario puede ingresar al sistema las veces que estime conveniente.

- Selecciona alternativa de página Web:
Una vez ingresado a la página Web, el sistema requiere actualizar la información en la base de datos, donde se entregan 4 opciones para la carga de información:
 - a. Información de todos los equipos (hasta la hora de la consulta)
 - b. Información por ESN hasta esa hora
 - c. Información para días anteriores
 - d. No actualizar información

Si el usuario selecciona “No actualizar información” (d), el sistema se cierra.

Si el usuario selecciona a, b o c, el sistema continúa.

- Revisión estado información:

El sistema muestra al usuario un análisis exploratorio de los datos resumidos, cantidad de registros total, errores detectados, registros existentes y periodo de carga, donde el Ingeniero de Planificación debe aceptar o rechazar la muestra por el sistema.

Si la información se encuentra OK, este avanza a la siguiente tarea del flujo, en caso contrario, se vuelve a solicitar la información para verificar los resultados obtenidos.

- Solicitud opción MPred:

Cuando los datos se encuentran “OK” el sistema ofrece la opción de ejecutar (o no) el o los modelos matemáticos predefinido en el sistema, en caso de que el usuario no desee ejecutar el cálculo de mantenimiento predictivo puede cancelar la operación y terminar el programa.

- Revisión de resultados:

Al finalizar el procesamiento de datos, el resultado será un listado de ESN con su probabilidad de ocurrencia asignada y el nodo que entrega este resultado en caso de que corresponda.

El usuario tiene la opción de descargar (o no) la información en formato Excel. Por último, se le pregunta al usuario si requiere realizar un nuevo análisis, si su respuesta es negativa, el sistema envía la información al siguiente paso del proceso superior (Determinación Plan Mantenimiento Predictivo, anexo 14.1)

A continuación se detallan los pasos que debe realizar el sistema (Pista: Sistemas):

- Despliegue interfaz de usuario:

El sistema será capaz de enviar instrucciones al browser del usuario para que pueda ser desplegada en la pantalla una interfaz acorde a lo que se requiere, como ejemplo se muestra la siguiente figura:

The screenshot shows a web-based user interface for 'SPECTO Sistema de Monitoreo Remoto de Motores'. The header includes the Cummins Chile logo and the system name. The main content area is titled 'Selección información a procesar' and contains two filter sections. The 'Filtro Motores' section has two radio buttons: 'Todos los equipos' (selected) and 'Filtro por ESN' with an adjacent empty text input field. The 'Filtro Fecha/Hora' section has two radio buttons: 'Fecha y hora actual sistema: 25-03-2016 - 14:18:00' (selected) and 'Indicar Fecha' with an adjacent date selection dropdown menu. At the bottom are 'Aceptar' and 'Cancelar' buttons.

Figura 6.6: Despliegue interfaz de usuario. Fuente: Elaboración propia.

- Enviar solicitud *data* nueva:

De acuerdo a la selección del usuario, el sistema genera una *query* en lenguaje SQL, el que contiene la instrucción de acuerdo a lo indicado en el paso anterior. En pseudocódigo queda establecido como:

```

1 Función Filtro_data(seleccion_data, seleccion_fecha)
2 //seleccion_data: corresponde a opción elegida (Todos los ESN o ESN específico)
3 //seleccion_fecha: especifica fecha actual (incluyendo hora de sistema) o fecha
4 //pasada (desde las 0:00 hrs hasta las 23:59:59 hrs)
5 //
6 //Suponiendo que a vista entregada pr a base de datos Specto se llama "datos_specto"
7 //el cual contiene 82 campos
8
9 datos_a_extraer <- Select * from 'datos_specto'
10 Si seleccion_data = 'todos los equipos' Entonces
11     Si seleccion_fecha = 'inf.diaria a la hor del sistema' Entonces
12         datos_a_extraer <- datos_a_extraer + ' WHERE fecha_AVL' = ' + fecha/hora_sistema
13     FinSi
14
15     Si seleccion_fecha = 'especificar fecha' Entonces
16         datos_a_extraer <- datos_a_extraer + 'WHERE fecha_especificada
17     FinSi
18 FinSi
19
20 Si seleccion_data = 'filtro por ESN' Entonces
21     Si seleccion_fecha = 'inf.diaria a la hor del sistema' Entonces
22         datos_a_extraer <- datos_a_extraer + ' WHERE fecha_AVL' = ' + fecha/hora_sistema
23     FinSi
24
25     Si seleccion_fecha = 'especificar fecha' Entonces
26         datos_a_extraer <- datos_a_extraer + 'WHERE fecha_especificada
27     FinSi
28 FinSi
29
30 //Este código entrega Query en lenguaje SQL para filtrar información solicitadaal repositorio
31 //Specto.

```

Ilustración 6.1: Seudocódigo Filtro_data. Fuente: Elaboración propia.

- Recepción y almacenamiento de información:

Se encarga de recibir la información de la consulta anterior y guardarla en un repositorio alternativo histórico, lo anterior se realiza de la siguiente forma:

*Considerando una base de datos existente llamada base_de_datos_analisis con una tabla de hecho llamada Info_specto que contiene los mismos campos que la vista anterior, con un campo auto incrementable extra.

La información es almacenada como se muestra a continuación:

```

1 Funcion Recepcion_y_almacenaje_data(datos_a_extraer)
2
3 estado_actualizacion <- 'Falso'
4
5     INSERT INTO `Info_specto`
6     SELECT * from `datos_a_extraer`
7
8 //Código que actualiza base de análisis
9 estado_actualizacion <- 'Verdadero'

```

Ilustración 6.2: Seudocódigo función Recepción y almacenaje data. Fuente: Elaboración propia.

- Envío estado de actualización información:

Al realizar la solicitud de información, el sistema debe entregar al usuario un listado de información exploratoria simple, que está enfocada en entregar la cantidad de registros por columna, promedio, mínimo y máximo.

El seudocódigo que define lo anterior se muestra a continuación:

```

1 Funcion Datos_exploratorios_simple(datos_a_extraer)
2
3 //k indica la cantidad total de variables (campos)
4 //i indica la cantidad de registros
5 //j indica la variable j
6 //r indica la cantidad de registros totales
7
8 cuenta_registros <- 0
9 i <- 1
10 j <- 1
11 suma <- 0
12
13 Para j <- 1 Hasta k Hacer
14     Min[j] <- datos_a_extraer[i,j]
15     Max[j] <- datos_a_extraer[i,j]
16     Suma <- 0
17     Prom <- 0
18
19     Para i <- 1 hasta r Hacer
20         Si datos_a_extraer[i,j] != 'vacio' Y datos_a_extraer[i,j] != N/A Entonces
21             cuenta_registros <- cuenta_registros + 1
22         FinSi
23         Si Min[j] > datos_a_extraer[i,j] Entonces
24             Min[j] <- datos_a_extraer[i,j]
25         FinSi
26         Si Max[j] < datos_a_extraer[i,j] Entonces
27             Max[j] <- datos_a_extraer[i,j]
28         FinSi
29         suma <- suma + datos_a_extraer[i,j]
30     FinPara
31
32     Prom <- suma/cuenta_registros
33     Matriz_resultados[j] <- (j,cuenta_registros,Min,Max, Prom)
34 FinPara
35
36 Enviar Matriz_resultados

```

Ilustración 6.3: Seudocódigo Envío estado de actualización información. Fuente: Elaboración propia.

- Filtro de información y entrega a software estadístico:

Cuando la información es descargada desde el servidor, trae todos los campos de la vista habilitada (todos los valores de los sensores del motor

monitoreados), sin embargo, de acuerdo a los modelos matemáticos procesados, sólo un conjunto de campos son seleccionados para el análisis. Lo anterior se puede ver en el siguiente seudocódigo:

```
1 Procedimiento Entrega_info_software_estadistico(datos_a_extraer)
2
3 //Sea Campos_validos un arreglo con los campos válidos para el análisis predictivo.
4
5 datos_software_estadistico <- 'Select InList(Campos_validos) from datos_a_extraer
6
7 Entrega datos_software_estadistico
```

Ilustración 6.4: Seudocódigo Entrega información software estadístico. Fuente: Elaboración propia.

- Recepción y almacenamiento de resultados:
El sistema estadístico debe ser capaz de entregar una predicción basado en un ESN, clasificación y probabilidad de ocurrencia, sin embargo, dependiendo del *layout* de salida del software será creada la base_de_datos_resultados.
- Despliegue resultados:
Implica tomar lo que fue almacenado en base_de_datos_resultados y entregar lo que al usuario le interesa, como primeras reglas de despliegue de resultados se debe cumplir lo siguiente:
 - Mostrar ESN
 - Mostrar Fecha
 - Mostrar Hora
 - Mostrar nodo de clasificación
 - Mostrar probabilidad de ocurrencia

Por la naturaleza de los datos, el sistema en seudocódigo debe realizar lo siguiente:

```

1 Funcion Despliegue_resultados(base_de_datos_resultados)
2
3 //k es la cantidad de ESN
4
5 k <- 1
6 i <- 0
7
8 ESN_existentes <- Select DISTINCT(ESN) FROM 'base_de_datos_resultados'
9
10 Para i <- 1 hasta k Hacer
11     Si ESN_existentes[i] = base_de_datos_resultados.ESN Entonces
12         Prob[i] <- Guardar mayor probabilidad de ocurrencia
13         Fecha[i] <- Guardar fecha mayor probabilidad de ocurrencia
14         Hora[i] <- Guardar Hora mayor probabilidad de ocurrencia
15     FinSi
16 FinPara
17
18 Despliegue_resultados <- [ESN, Prob, Fecha, Hora]

```

Ilustración 6.5: Seudocódigo Despliegue de resultados. Fuente: Elaboración propia.

Cuando el sistema se ejecuta de forma automática (proactiva):

Se debe considerar:

- Como proceso automático será configurado para que sea ejecutado diariamente cada 6 horas (4 veces al día).
 - Este proceso analiza información que no ha sido procesada, para no exigir ni desperdiciar capacidad de procesamiento (analiza información incremental).
- Revisión *data* existente:

El sistema determina que información no ha sido analizada mediante un *flag* especial en la base de datos, por otro lado, también determina el rango horario de información extraída, determinando los rangos de tiempo donde existe *data*, esto evita que la información sea descargada nuevamente (duplicación de información), para lo anterior, el sistema debe establecer la hora y fecha actual para determinar el rango de tiempo a descargar:

$$\text{Intervalo de descarga} = (T_{\text{sistema}} - T_{\text{máximo}}) + \text{información no analizada}$$

Donde:

$T_{sistema}$: Hora del sistema al momento de estimar la descarga.

$T_{máximo}$: Hora máxima de información ya descargada.

información no analizada : Corresponde a la información que no tiene el *flag* activado de información analizada.

- Solicitud información nueva

Esta tarea utiliza la función *Filtro_data*, donde los parámetros a utilizar son:

- *Selección_data*: Todos los ESN.
- *Selección_fecha*: [$T_{sistema}, T_{máximo}$]: Fecha sistema

- Recepción información nueva:

Funciona de la misma forma en que trabaja Recepción y almacenamiento de la información.

- Selección muestra aleatoria:

Es generada una muestra aleatoria simple para cada uno de los ESN que tienen información y que no han sido analizados con el 30% de la información total, siendo esta información guardada en una vista simple llamada *análisis_info_automatico*.

- Envío información software estadístico

Realiza el mismo trabajo que filtro de información y entrega al software estadístico.

- Recepción de resultados

Realiza el mismo trabajo que recepción y almacenamiento resultados.

- Análisis de resultados

El sistema no despliega resultados, sin embargo, los analiza en el contexto de que esta modalidad no generará ninguna alerta si su probabilidad de ocurrencia no es superior a un porcentaje preestablecido, el cual, actualmente está definido sobre un 75%, este porcentaje es inicial para comenzar a analizar el comportamiento de esta modalidad, en caso de que esta probabilidad de ocurrencia sea generado por el modelo predictivo, el sistema enviará una alerta directamente al ingeniero de planificación.

7 Arquitectura de Sistemas

7.1 Diagrama de paquetes

En este punto, ya se encuentra definida la lógica de negocio, como el sistema deberá determinar la demanda de motores por servicio de mantenimiento.

Se debe implementar una solución que sea capaz de realizar las acciones que fueron mencionadas, actividades como ejecución de modelos predictivos, consultas de bases de datos, actualización de información, entre otras, las cuales se deben dejar plasmados para que cualquier persona con conocimientos en UML los pueda entender e interpretar. Se dejará documentada la arquitectura TI que permitirá implementar la solución, por lo que será definido el Diagrama de Paquetes, el cual agrupa de una manera simple todas las actividades del sistema como se ve en el siguiente diagrama:

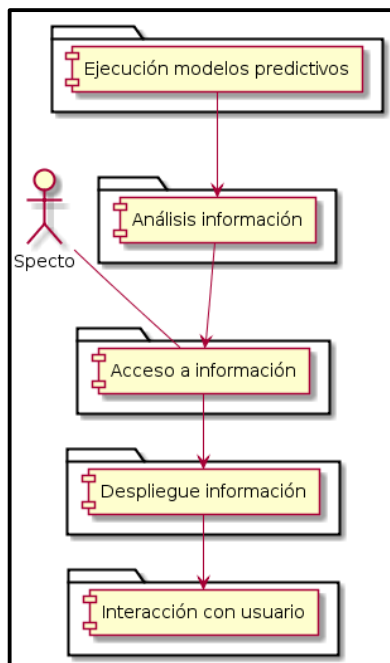


Figura 7.1: Diagrama de paquetes. Fuente: Elaboración propia.

Estos paquetes reflejan las principales actividades del sistema, donde destaca la “ejecución de modelos predictivos”, contendrá todas las clases necesarias para la integración entre la solución web y el procesamiento de datos para la ejecución del

modelo predictivo determinado, el paquete de “análisis de información” se preocupa de agrupar y designar la información que será guardada en la base de datos. “Acceso a información” es quién guarda y consulta información a las bases de datos, ya sean internas o externas, también se encuentran los paquetes de “despliegue de información”, se preocupa de todo lo correspondiente a *layout* de información en pantalla y por último el paquete de “interacción con usuario” que toma todas las necesidades de éste mediante la interacción con la interfaz gráfica.

7.2 Casos de uso

A continuación, se muestra en la Figura 7.2 el diagrama de casos de uso para la interacción del sistema y el ingeniero de planificación:

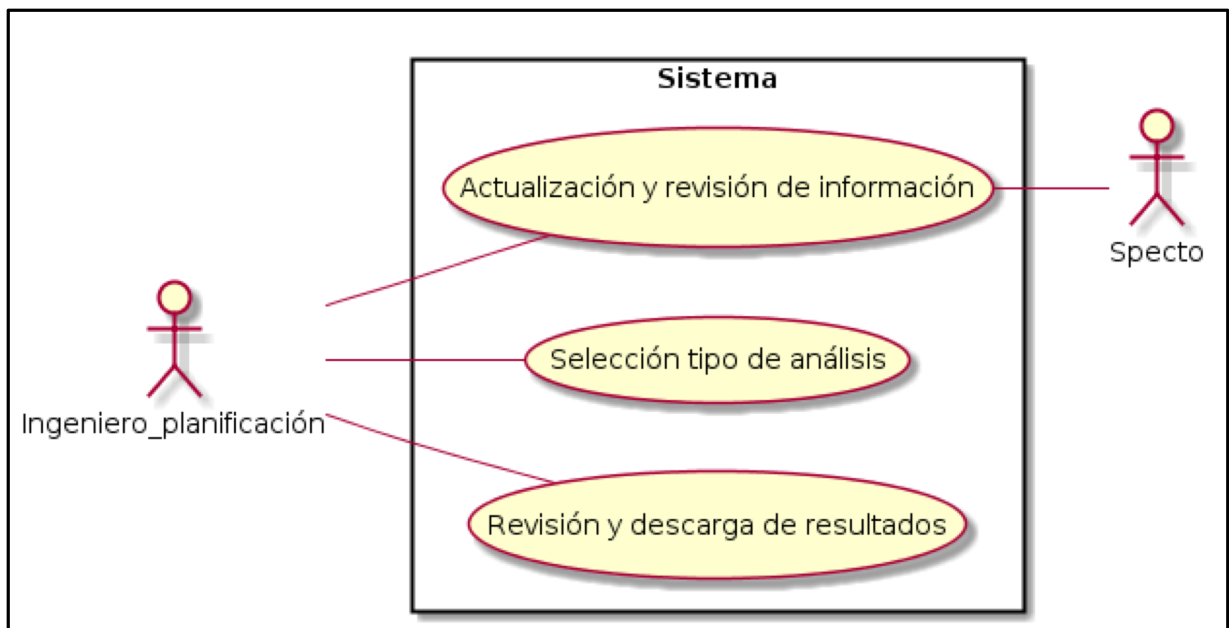


Figura 7.2: Diagrama Casos de uso: Interacción sistema e Ingeniero Planificación.
Fuente: Elaboración propia.

El primer caso de uso corresponde a la interacción del Ingeniero de planificación y el sistema, donde “Actualización y revisión de información”, este caso de uso ejecuta todos los procesos relacionados a la actualización de información que se encuentran en las bases de datos para analizar con los modelos predictivos, es importante ya que

en caso que el usuario no desee actualizar, el software es detenido, ya que sin información nueva, no será necesario utilizar recursos del sistema en ejecutar el análisis, además, entrega al usuario un resumen de la información nueva que se agregará, donde el usuario debe indicar la aceptación de la información actualizada de acuerdo a estadística descriptiva simple que es entregada por el sistema para que el usuario determine si será utilizada esta información o no en la ejecución del análisis.

“Selección tipo de análisis”, es donde el usuario tiene la facultad de seleccionar el pronóstico de todos los equipos de los cuales el sistema recogió información o puede filtrar de los que el sistema ofrezca (que corresponderá a la información de los motores que tienen información al día de la actualización de la *data*).

“Revisión y descarga de resultados”, permite al usuario aceptar los resultados entregados por el sistema de acuerdo a la opción entregada anteriormente, permitiendo dar una mirada a los resultados y solicitar reprocesar en caso de que sea necesario. Además, permite al usuario descargar los resultados a su equipo computacional para que lo pueda procesar como estime conveniente.

El segundo caso de uso corresponde al control que realiza el sistema sobre la planificación y la operación, el cual permitirá generar alarmas de acuerdo a parámetros establecidos. A continuación se muestra el caso de uso:

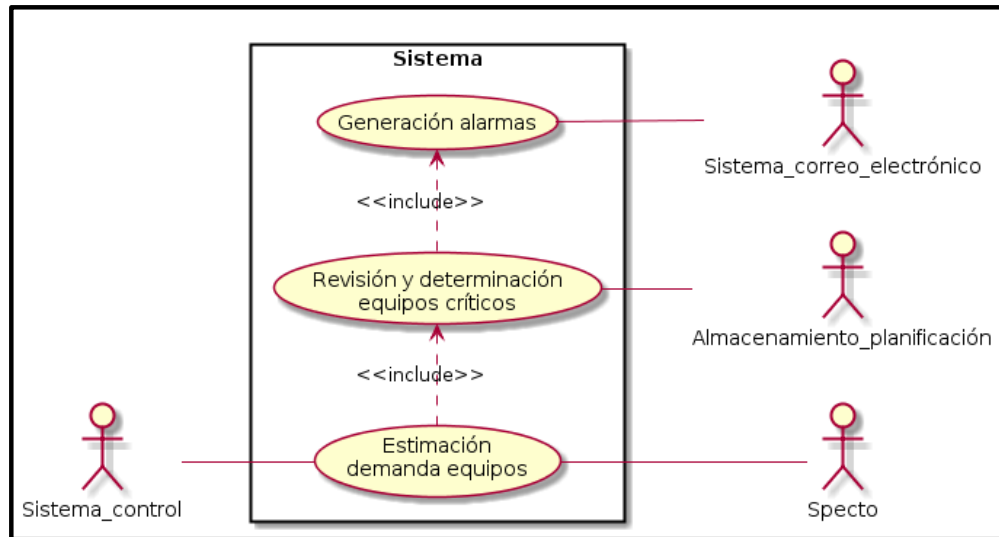


Figura 7.3: Diagrama Casos de uso: Sistema de Control. Elaboración propia.

Existen los casos de uso asociados al Sistema de Control, este sistema será encargado de entregar alarmas de acuerdo a parámetros que serán establecidos en el desarrollo de la tesis, basándose en el estado de las variables del equipo.

El primer caso de uso es “Estimación demanda de equipos”, esta tarea es ejecutada de forma automática por el sistema cada 6 horas, esto permite una frecuencia de 4 veces al día, siendo una cantidad menor de horas a las que corresponden a un turno, lo anterior busca la proactividad del sistema, se debe evitar el uso indiscriminado del sistema para que no sature los servidores de base de datos en productivo con solicitudes de información recurrentes.

Para determinar los equipos candidatos, estos son establecidos de acuerdo a la probabilidad de ocurrencia, que por el momento se encuentra como mínimo el 75% (este parámetro se irá ajustando a medida que se conozca de mayor forma el comportamiento de la flota basando el análisis en la información generada por el modelo predictivo), el sistema debe ser capaz de determinar si estos equipos ya fueron programados para recibir mantenimiento (en el caso de que el usuario a través de la ejecución del proceso de mantenimiento predictivo determine el listado de equipos a reparar), para discriminar entre los equipos que debe emitir alerta y los que no,

además, se encarga de definir cuáles serán las alarmas (y por ende los equipos) a emitir, entregando la información al usuario, este caso de uso lleva por nombre “Revisión y determinación equipos críticos”. Por último, existe el caso de uso “Generación de Alarmas”, que se encarga de enviar un informe al sistema de correo electrónico para que sean informadas las personas responsables.

7.3 Diagramas de sistemas y de clases

Serán desarrollados el diagrama de secuencia de sistema para cada uno de los casos de uso, donde el sistema se utiliza como una caja negra, luego, estos diagramas se pueden aperturar en diagramas de sistemas extendidos, donde se muestra la secuencia de clases del sistema, y por último el diagrama de clases que muestra la relación entre ellas.

7.3.1 Diagrama de secuencia de sistema

7.3.1.1 Casos de uso: Interacción sistema e Ingeniero Planificación

7.3.1.1.1 Actualización y revisión de información

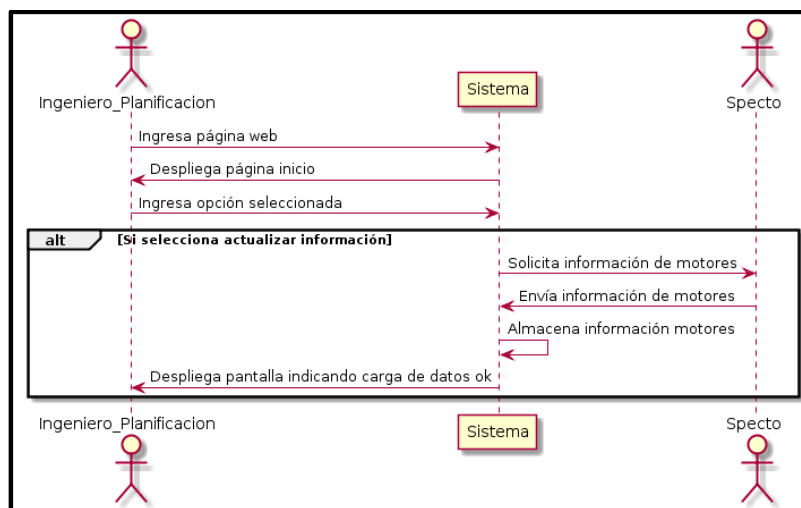


Figura 7.4: Diagrama de secuencia de sistema: Actualización y revisión de información.
Fuente: Elaboración propia.

7.3.1.1.2 Selección tipo de análisis

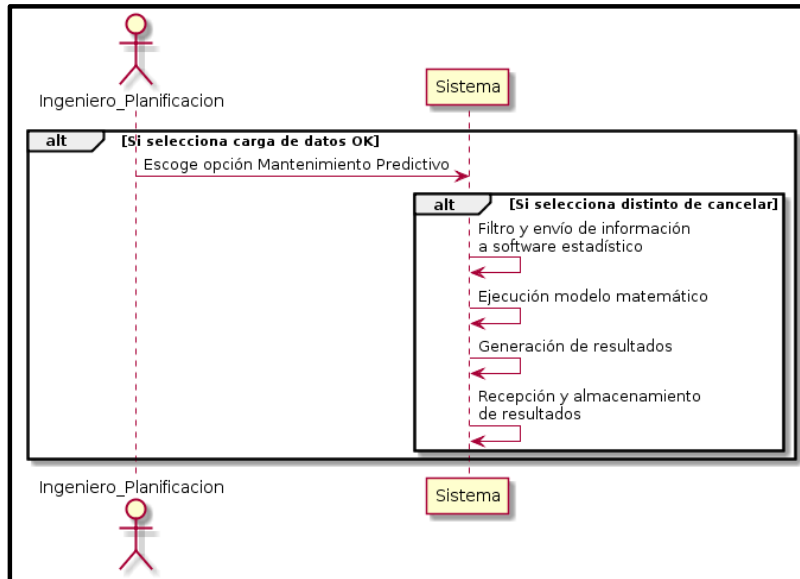


Figura 7.5: Diagrama de secuencia de sistema: Selección tipo de análisis. Fuente: Elaboración propia.

7.3.1.1.3 Revisión y descarga de resultados

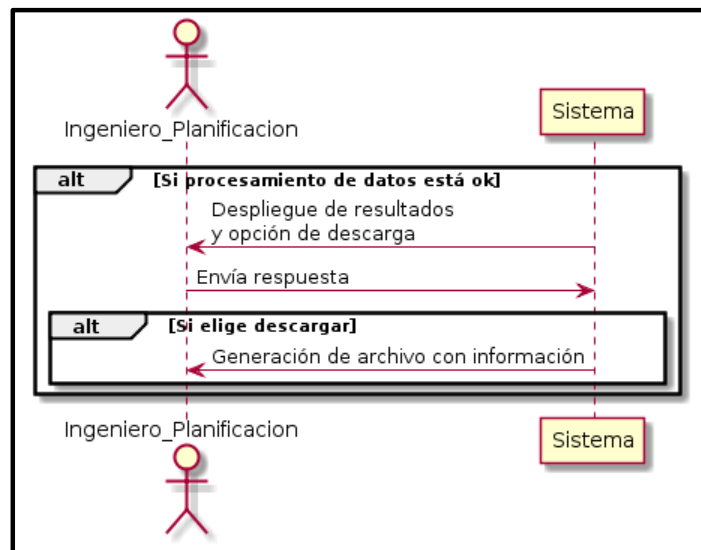


Figura 7.6: Diagrama de secuencia de sistema: Revisión y descarga de resultados. Fuente: Elaboración propia.

Estos diagramas muestran cómo funciona cada caso de uso donde la interacción con el sistema y el usuario es lo principal.

Para mayor información, puede consultar el anexo 14.14 Narrativa casos de uso: Interacción sistema e Ingeniero Planificación.

A continuación, se muestra un ejemplo de interfaz que podría ser utilizado para la implementación de esta aplicación, para lo cual iré utilizando los flujos dentro del diagrama de sistema:

a. Ingresa página web:

El acceso se hará desde la siguiente dirección, el cual corresponde a una ubicación de intranet, la cual el usuario debe ingresar en el navegador de su computador:

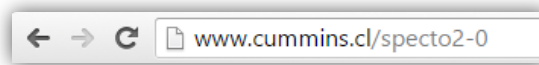


Figura 7.7: Dirección web. Fuente: Elaboración propia.

b. Despliega página inicio:

El sistema desplegará la siguiente página de inicio para el usuario:



Figura 7.8: Inicio Spectro 2.0. Fuente: Elaboración propia.

Donde el usuario iniciará el sistema haciendo click en la flecha. Al iniciar el sistema, aparecerán las siguientes opciones:



Figura 7.9: Actualización base de datos. Fuente: Elaboración propia.

Es en este paso donde el analista puede decidir el tipo de información necesita actualizar para procesar.

- c. Despliega pantalla indicando carga de datos ok:

Cuando el usuario ha seleccionado una opción de las que se presentan en el paso anterior, el sistema arrojará un resumen como el que se muestra a continuación:



Figura 7.10: Actualización base de datos (resumen). Fuente Elaboración propia.

El analista deberá aceptar o rechazar el resumen que se muestra, esto permite procesar información que se encuentra dentro de los parámetros que el analista está utilizando.

- d. Despliega opciones ejecución Mantenimiento Predictivo:

Cuando el analista indica que la información se encuentra OK, el sistema procede a desplegar las siguientes opciones:



Figura 7.11: Selección algoritmo data Mining. Fuente: Elaboración propia.

e. Despliegue de resultados y opción de descarga:

Al seleccionar una de las opciones de procesamiento de datos, se entrega un resumen en pantalla:

Faena	ESN	Probabilidad falla	Fecha	Nodo (si corresponde)
Radomiro Tomic	318273645	80%	14-05-2015	N/A
Collahuasi	318745343	70%	14-05-2015	N/A
Los Pelambres	313484905	60%	14-05-2015	N/A
El Soldado	317463423	50%	14-05-2015	N/A
Quebrada Blanca	746353428	40%	14-05-2015	N/A
Andina	376352423	30%	14-05-2015	N/A
Cerro Colorado	948373625	20%	14-05-2015	N/A
Zaldivar	746352436	10%	14-05-2015	N/A

Figura 7.12: Resumen Resultados. Fuente: Elaboración propia.

Donde se muestran resultados relevantes, sin embargo, el sistema permite descargar en formato Excel la predicción de toda la información.

f. Despliegue nuevo análisis:

Por último el sistema permite volver a procesar la información:



Figura 7.13: Procesar más información. Fuente: Elaboración propia.

7.3.1.2 Casos de uso: Sistema de control

7.3.1.2.1 Estimación demanda equipos

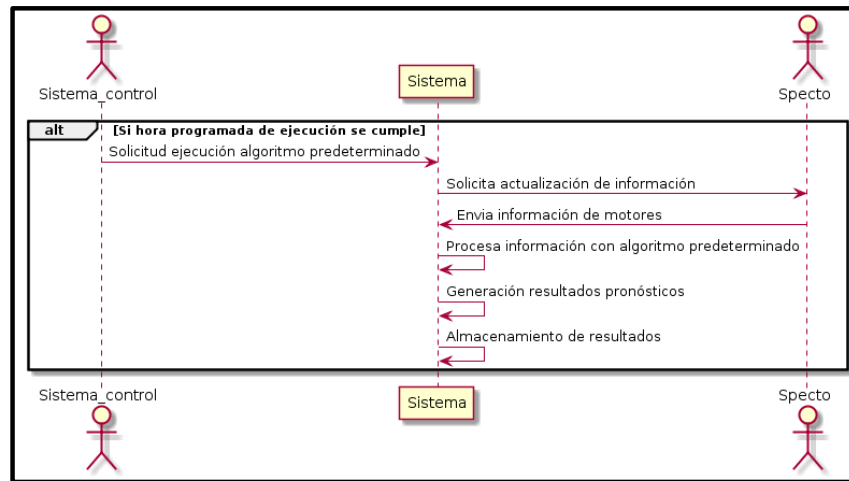


Figura 7.14: Diagrama de secuencia de sistema: Estimación demanda equipos. Fuente: Elaboración propia.

7.3.1.2.2 Revisión y determinación equipos críticos

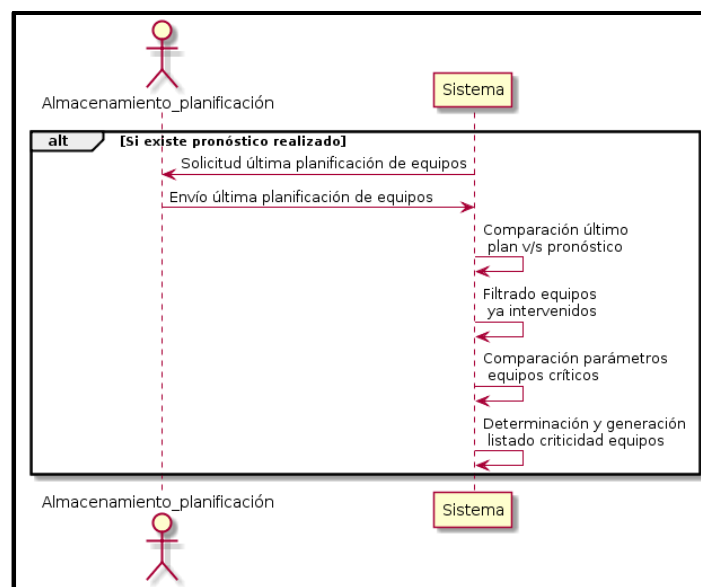


Figura 7.15: Diagrama de secuencia de sistema: Revisión y determinación de equipos críticos. Fuente: Elaboración propia.

7.3.1.2.3 Generación alarmas

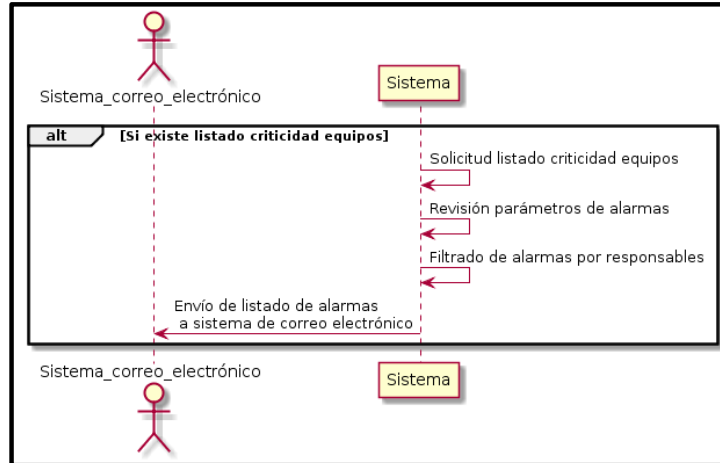


Figura 7.16: Diagrama de secuencia de sistema: Generación de alarmas. Fuente: Elaboración propia

Para mayor información puede consultar el anexo 14.15 Narrativa casos de uso: Sistema de control.

7.3.2 Diagrama de secuencia de sistema extendido

7.3.2.1 Casos de uso: Interacción sistema e Ingeniero Planificación

7.3.2.1.1 Actualización y revisión de información

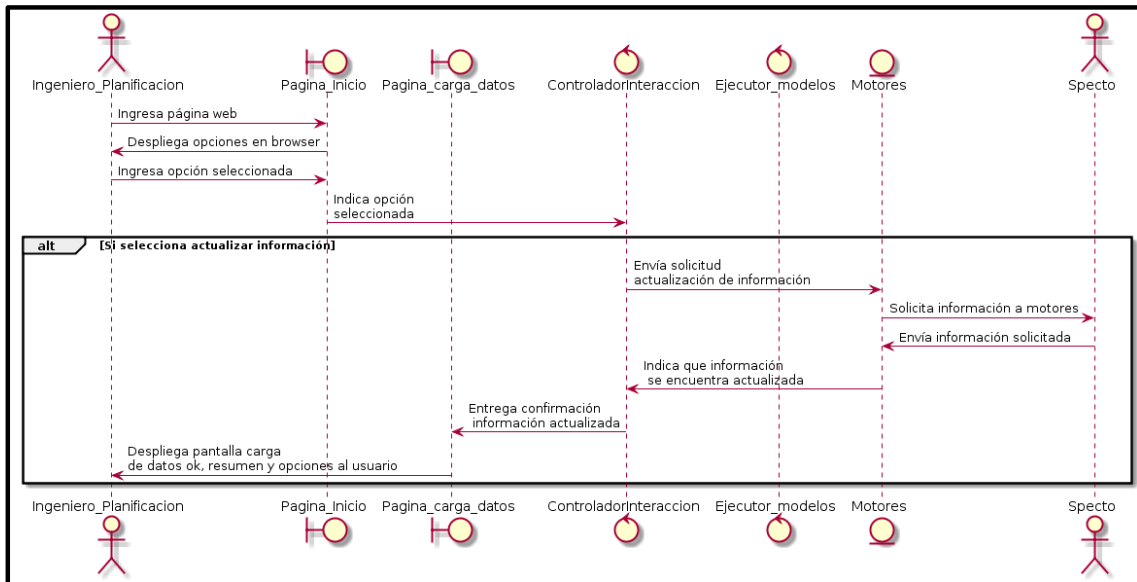


Figura 7.17: Diagrama de secuencia extendido: Actualización y revisión de información. Fuente: Elaboración propia.

7.3.2.1.2 Selección tipo de análisis

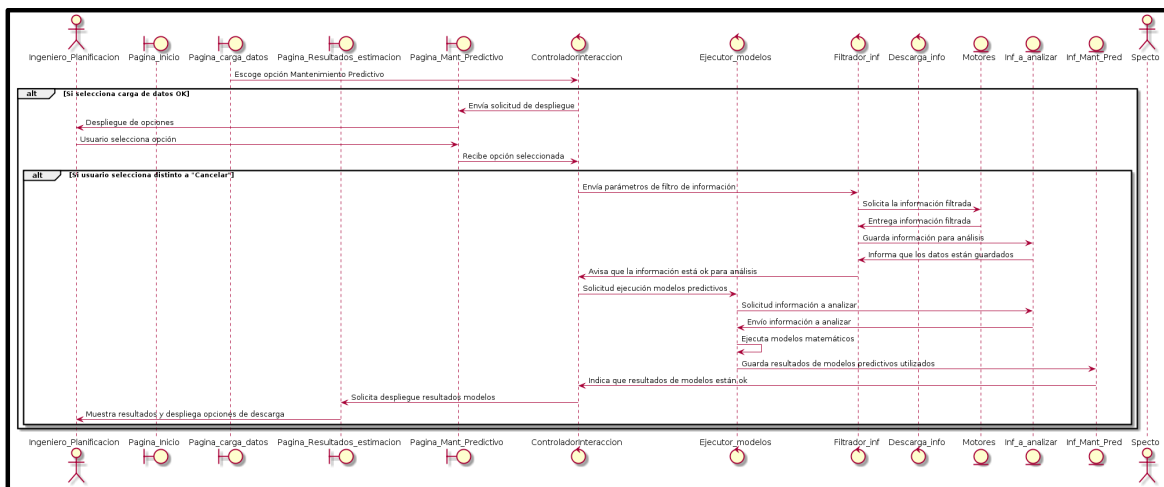


Figura 7.18: Diagrama de secuencia extendido: Selección tipo de análisis. Fuente: Elaboración propia.

7.3.2.1.3 Revisión y descarga de resultados

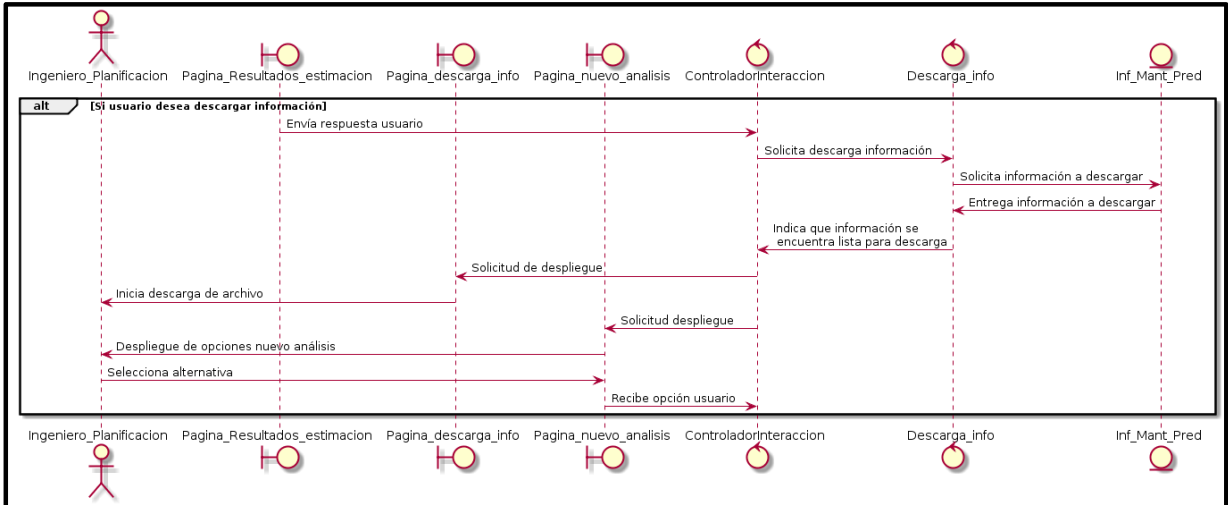


Figura 7.19: Diagrama de secuencia extendido: Revisión y descarga de resultados. Fuente: Elaboración propia.

7.3.2.2 Casos de uso: Sistema de control

7.3.2.2.1 Estimación demanda equipos

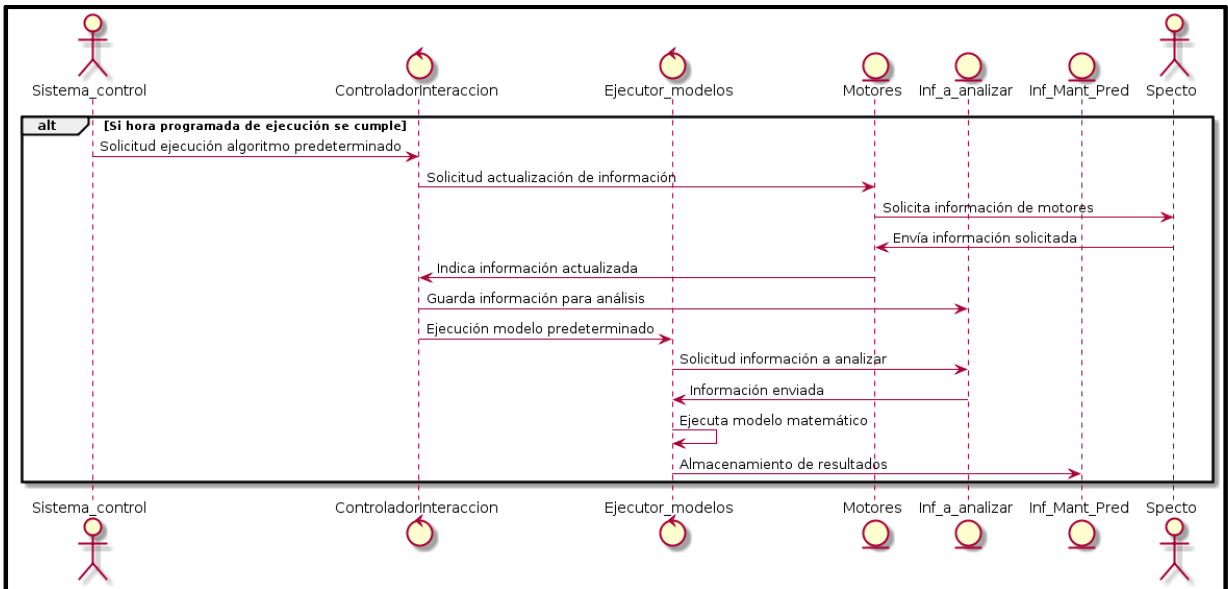


Figura 7.20: Diagrama de secuencia extendido: Estimación demanda equipos. Fuente: Elaboración propia.

7.3.2.2 Revisión y determinación equipos críticos

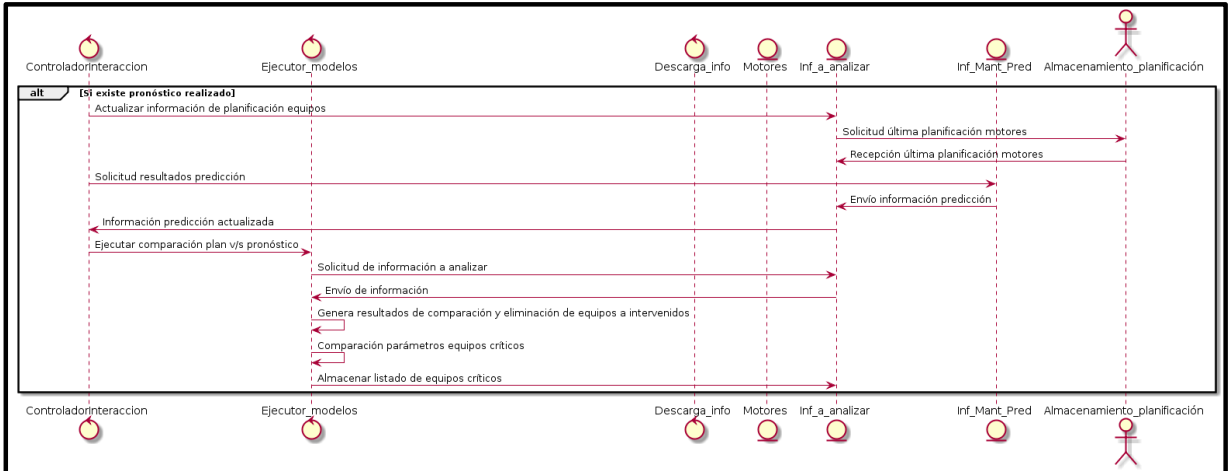


Figura 7.21: Diagrama de secuencia extendido: Revisión y determinación equipos críticos. Fuente: Elaboración propia.

7.3.2.2.3 Generación alarmas

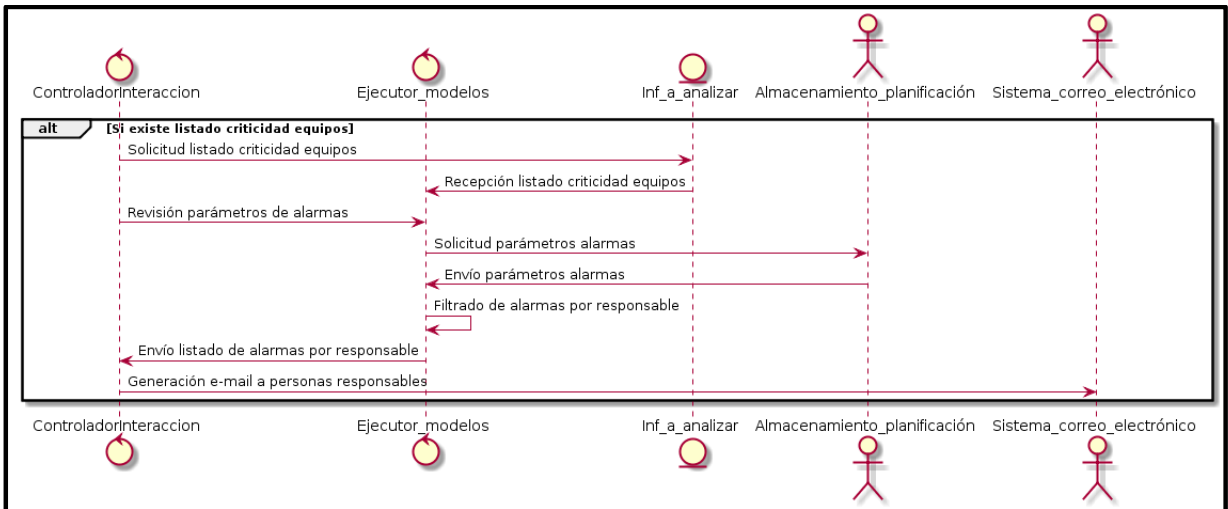


Figura 7.22: Diagrama de secuencia extendido: Generación de alarmas. Fuente: Elaboración propia

Se muestran las clases propuestas que hacen posible las acciones descritas. Destacan las clases de control “ControladorInteraccion”, quién gestiona y coordina todas las actividades y por ende a las clases, “Ejecutor_modelos” se concentra en las

lógicas de ejecución para el modelo predictivo, “Filtrador_inf” quién se encarga de filtrar la información proveniente del sistema Specto, de acuerdo a las necesidades del usuario, por último, “Descarga_info” que es el encargado de preparar la información cuando el usuario desea descargar la información a su computador.

7.4 Diagrama de clases

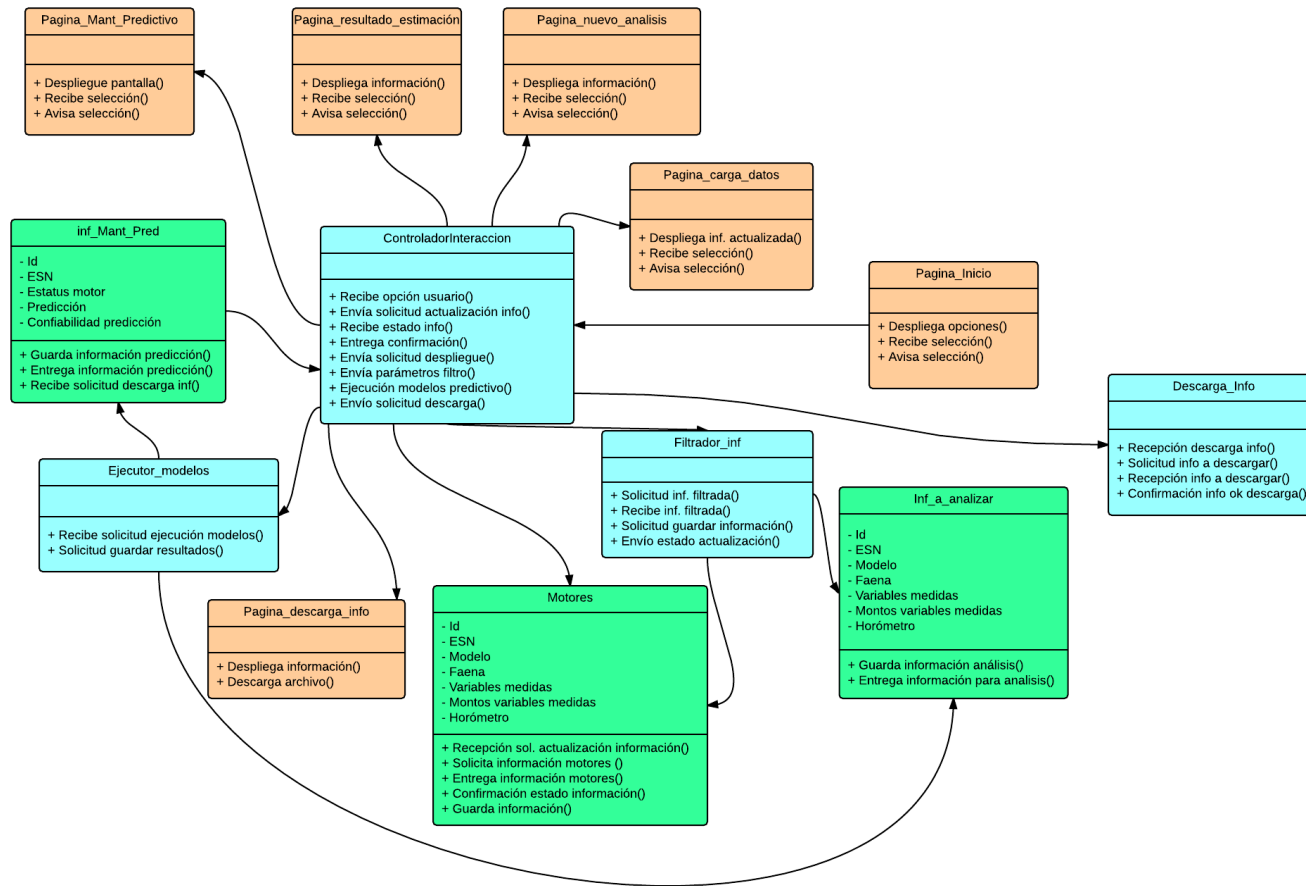


Figura 7.23: Diagrama de clases. Elaboración propia.

7.4.1 Base de datos

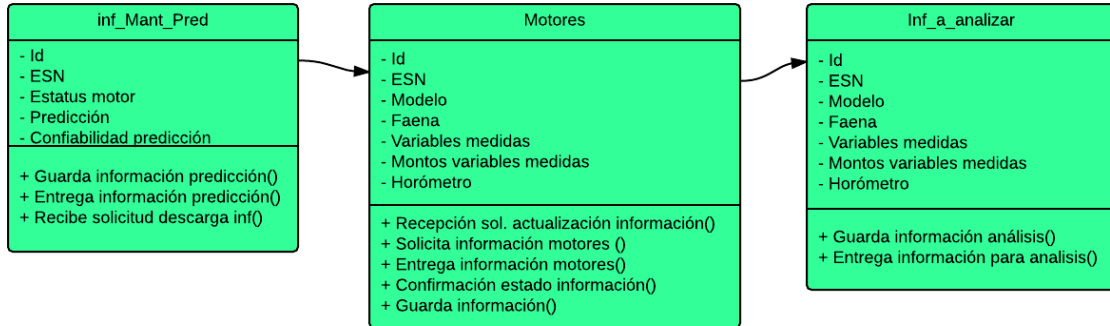


Figura 7.24: Relación Entity del sistema. Elaboración propia.

Como consecuencia de los diagramas de clases, las relaciones que existen entre las clases y los diagramas de sistemas, se genera este diagrama de base de datos, que solo considera las entidades creadas, por lo que será este el inicio de la creación de la base de datos física.

7.4.2 Diagrama de despliegue

El diagrama de despliegue es clave para indicar al lector cual será el ambiente de desarrollo en el cual se encontrará la aplicación, este diagrama indica el software, los medios de comunicación, las especificaciones más relevantes para la implementación de la solución diseñada. Para la solución completa, mediante la información recopilada de lo que deberá ser implementado, se encuentra en el siguiente diagrama:

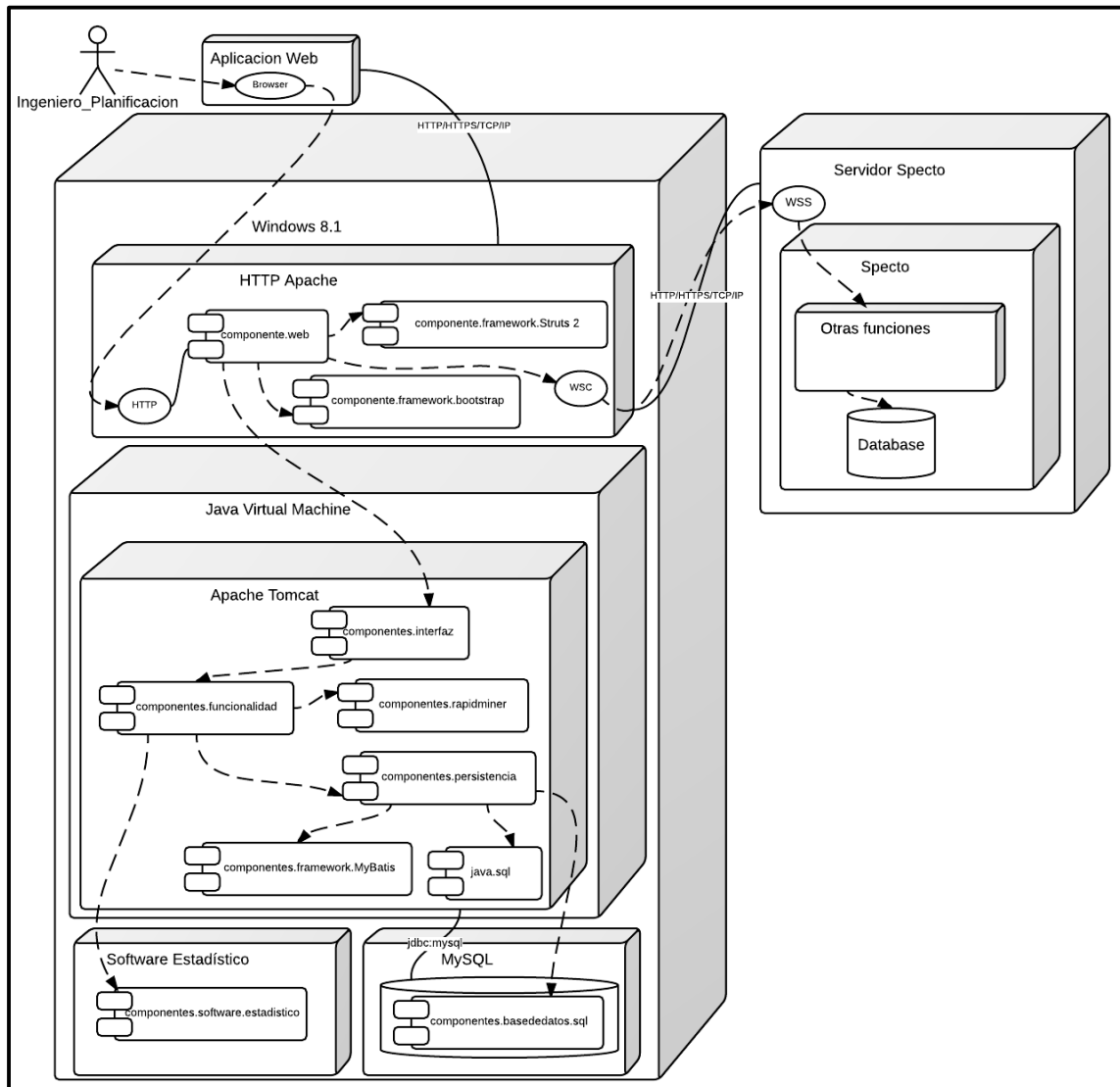


Figura 7.25: Diagrama de despliegue. Fuente: Elaboración propia

Se puede ver que el servidor de la aplicación web funcionará bajo Windows, el cual tendrá como motor de base de datos MySQL, por otro lado tendrá como servidor web Apache Tomcat, y todo esto ejecutado por Java, por otra parte, el servidor deberá contar con el software estadístico para ejecutar los procesos que sean establecidos. El usuario se conectará con la aplicación vía web a través de su browser. Por otro lado,

la aplicación se conectará a bases de datos externos a través de Web Services, los cuales sólo podrán consultar información.

8 Gestión del cambio

Los proyectos ejecutados en las organizaciones, luego de ser diseñados y planificados en papel, deben ser materializados con las personas que participan en el proyecto, donde es muy importante contar con el involucramiento y convencimiento de todos los actores importantes para que sea una implementación exitosa.

En este capítulo se desarrollarán conceptos importantes para alcanzar el éxito en la implementación de este proyecto, que deben ser tomados en cuenta para evitar que sólo se convierta otra iniciativa incompleta sobre el escritorio de alguien con la imagen de que éste no fue ejecutado completamente debido a que el proyecto no tenía futuro o por una mala planificación.

8.1 Contexto de la empresa

Distribuidora Cummins Chile S.A. es una empresa proveedora de servicios para la alta minería. Esta industria se ha visto muy afectada por el actual escenario económico, afectando principalmente a países como Chile, cuya economía está basada en la comercialización de cobre. Lo anterior genera que las empresas tomen medidas necesarias para ser eficientes en sus procesos productivos, revisando procesos y procedimientos, reestructurando áreas, distribuyendo tareas, buscando economías de escala que les permita operar de la forma más eficiente posible, donde el primer análisis que se genera es medir la capacidad laboral actual versus la necesidad de servicios que se busca cubrir, convirtiéndose la disminución de personal en una de las primeras medidas a tomar por las altas jerarquías del holding.

De acuerdo al escenario expuesto, se genera un ambiente de incertidumbre dentro de la organización a tal punto que aparecen ciclos viciosos que afectan directamente al estado emocional de las personas y por ende su predisposición a cooperar en nuevas iniciativas, ya que el principal foco del trabajador será realizar bien su trabajo regular para no ser desvinculado (no tomar riesgos), ya que en el caso de ser evaluado, su mayor evaluación será con respecto a sus logros en las tareas ya asignadas (y

conocidas), además, la oferta laboral en el país no es de las mejores, por lo que es un tema muy complejo y sensible de tratar.

Por otro lado, existe la necesidad de disminuir costos y gastos a nivel transversal en el holding, holding donde se encuentra inmersa Distribuidora Cummins Chile, por lo que se están realizando los mayores esfuerzos para disminuir todo lo que requiera flujo de caja, esto, como medida de prevención ante un cambio drástico en el ambiente tributario del país, junto con disminuir las inversiones que no sean estrictamente necesarias, todo lo anterior se está realizando a nivel holding, debido a directrices desde casa matriz que se encuentra en Japón.

8.2 Creación de sentido y narrativas

Para lograr que la implementación del proyecto sea exitoso, es necesario que todos los actores fundamentales se encuentren involucrados en el proyecto, ya que de otra manera pueden generar escenarios en los cuales no se sientan partícipes y por lo mismo, se conviertan en obstáculos muy difíciles de sortear en procesos posteriores.

A continuación se muestra un esquema de las personas que se encuentran considerados dentro de la estructura del proyecto:

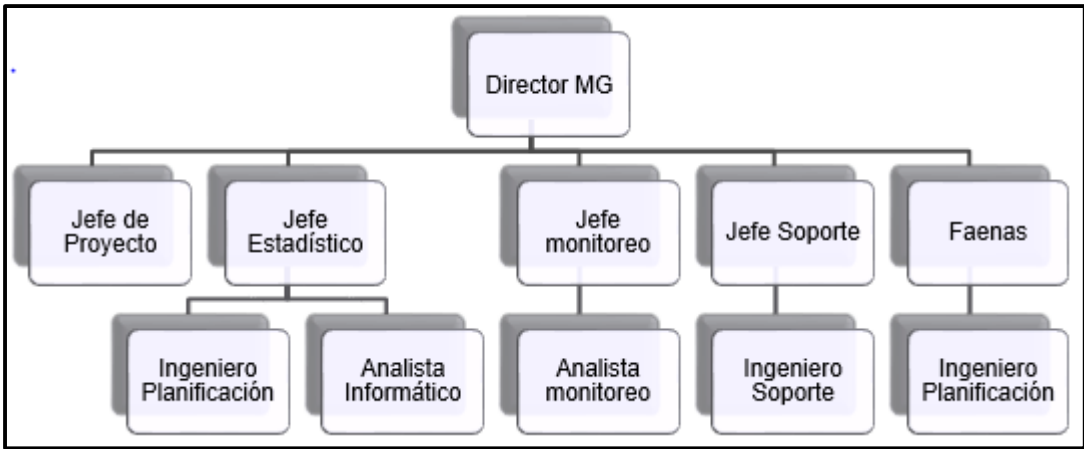


Figura 8.1: Actores relevantes implementación proyecto. Fuente: Elaboración propia.

En la Figura 8.1, dentro del proyecto se encuentran involucrados todos los actores que participarán en su implementación, pero que además, continuarán utilizando el sistema en el tiempo. Existe un gran *sponsor* del proyecto, ya que es necesario para que la predisposición a participar y aportar al proyecto se vea inclinada a favor desde el inicio.

El éxito o fracaso de implementación será definido en mayor medida si los participantes se sienten parte o no del proyecto, a pesar de existir el Director MG (*Mining Group*) como principal *sponsor*, es necesario involucrar a la gente mediante la incorporación de un sentido común, que sea grupal, para convencer a los participantes que es lógico que se encuentren allí participando ya que será provechoso en el futuro, y no que es una imposición de su jefe directo. Para cumplir con lo anterior, se deben generar las narrativas para cada uno de los participantes, que se muestran a continuación:

- Director MG (*Mining Group*): “El proyecto logrará disminuir las fallas imprevistas, apuntando a evitar las fallas catastróficas (*overhaul*), esto mejorará la percepción de los clientes con respecto a los motores Cummins y a la confiabilidad del servicio que es entregado, un servicio con características claramente diferenciadas que serán muy valoradas por el cliente, lo cual, además, beneficiará directamente en la generación de nuevos contratos de servicio en faenas donde la competencia aún se encuentra presente, lo anterior permitirá capturar mayores ingresos y mejorar el retorno como consecuencia directa del proyecto”.
- Jefe estadístico: “El sistema que será implementado será capaz de recoger y analizar información en línea, permitiendo generar y corroborar la información que actualmente es obtenida de forma manual, integrando sistemas a nivel nacional, en todas las faenas, este será el primer paso para que todos los sistemas existentes converjan en una plataforma única, donde es necesario el conocimiento de la gente que realmente conoce el negocio y la operación de las faenas.”

- Jefe de proyecto: “Este proyecto es innovador dentro de la empresa, el cual involucra mucha coordinación entre las áreas técnicas y operacionales, por lo mismo es un proyecto de alto impacto en la empresa y de alta visibilidad en el holding”.
- Jefe de monitoreo: “Este proyecto entregará una herramienta que funcionará de manera proactiva, junto con recibir solicitudes de usuario, de forma paralela a sus actuales procesos, por lo que busca ser de ayuda al momento que realicen sus tareas periódicas, es por lo anterior, que es sumamente necesaria su ayuda ya que existen aspectos técnicos que un modelo matemático jamás podrá explicar, por lo que la ejecución de este proyecto no puede avanzar de forma correcta sin su ayuda, para establecer y afinar las alarmas que realmente sean de interés para sus clientes.
- Jefe de soporte: “Este proyecto nace directamente como una herramienta directa para esta área en particular, el objetivo final es identificar información oculta que se encuentra en la información del sistema, es por esto que el desarrollo del modelo matemático debe ser hecho muy de cerca con esta área, ya que una mala interpretación de información puede generar un proyecto (y un malgasto de dinero) que no entregue los resultados requeridos, y que no sirva a ninguna de las áreas involucradas”.
- Ingeniero de planificación: “Es importante la incorporación del uso de esta herramienta dentro de sus procesos de trabajo, debido a que ayudará a generar mejores planes de mantenimiento, es por esto, que su participación es muy importante ya que es necesario definir los requerimientos de los usuarios finales que utilizarán la herramienta, de esta forma se convierte en un apoyo para quienes deban determinar el plan de mantenimiento de las faenas”.

- Analista monitoreo: “El proyecto entregará información de forma proactiva y en tiempo real (o de acuerdo a la frecuencia que sea determinada) del estatus del motor, lo cual mejora en mayor medida los diagnósticos que serán entregados por el área, logrando que sus acciones tengan efectos favorables en las áreas/faenas a las que se encuentren asesorando/informando”.
- Ingeniero de soporte: “Este proyecto dentro de sus otros entregables, permitirá analizar la historia de los equipos, de acuerdo a la necesidad de cada caso, se podrá generar información histórica capaz de ser utilizada como input de un proyecto que preceda al actual, generando conocimiento valioso para el negocio, permitiendo asesorar de mejor forma al negocio y sus clientes”.

8.3 Factores críticos de éxito

Los factores críticos de éxito relacionados a la gestión del cambio, para efectos de implementación del proyecto, generalmente están asociados a la disponibilidad, predisposición, ánimo, entre muchos otros aspectos que afectan a las personas, en especial, a las personas que se encuentran involucradas directamente en el proyecto, ya que son ellos quienes deberán ayudar a impulsar que el proyecto avance.

A continuación se detallan los factores de éxito que son críticos en esta fase:

- Involucramiento *sponsor*
Es muy importante que el *sponsor*, en este caso, el Director del *Mining Group* se encuentre alineado y motivado con el proyecto, ya que él será el principal actor para las demás áreas, será el encargado del financiamiento, por lo que si este riesgo no se encuentra tratado, el proyecto completo puede fallar.

- Apoyo mandos medios y analistas
Si bien el involucramiento del *sponsor* es importante, para el desarrollo del proyecto es indispensable el apoyo de todos los demás involucrados, son ellos quienes tienen los conocimientos técnicos, los requerimientos necesarios que deben ser cubiertos, junto con los “dolores” que pueden tener y que el proyecto es capaz de cubrir, por lo que deben ser motivados y sentirse parte del proyecto como un equipo.
- Creación de coalición y autoridad
Para que el proyecto avance en el tiempo aunque los actores cambien, debe ser generado un ambiente de coalición entre los participantes de forma que cuando llega una persona nueva los demás hagan sentir que ya existen lineamientos establecidos con respecto al proyecto y que todos los actores se sientan empoderados de estos lineamientos y objetivos, de esta forma, el traspaso será hecho de forma natural, logrando una sensación de pertenencia para cada uno de los involucrados.

8.4 Aspectos a conservar

Este proyecto no afecta directamente la forma de trabajar que utiliza el *Mining Group*, solamente añade algunos pasos extra a lo que existe, sin embargo, por el carácter de la herramienta a desarrollar, se requiere que la gente con la que se inicia este proyecto sea la que continúa durante un periodo de puesta en marcha, para realizar los ajustes necesarios y evitar que el objetivo y/o alcance sea modificado.

En resumen los aspectos que deben ser conservados son:

- Personal existente en los cargos actuales son considerados idóneos y claves en la culminación del proyecto.

- Procesos de planificación de mantenimiento actuales (Preventivo y Correctivo) son base para establecer el proyecto.
- Se debe mantener el clima de confianza existente en la forma de trabajar en cada área pese a que el sistema arroje resultados distintos a los que se utilizan como parámetros, esto, ya que se requerirá de toda la capacidad crítica y analítica actual de cada área para interpretar los resultados.
- Mantener las mejores prácticas existentes en el *Mining Group*.

8.5 Gestión de poder

Todo proyecto en vías de diseño, desarrollo e implementación que se encuentre en una organización, sea cual sea su tamaño, pero que involucre a más de un área, requiere de ejes que puedan gestionar el poder que poseen de forma intrínseca de manera efectiva y natural. Para lograr esto, es importante diseñar un plan que contenga las responsabilidades de cada uno de forma clara, de tal forma que todos las conozcan y puedan gestionar sus tareas y responsabilidades de forma clara, sencilla y acorde a las necesidades del proyecto.

A continuación se especifica los distintos poderes distribuidos según los actores relevantes:

- *Director Mining Group*: Encargado de entregar el financiamiento relacionado directamente con el proyecto, además, de supervisar (en menor grado) los estados de avance y tomar decisiones cuando entre las áreas no sea posible.
- *Jefe de Proyecto*: Encargado de entregar lineamientos definidos para que la implementación y desarrollo del proyecto sea lo más fluida posible, estará encargado de supervisar los estados de avance (alto grado), coordinar áreas y tareas necesarias para la ejecución exitosa del proyecto.

- Jefe de monitoreo: Encargado de entregar los parámetros necesarios de configuración y ajuste del sistema a medida que se requiera para la obtención de resultados, además, debe generar requerimientos explícitos de uso de la herramienta.
- Jefe de soporte: Encargado de interpretar información entregada por el modelo, definir si esta información es útil o no de acuerdo al conocimiento de los equipos, solicitar ajustes al modelo matemático en caso de ser necesario.
- Ingeniero de planificación: Encargado de generar requerimientos de uso para el sistema, estableciendo opciones requeridas junto con los resultados esperados.
- Analista informático: Encargado de generar documentación relevante con respecto al sistema informático generado para la implementación del proyecto, además, es el encargado de recibir los requerimientos de las distintas áreas y reunirse con los proveedores.
- Analista Monitoreo: Es el encargado de analizar los resultados y la usabilidad del sistema, establece modificaciones y mejoras al sistema para que se encuentren en línea con los procesos actuales del sistema.
- Ingeniero soporte: Encargado de revisar la calidad de la información de salida y comparar con los resultados actuales, esto permite mejorar las bases de conocimiento existentes y por otro lado permite medir la efectividad del modelo matemático.

9 Prueba de concepto

Son presentados los resultados obtenidos al aplicar distintos modelos de clasificación sobre los datos de prueba, buscando demostrar los resultados y la metodología en que se debe desarrollar el modelo matemático. Los datos de prueba corresponden a 105.419 registros que son obtenidos a través de los sensores de motores en faena Collahuasi, desde su sistema Specto, estos registros se encuentran clasificados en 2 categorías: Equipo con falla (1) y Equipo sin falla (0).

Los análisis y muestras de datos de prueba se encuentran basados en la información que el área de estadística del *Mining Group* tiene clasificado por sistemas relevantes dentro de los motores, de esta forma tienen clasificadas las fallas en todas las faenas en las que Cummins se encuentra presente.

La información utilizada para la base de entrenamiento, como para la de prueba corresponde a las fechas entre el 1 de Septiembre al 31 de Octubre 2015, descargado directamente desde el sistema Specto.

Para trabajar la información, se montó un servidor de base de datos MySQL donde fueron desarrollados las consultas para analizar la información de forma eficiente.

9.1 Consideraciones particulares de la evaluación

9.1.1 Faenas objetivos

Se establecen las faenas que estarán dentro del universo de estudio para esta tesis serán aquellas que cuenten con un contrato MARC (*maintenance and repair contract*), quedando excluidos los contratos LPP (*labor plus parts*) y asistencias técnicas (AT).

9.1.2 Faena a estudiar

En el contexto de la prueba de concepto que será desarrollado en esta tesis, se revisa la información existente relacionada a los tiempos de fallas en todas las faenas para determinar la flota a estudiar.

Los principales factores que deben ser tomados en cuenta son tiempo de falla y frecuencia, ya que son fundamentales para obtener la información requerida a analizar.

A continuación se muestran las faenas ordenadas por frecuencia de falla y tiempo acumulado:

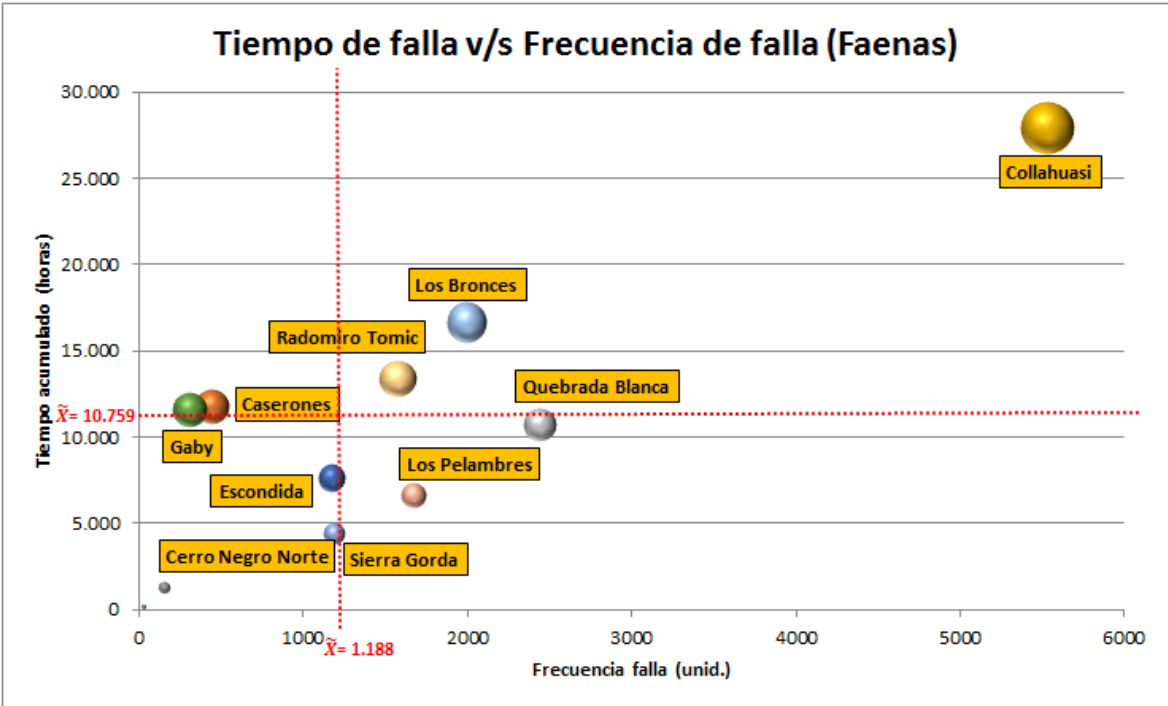


Gráfico 9.1: Tiempo vs Frecuencia de falla por faena. Fuente: Elaboración propia.

El Gráfico 9.1 presenta en el eje X la frecuencia de fallas por faena, en el eje Y las horas acumuladas, el tamaño de la burbuja representa la cantidad de horas acumuladas por faena, que permite revisar de manera visual la faena que contiene mayor información, esto es importante debido a que a mayor cantidad de horas de falla, mayor es la cantidad de información que podrá ser utilizada como target para los análisis posteriores.

De acuerdo al análisis de la información, faena Collahuasi cuenta con mayor cantidad de horas de falla acumulada, por lo que es en esta faena donde se desarrollará esta tesis.

Es importante mencionar que esta faena posee la mayor flota de Cummins Chile, y es por esto que se muestra como un punto atípico en el comportamiento de los parámetros medidos.

9.1.3 Sistema a estudiar

Las fallas son clasificadas en varios sistemas, los cuales se muestran a continuación:

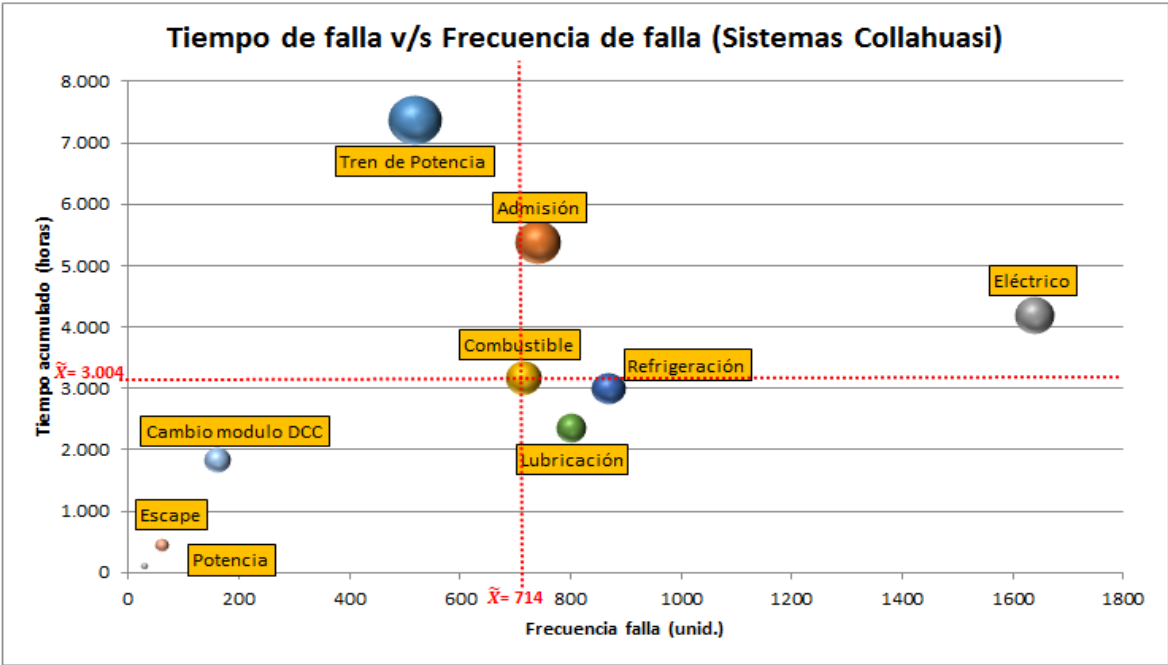


Gráfico 9.2: Tiempo vs Frecuencia de falla por sistema. Fuente: Elaboración propia.

Del Gráfico 9.2 se puede ver que el sistema con mayor frecuencia de fallas es el “eléctrico”, sin embargo, el sistema que aporta con la mayor cantidad de tiempo acumulado es “tren de potencia”, tiene menos frecuencia de falla, sin embargo, tiene la mayor cantidad de información para el desarrollo del modelo matemático, por lo que es seleccionado para comenzar con los análisis propuestos.

9.1.4 Muestra y variables a analizar

De la revisión de información, se establece que la muestra total de información a descargar corresponde a 380 archivos en formato Excel, los cuales son identificados con las siguientes variables:

- ESN : Engine Serial Number
- Fecha inicio : corresponde al día de inicio de la falla.
- Fecha término : corresponde al día de término de la falla.
- Hora inicio : corresponde a la hora de inicio de la falla.
- Hora término : corresponde a la hora de término de la falla.
- Tiempo recalculado: corresponde al total de tiempo de la falla.
- Equipo : Modelo de equipo donde está montado el motor.
- Unidad : Unidad donde está montado el motor.
- N° serie equipo : número de serie de la unidad donde está montado el motor.

Un ejemplo de esta fuente de datos se puede ver en la figura siguiente:

Descarga información Tren de potencia										
Id	ESN	Fecha Inicio	Fecha Termino	Hr Inicio	Hr Termino	Frecuencia real	Tiempo ReCalculado	Equipo	UNIDAD	N° SERIE EQUIPO
1		10-02-2014	10-02-2014	13:50:00	18:20:00	1	4,5			
2		16-04-2014	17-04-2014	19:00:00	1:30:00	0	6,5			
3		19-04-2014	19-04-2014	10:00:00	18:30:00	0	8,5			
4		20-04-2014	20-04-2014	4:30:00	6:30:00	0	2,0			
5				11:00:00	18:30:00	0	7,5			
6			21-04-2014	21:00:00	6:00:00	0	9,0			
7		21-04-2014	21-04-2014	13:30:00	18:00:00	0	4,5			
8		22-04-2014	23-04-2014	22:00:00	1:30:00	0	3,5			
9		25-04-2014	25-04-2014	8:30:00	18:10:00	0	9,7			
10		10-10-2014	10-10-2014	14:00:00	18:30:00	1	4,5			
11			11-10-2014	20:00:00	6:30:00	0	10,5			
12		11-10-2014	11-10-2014	8:00:00	12:45:00	0	4,8			

Figura 9.1: Ejemplo planilla descarga información. Fuente: Elaboración propia.

Por otro lado, tenemos las variables que serán analizadas (dependiendo de la disponibilidad de esta información en el sistema), éstas son extraídas directamente del sistema.

Algunas de las variables que serán consideradas en el análisis inicial se muestran a continuación:

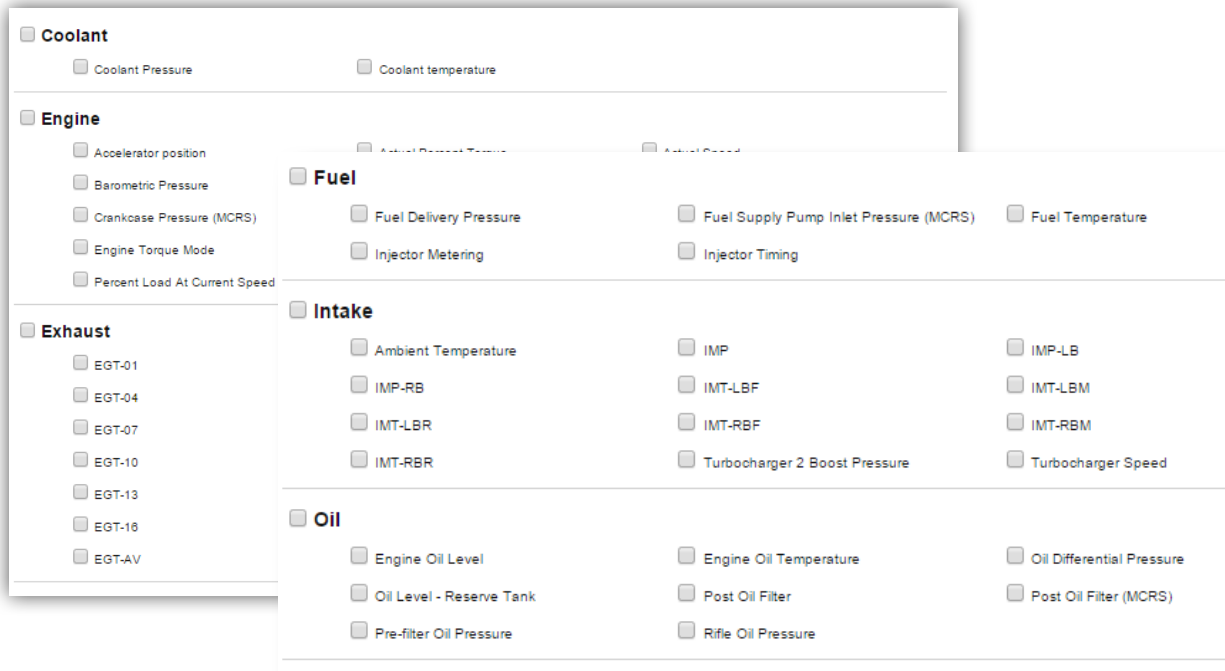


Figura 9.2: Variables sistema Spectro. Fuente: Elaboración propia.

9.1.5 Evaluación de resultados

Para evaluar los resultados que el modelo predictivo entregará, será utilizada información real a la fecha, para contrastar la realidad en comparación con el pronóstico. La *data* que será utilizada para entrenar el modelo matemático corresponde al periodo Septiembre 2015 a Octubre 2015.

Importante es mencionar que las fallas catastróficas se encuentran identificadas, por lo que también podrá establecer la causa que las produjo.

Lo mencionado, será la forma de contrastar el resultado con la realidad, sin embargo, también deberá ser establecido el procedimiento para analizar el error del modelo y definir cierto rangos que sean aceptables (en el error) para que el modelo sea válido estadísticamente.

9.2 Datos

9.2.1 Definición descarga de información

Para realizar un análisis de la información existente en el sistema Specto, los esfuerzos serán enfocados en faena Collahuasi, ya que la mayor parte de la información confiable en el sistema corresponde a esta faena.

El modelo predictivo busca estimar cuando un equipo de la flota de esta faena puede fallar, por lo que se define que la base debe contener información de acuerdo a los reportes de falla del área de soporte técnico del *Mining Group*.

Se cuenta con restricciones técnicas de uso y descarga de la información dentro de la base de datos Specto. La unidad de negocio, en especial el área a cargo del sistema, es quien tiene la información disponible en el sistema Specto y ésta se encuentra dividida en 3 partes:

- Información disponible en Specto productivo: corresponde a 1 mes de antigüedad de la información de los distintos motores.
- Versión histórica de Specto, la cual tiene disponible hasta 6 meses móviles.
- Información histórica (mayor a 6 meses de antigüedad) no es posible acceder a menos que se solicite de manera especial a los encargados de la base de datos (no disponible de manera online).

Para la definición de información a descargar y analizar, se establece que deben ser analizados todas las reparaciones imprevistas (RI), reparaciones programadas (RP),

continuación reparaciones imprevistas (CRI) y continuación reparación programadas (CRP) de los equipos que se encuentran en faena Collahuasi.

La información a analizar tiene 2 componentes principales:

1. El registro de las fallas de motores son obtenidos de una base de datos Access correspondiente a la faena en estudio, ésta se alimenta por usuarios de faena de forma directa. Debido a las restricciones indicadas, se establece que la información a descargar será la correspondiente a las fechas entre el 01.09.2015 y el 31.10.2015.

De acuerdo al rango definido, la muestra de información a descargar corresponde a 380 archivos, se muestra un ejemplo del detalle a considerar en la descarga de información:

Fecha Inicio	Fecha Termino	Hr Inicio	Hr Termino	Esn	Tipo	Unidad
01-may-12	01-may-12	7:00:00	13:00:00	66301387	930 SE	77
01-may-12	01-may-12	14:30:00	18:00:00	66301387	930 SE	77
01-may-12	02-may-12	20:00:00	1:00:00	66301387	930 SE	77
02-may-12	02-may-12	8:00:00	13:00:00	66301387	930 SE	77
02-may-12	02-may-12	10:50:00	12:00:00	66302150	930 SE	89
02-may-12	02-may-12	14:30:00	18:00:00	66301387	930 SE	77
02-may-12	03-may-12	20:00:00	1:00:00	66301387	930 SE	77

Tabla 9.1: Planilla maestra información falla de equipos

2. Para la descarga de información desde Specto, los parámetros de entrada son "Faena", "fecha", "tipo" y "unidad". Para el caso del campo "tipo", este se refiere al tipo de flota, que en este caso corresponde a la flota de camiones 930 SE, y la unidad corresponde al número que identifica al camión dentro de esta flota de manera única.

Debido a inconsistencias en la información reportada por la faena, y considerando que el ESN (*engine serial number*) es el número mandatorio por motor, se realizó una cuadratura de esta información con respecto al sistema, arrojando que la muestra disponible disminuye de 380 a 206 archivos a descargar.

9.2.2 Definición Target

El objetivo de este trabajo es entregar herramientas que permitan identificar uno o varios patrones de comportamiento para detectar fallas de forma anticipada, por lo que se busca enseñar a este modelo predictivo ambos casos, que sea capaz de clasificar de acuerdo a la información ingresada, para lo anterior se define lo siguiente:

- Cuando el valor del campo Target es 1, indica que el equipo tuvo problemas en ese periodo de tiempo.
- Cuando el valor del campo Target es 0, indica que el equipo no tuvo problemas en ese periodo de tiempo.

Se define según la Tabla 9.1, que cada archivo contiene información correspondiente a los valores de los sensores para la fecha indicada, y la información corresponde a todo un día de operación de ese motor en particular.

Para que esta información tenga valor en cuanto al objetivo del trabajo, se debe incorporar la variable que indica cuando el motor falló o se encuentra funcionando en estado normal.

9.2.3 Creación Base de datos

Debido a la cantidad de archivos Excel, junto con las características de este trabajo, obliga a que esta información sea almacenada en una base de datos. A modo de resumen, son indicadas las características de esta base de datos:

- Motor de base de datos : MySQL
- Nombre Administrador base de datos : phpMyAdmin
- Tamaño : 930 Mb
- Registros : 2.151.943 líneas
- Campos : 82

El primer análisis exploratorio realizado a los datos, están relacionados al campo más importante que es el Target 1, se puede obtener los siguientes resultados a nivel agrupado:

Target	Cantidad	%
0	1.930.663	89,72%
1	221.280	10,28%
Total	2.151.943	100,0%

Tabla 9.2: Balance muestra base entrenamiento 1. Fuente: Elaboración propia.

En la Tabla 9.2 , la proporción entre la clase de Target (1) y la no Target (0) es de un 10,28% vs un 89,72% respectivamente. Para evitar el problema de muestra desbalanceada, se aplicó la metodología de generar una submuestra de forma aleatoria de la clase “no Target”, donde el objetivo fue generar una submuestra cuya distribución para ambas clases fuera parecida (50%/50%).

Posterior al muestreo, la base de entrenamiento fue reducida, quedando con las siguientes características:

- Motor de base de datos : MySQL
- Nombre Administrador base de datos : phpMyAdmin
- Tamaño : 189,5 Mb
- Registros : 437.042 líneas
- Campos : 82

La literatura sugiere aplicar este método para contrarrestar el desbalanceo de información, ayudando en la práctica a mejorar el rendimiento durante el proceso de revisión y análisis estadístico de la información para la determinación de la base de entrenamiento.

9.2.4 Análisis exploratorio de la base de datos

Al revisar los 82 campos en búsqueda de “missing” fue creada la tabla indicada en el 14.5 Selección campos a considerar en base de datos para la determinación de los campos que se mantienen o se eliminan, fue definido lo siguiente:

- Todos aquellos campos que tienen menos del 99% de los registros serían eliminados.

El criterio anterior permite privilegiar la cantidad de registros en la base de datos en comparación a los campos, por lo que el listado de campos preliminar son los que en el campo llamado “Estado” tienen el valor “Mantener” en el anexo 14.5, disminuyendo la cantidad de campos de 82 a 52.

Los campos que se mantienen, pero no tienen el 100% de la información, son analizados de acuerdo al registro (en la base de datos: “id_datos”) estableciendo que:

- Todos aquellos registros que no tengan los 52 campos con información disponible serán eliminados.

Con el criterio anterior, se logra obtener una base de datos con el 100% de la información disponible, por lo tanto, las características de la base de datos son:

- Motor de base de datos : MySQL
- Nombre Administrador base de datos : phpMyAdmin
- Tamaño : 183,3 Mb
- Registros : 421.677 líneas
- Campos : 52

El balance de la base de datos de entrenamiento queda de la siguiente forma:

Target 1	Cuenta registros	%
0	209.354	49,65%
1	212.323	50,35%
Total	421.677	100,00%

Tabla 9.3: Balance base de datos entrenamiento 2.

En este punto del análisis, la base de datos de entrenamiento en los 52 campos y en sus 421.677 filas tiene *data* (no existen valores vacíos), sin embargo, quedan 2 análisis exploratorios antes de comenzar con el análisis estadístico.

9.2.5 Calidad de información agrupada:

De acuerdo al análisis gráfico que se muestra en el anexo 14.6, el único campo que es eliminado es PowerHP, debido a que sólo tiene valor 0, lo cual no agrega valor para la predicción.

Existen otros campos que tienen comportamiento similar, referente a la cantidad de registros con valor 0, algunos campos incluso poseen el 50% de los registros en este valor (200.000 registros aproximadamente), sin embargo, no son eliminados debido a que tienen información muy variada que podría agregar valor al momento de clasificar los datos reales, por lo que de acuerdo a los resultados y confiabilidad de la predicción, serán considerados estos campos como factores influyentes para la mitigación del error (en caso de que exista).

9.3 Árbol de decisión (CHAID Exhaustivo)

El Árbol de decisión fue ejecutado con el software SPSS Statistics 19 como primer filtro, ya que según las características de este algoritmo, ayuda a determinar los campos que son relevantes y los que no son relevantes para el modelo, por lo que se convertirá en un filtro para los algoritmos siguientes.

El algoritmo de árbol de decisión utilizado es CHAID Exhaustivo, al cual se le entregan las 52 variables iniciales para su tratamiento, donde el algoritmo consideró los siguientes campos como relevantes:

Resultados	Variables independientes incluidas	AVL, Fuel_TemperatureF, Coolant_PressurePSI, Pedal, EGT08F, EGT06F, IMTRBRF, EGT17F, Post_Oil_FilterPSI, Engine_Oil_TemperatureF, EGT14F, EGT04F, IMPLBPSI, RPM, Coolant_temperatureF, IMTRBFF, EGT10F, EGT15F, EGT01F, EGT11F, IMTLBFF, EGT16F, EGT03F, Fuel_rateLh, EGT02F, F_de_Carga, EGTAVF, EGT13F, Battery_potentialV, Estado_Motor, EGT07F, Injector_TimingPSI, Barometric_PressurePSI, Rifle_Oil_PressurePSI, Crankcase_PressureHPInH2O, Percent_Load_At_Current_Speed, Injector_MeteringPSI, Actual_Percent_Torque_, Potencia
	Número de nodos	4648
	Número de nodos terminales	2940
	Profundidad	6

Tabla 9.4: Tabla Variables independientes consideradas en modelo matemático. Fuente: Elaboración propia.

De acuerdo a la Tabla 9.4, la base de datos disminuyó de 52 a 38 variables independientes, las cuales se encuentran indicadas en la Tabla 9.4.

A continuación se muestra la matriz de confusión obtenida para este modelo:

Observado	Pronosticado		Porcentaje correcto
	Target		
Target	0	1	
0	38.339	6.357	73,10%
1	14.109	46.614	88,00%
Porcentaje Global			80,59%

Tabla 9.5: Matriz de confusión Árbol de decisión CHAID Exhaustivo. Fuente. Elaboración propia.

Se puede ver que el modelo global tiene un 80,6% de probabilidad de acertar en su predicción, sin embargo, mayor es la probabilidad de acierto para el caso de la predicción del target 1 (Equipo con falla), con un 88% de probabilidad de acierto.

9.4 Regresión Logística

Regresión logística fue realizada con el software matemático SPSS Statistics 19, donde en la siguiente tabla se muestran los resultados obtenidos, es importante revisar los P-valores para cada variable:

Variables independientes	B	E.T.	Wald	gl	Sig.
IMTRBFF	-,128	,002	4882,397	1	,000
IMTRBRF	,114	,002	3892,268	1	,000
EGT04F	-,003	,000	1472,065	1	,000
EGT14F	,003	,000	1673,424	1	,000
Post_Oil_FilterPSI	,014	,000	2092,410	1	,000
Coolant_PressurePSI	-,050	,001	1428,406	1	,000
EGT06F	,003	,000	1271,346	1	,000
EGT17F	-,002	,000	756,484	1	,000
IMTLBFF	,013	,000	743,875	1	,000
EGT10F	-,002	,000	526,210	1	,000
EGT07F	-,002	,000	520,540	1	,000
Crankcase_PressureHPlinH2O	,053	,002	501,781	1	,000
Coolant_temperatureF	-,012	,001	369,319	1	,000
Fuel_TemperatureF	,007	,000	329,070	1	,000
Engine_Oil_TemperatureF	,008	,000	293,745	1	,000
Pedal	,005	,000	275,954	1	,000
Constante	-3,941	,252	244,732	1	,000
EGT08F	-,001	,000	212,364	1	,000
Battery_potentialV	,087	,007	147,527	1	,000
Barometric_PressurePSI	,251	,021	143,498	1	,000
Estado_Motor			135,228	4	,000
Injector_TimingPSI	-,001	,000	95,873	1	,000
EGTAVF	,002	,000	82,047	1	,000
Estado_Motor(3)	-1,958	,228	73,958	1	,000
Rifle_Oil_PressurePSI	-,008	,001	71,589	1	,000
Estado_Motor(2)	,110	,015	54,059	1	,000
EGT13F	,000	,000	46,673	1	,000
RPM	,000	,000	42,555	1	,000
IMPLBPSI	-,003	,001	38,335	1	,000
EGT01F	,000	,000	27,318	1	,000
EGT03F	,000	,000	24,462	1	,000
Actual_Percent_Torque_	-,012	,003	13,940	1	,000
EGT11F	,000	,000	8,596	1	,003
Estado_Motor(1)	,274	,105	6,857	1	,009

Tabla 9.6: Modelo Regresión Logística. Fuente: Elaboración propia.

Los P-valores para las variables independientes son menores a 0,05, por lo tanto, son estadísticamente significativos.

A continuación se muestra matriz de confusión para la regresión logística:

Observado	Pronosticado		Porcentaje correcto
	Target		
Target	0	1	
0	29.161	17.958	55,60%
1	23.287	35.013	66,10%
Porcentaje Global			60,88%

Tabla 9.7: Matriz de confusión Regresión Logística. Fuente: Elaboración propia.

El modelo tiene un rendimiento de 60,88% global, por otro lado, para el objetivo de este trabajo la probabilidad de acierto es un 66,10%.

9.5 Support Vector Machines

El modelo de *Support Vector Machine* fue desarrollado utilizando el software R, en específico utilizando la librería “e1071”, donde fueron entrenados los modelos de *Support Vector Machine* con *Kernel* Lineal, Radial y Sigmoideal.

9.5.1 Kernel Lineal

A continuación se muestra la matriz de confusión correspondiente:

Observado	Pronosticado		Porcentaje correcto
	Target		
Target	0	1	
0	25.539	26.984	64,82%
1	13.861	39.035	59,13%
Porcentaje Global			61,25%

Tabla 9.8: Matriz de confusión Kernel Lineal

9.5.2 Kernel Radial

La matriz de confusión correspondiente a este *Kernel* es:

Observado	Pronosticado		Porcentaje correcto
	Target 0	Target 1	
0	34.701	10.455	66,25%
1	17.680	42.583	80,29%
Porcentaje Global			73,31%

Tabla 9.9: Matriz de Confusión Kernel Radial. Fuente: Elaboración Propia.

9.5.3 Kernel Sigmoideal

La matriz de confusión correspondiente a este *Kernel* es:

Observado	Pronosticado		Porcentaje correcto
	Target 0	Target 1	
0	26.286	26.163	50,12%
1	26.162	26.808	50,61%
Porcentaje Global			50,36%

Tabla 9.10: Matriz de Confusión Kernel Sigmoideal. Fuente: Elaboración propia.

9.6 Comparación de modelos

9.6.1 Análisis de resultados

9.6.1.1 Métricas de evaluación

A continuación se muestra una tabla comparativa con las métricas indicadas en el punto 4.8.1, las cuales se resumen a continuación:

	Accuracy	Sensivity	Specificity	Precision	Recall	F-measure
Decision Tree	80,59%	85,78%	76,77%	73,10%	85,78%	78,93%
LOGIT	60,88%	61,89%	60,06%	55,60%	61,89%	58,58%
SVM - Lineal	61,25%	48,62%	73,80%	64,82%	48,62%	55,57%
SVM - Radial	73,31%	76,85%	70,66%	66,25%	76,85%	71,15%
SVM - Sigmoideal	50,36%	50,12%	50,61%	50,12%	50,12%	50,12%

Tabla 9.11: Comparación Modelos de Predicción – Métricas. Fuente: Elaboración propia.

De la tabla anterior, se explicará el modelo que mejor se ajusta a cada métrica para finalmente indicar cuál es el modelo seleccionado, de acuerdo a la justificación indicada.

- *Accuracy:*

Es evidente pensar que el modelo a seleccionar debiese ser *Decision Tree* debido a su alto nivel de acierto en la predicción (80,59%), sin embargo, debido a la naturaleza de cálculo de este indicador y lo que persigue este proyecto, deben ser analizados las demás métricas.

- *Sensitivity:*

Por la forma de cálculo, lo importante es que el modelo tenga un nivel alto de sensibilidad (85,78%), esto ya que al disminuir FN, aumentan los TN, que es lo que busca este proyecto. Para este índice *Decision Tree* continúa siendo el mejor modelo.

- *Specificity:*

Se busca que este indicador sea lo más alto posible, debido a que los FP pueden generar trabajo extra en desmedro de otros equipos. Nuevamente es *Decision Tree* quien tiene el mejor índice (76,77%), aunque en este caso, *Support Vector Machine* con *kernel* Lineal estuvo muy cerca (73,80%), sin embargo, debido al *Accuracy* de ambos (80,59% DT vs 61,25% SVM – Lineal), *Decision Tree* continúa siendo la mejor opción.

- *Precision:*

Se busca el mejor indicador ya que es en esta componente de la matriz de confusión donde se pueden generar altos costos por error de clasificación. *Decision Tree* lidera con un 73,10%.

- *Recall*:
Se busca el indicador más alto, debido a que se logra optimizar el uso de los recursos, evitando el malgasto de recurso humano por mala clasificación. *Decision Tree* es quién tiene el mejor indicador con un 85,78%.
- F-measure:
Esta métrica se encarga de medir el grado de influencia de *Recall* y *Precision*, de acuerdo a lo indicado en los puntos anteriores, el mejor valor para esta métrica lo entrega *Decision Tree* con un 78,93%.

En conclusión, por medición de las métricas, se elige *Decision Tree* como el mejor modelo de acuerdo a sus resultados.

9.6.1.2 Función de costos

La función de costo se encuentra dada por:

$$f_{\text{costo total}}(TP, FP, TN, TF) \\ = f_{\text{costo TP}}(TP) + f_{\text{costo FP}}(FP) + f_{\text{costo TN}}(TN) + f_{\text{costo TF}}(TF)$$

Donde $f_{\text{costo TP}}(TP) = 0$

Es importante mencionar que la función de costo sirve para evaluar los casos reales, cuando el modelo se encuentre en funcionamiento, sin embargo, las matrices de confusión no considera los supuestos de entrega de información al usuario, si bien sirve de manera referencial para medir y evaluar el modelo, no serán los costos reales de implementación debido a la naturaleza de los datos utilizados para el entrenamiento del modelo predictivo.

En la siguiente tabla se muestra el cálculo de los costos de acuerdo a las funciones establecidas en el punto 4.8.2:

	Predicción				Costo (M USD)				
	TP	FN	FP	TN	TP	FN	FP	TN	Costo Total
Decision Tree	38.339	6.357	14.109	46.614	0	1.831	456.831	1.509.367	1.968.028
LOGIT	29.161	17.958	23.287	35.013	0	5.172	754.028	1.133.713	1.892.913
SVM - Lineal	25.539	26.984	13.861	39.035	0	7.771	448.816	1.263.945	1.720.532
SVM - Radial	34.701	10.455	17.680	42.583	0	3.011	572.475	1.378.828	1.954.314
SVM - Sigmoideal	26.286	26.163	26.162	26.808	0	7.535	847.120	868.037	1.722.692

Tabla 9.12: Resultados función de costo. Fuente: Elaboración propia.

El algoritmo que mayor costo genera es *Decisión Tree*, esto se debe principalmente a que detecta de mejor forma los *True Negative* que los demás algoritmos, siendo esta función de costo una de las más altas. Por otro lado, para el caso de los *False Negative*, *Decision Tree* es quién genera el menor costo en comparación los demás algoritmos y por último, para los *False Positive* es el segundo mejor costo, después de *Support Vector Machine – Kernel Lineal*.

Como conclusión de esta evaluación, si bien, *SVM – Lineal* genera el menor costo en comparación a *Decision Tree*, este algoritmo tiene mayor probabilidad de error, por lo que no ayuda al objetivo del proyecto, que es detectar de forma efectiva las fallas, antes de que estas ocurran.

Por lo tanto, como conclusión, *Decision Tree* es la mejor alternativa para implementar una solución tecnológica.

9.7 Validación de modelo predictivo

9.7.1 Faena y sistema analizado

Dentro de los contratos MARC que posee Distribuidora Cummins Chile S.A., Collahuasi es la faena que tiene la mayor flota disponible para el cliente como también para el sistema Specto, para lo cual es aprobado el análisis de las fallas detectadas en el periodo de tiempo indicado anteriormente, además, es importante mencionar que esta faena posee la mayor flota de equipos QSK60 y QSK45, siendo estos modelos los de interés para el análisis y objetivo de este proyecto.

Dentro de los motores Cummins, el área de soporte técnico diferencia 9 sistemas principales, los cuales son críticos para la operación normal de los equipos, estos son:

- Potencia
- Escape
- Cambio módulo
- Lubricación
- Combustible
- Refrigeración
- Eléctrico
- Admisión
- Tren de potencia

Se analizaron los tiempos de falla y frecuencia de las fallas de todos los equipos para cada sistema, donde los más relevantes son los eléctrico, admisión y tren de potencia.

Para el caso del sistema eléctrico, existe una alta frecuencia en fallas, sin embargo, no es el sistema que acumula la mayor cantidad de tiempo, siendo un factor crítico para alimentar el modelo matemático y su entrenamiento, por otro lado, el área de soporte técnico indica que este sistema tiene la mayor parte de las fallas tipificadas, siendo el principal motivo de que, a pesar de que son directamente proporcional la frecuencia de fallas y de tiempo de falla, el aumento de una variable no afecta en la misma magnitud que la otra.

El sistema de admisión es uno de los sistemas que mayor tiempo de falla acumula junto con su frecuencia. Este sistema es crítico debido a que regula el ingreso de aire al sistema de combustión del motor, afectando la eficiencia del equipo (y de la operación), junto con el uso óptimo de recursos. Si bien este sistema tiene un alto tiempo acumulado y frecuencia, el área de servicio técnico prefiere realizar análisis posteriores para este sistema debido a la criticidad del sistema restante.

El sistema tren de potencia es seleccionado debido a su implicancia en componentes externos (operación del equipo) e internos (búsqueda de relaciones entre componentes). Del Gráfico 9.2 se desprende que este sistema no es el que mayor frecuencia de falla posee, sin embargo, es el sistema que mayor tiempo de falla acumula. Por otro lado, su importancia radica en los sistemas que afecta, ya sean los ejes, las transmisiones, entre otros, que se relacionan directamente con el desempeño del equipo en su conjunto (motor Cummins y equipo Komatsu), que puede generar la detención de la producción del cliente final.

9.7.2 Evaluación de resultados y modelo matemático

Teniendo en cuenta los puntos 9.1 y 9.7.1, se establece que la formulación del modelo matemático será desarrollada basándose en esos puntos, esto, debido a que los esfuerzos y recursos del área están dirigidos al almacenamiento y administración de la información junto con la continuidad operacional del sistema Specto.

El modelo matemático es considerado válido cuando la información utilizada para su desarrollo considera las fallas clasificadas como RI (reparaciones imprevistas), CRI (continuación reparación imprevista), RP (reparación programada) y CRP (continuación reparación programada) para la faena Collahuasi, cuya fuente de información para estas fallas debe ser el área de estadística del *Mining Group*.

Dentro de la flota de equipos existentes en Collahuasi se encuentran los modelos de motor QSK45 y QSK60, sin embargo, el modelo será desarrollado (inicialmente) en conjunto, sin diferenciación ya que para el *Mining Group* es importante determinar si las fallas son transversales o casos puntuales a los equipos (cuyo análisis no será realizado en este proyecto).

Para el análisis exploratorio de la información, se estableció que aquellas variables (campos) que tienen menos del 99% de la información serían eliminadas del desarrollo del modelo. Lo anterior, debido a que el negocio se encuentra en conocimiento de problemas de emisión de datos, por lo que necesita evaluar (de distintas formas) si

debe enfocar sus esfuerzos en mejorar el hardware de transmisión o si es suficiente realizar análisis con la información existente. Del proceso de limpieza, de 82 variables se disminuyó a 52 variables cumplan con los criterios definidos para ser incluidos dentro del análisis, siendo estas variables válidas para el área de soporte técnico y el área de estadística. Existen campos que dentro de sus registros tienen valores 0, siendo un caso extremo y por lo que fue eliminado el campo "PowerHP" (400.000 registros con valor 0), sin embargo, existen otros campos con aproximadamente 200.000 registros con valor 0 los cuales no fueron eliminados en el análisis exploratorio por petición del área de estadística, esto para que los modelos matemáticos discriminen su relevancia en los modelos finales.

De los modelos utilizados para el análisis de información, *Decision Tree* es el modelo que más acomoda al área de estadística, debido al seguimiento que permite realizar al usuario del sistema, la información que entrega a los técnicos, y el dinamismo y nivel de personalización que se puede lograr en este tipo de modelo matemático.

Es importante la información que entrega *Decision Tree*, sin embargo, también es importante conocer el grado de acierto del modelo, por lo que fueron propuestos por el alumno tesista (según la literatura existente) otros modelos matemáticos que el área de estadística aceptó como comparativo de su principal opción, que es *Decision tree*.

Dentro de los requerimientos más importantes del área de estadística, fue contar con la mayor cantidad de análisis, por lo que para lograr este objetivo se utilizó el algoritmo de *Decisión Tree* CHAID Exhaustivo, cuya principal característica es que no agrupa los nodos de forma binaria, sino que utiliza el estadístico de Chi-square para discriminar entre los distintos grupos de datos.

Los modelos matemáticos que fueron utilizados para comparar los resultados de clasificación de la información, fueron seleccionados de acuerdo a la principal característica de los datos y del objetivo de la tesis, que es la clasificación binaria.

El principal trabajo realizado para la generación de los modelos fue buscar (disminuir) los atributos correctos para el modelo, disminuyendo de 52 variables a 38 variables, donde las variables críticas para validar este modelo (por parte de las áreas de estadística y soporte técnico) corresponden a los sensores de gases de salida, clasificación de motor y estado de motor, las cuales si fueron incluidas.

Por otro lado, la matriz de confusión correspondiente a *Decision Tree* CHAID Exhaustivo entrega un 80,59% de *accuracy* global, además, *precision* para el caso de falla y no falla es de 88% y 73,10% respectivamente. Para el área de estadística del *Mining Group*, este modelo es aceptable cuando estos indicadores superan un 70%.

El resultado de los demás modelos matemáticos, como se puede ver en la Tabla 9.11, no fueron satisfactorios, sin embargo, el modelo de regresión logística permitió validar la relevancia de las variables que *Decision Tree* seleccionó, a pesar, de que los resultados del pronóstico no fueron satisfactorios, optando finalmente por *Decision Tree* como modelo a utilizar.

De los resultados obtenidos, si bien uno de los parámetros que el área de estadística más interesó fue el rendimiento del modelo matemático, las funciones de costo fue lo que permitió seleccionar definitivamente que algoritmo utilizar, ya que la traducción de los resultados es que con el modelo de predicción se logrará ahorrar en costos de mantenimiento en comparación al desempeño actual del proceso de mantenimiento sin apoyo computacional.

En conclusión, *Decision Tree* en cualquier caso fue superior y es el modelo predictivo seleccionado de acuerdo a estos análisis.

10 Evaluación Económica

En este capítulo se indican las inversiones y costos que deberán ser considerados para implementar el proyecto, finalizando con un flujo de caja.

10.1 Inversiones

Para el desarrollo del proyecto e implementación, se debe considerar lo siguiente como inversión:

- **Desarrollo modelo predictivo (externo):**

Se considera la opción de desarrollar el modelo predictivo de forma externa a la empresa, de esta forma se debe considerar como inversión dentro del flujo ya que el costo de desarrollo es de aproximadamente USD 27.000.

- **Software Predictivo:**

La habilitación del software que incluye una capacitación se considera como parte del software, el cual considera un monto inicial de USD 23.000.

- **Licencias Software Predictivo:**

Se considera el uso inicial de 6 licencias, 4 visualizador y 2 expertos para la implementación del sistema.

- **Desarrollo de plataforma web integración:**

Será necesaria la construcción de un sistema que permita a los usuarios utilizar la plataforma sin necesidad de acceder al software directamente de forma amigable, con cierto grado de usabilidad.

- **Capacitación software y modelo:**

Cuando el modelo y la plataforma sean desarrollados, es necesario la capacitación de los sistemas a los usuarios iniciales del sistema, para esto serán utilizados recursos internos que serán sacados de sus funciones regulares.

- **Servidor Software Predictivo:**

Por la carga de procesamiento que se requerirá al momento de ejecutar el modelo predictivo, se optó por adquirir un servidor dedicado para esta tarea.

- **Equipo interno proyecto:**

A pesar de que existirá un equipo externo desarrollando el modelo predictivo, existe un equipo interno en la empresa que estará encargado de proveer información, monitorear y velar por el buen resultado de lo que la empresa externa genere, para esto, se consideró un equipo de 6 personas provenientes del área TI y del área de monitoreo.

10.2 Beneficios

Debido a la naturaleza del proyecto se pueden distinguir 2 tipos de beneficios:

- **Tangibles**

- **Disminución de fallas menores:**

El proyecto fuerza a detectar cualquier tipo de falla, esto ya que desde el momento en que se enfoca inicialmente en las fallas que más tiempo consumen, deberá disminuir las fallas menores de acuerdo al conocimiento técnico del área.

- **Disminución de fallas imprevistas con desmontaje de motor:**

Este tipo de fallas es uno de los más costosos para Cummins, definiendo que esta es una de las oportunidades clave para diferenciarse de los demás competidores desde la eficiencia de la operación en faena.

- **Disminución costos de mantenimiento:**

A medida que el modelo predictivo comience a dar frutos, el conocimiento experto aumentará, lo que permitirá generar diagnósticos más precisos, y por ende, disminuyendo los costos de gestión y tiempo de mantenimiento.

- **Aumento índices de confiabilidad y disponibilidad:**

Al contar con mejor información para el desarrollo de los mantenimientos requeridos, la disponibilidad de los equipos aumentará debido al menor tiempo de mantenimiento (u optimización del trabajo) provocado por mejores decisiones en cuanto a los procedimientos a seguir, lo anterior también provocará que los equipos se mantendrán en operación por mayor tiempo en condiciones específicas.

- **Aumento productividad flota del cliente:**

El tiempo que disminuye en mantención, el cliente lo aprovecha en producción, el cual considera varias toneladas de producción “extra”.

- **Mayor conocimiento de la flota del cliente**

- Intangibles
 - **Mayor satisfacción del cliente:**

Debido principalmente a la proactividad que se hará presente en los distintos procesos de mantenimiento, y que se harán evidentes desde el momento en que el sistema sea implementado.
 - **Aumento de confianza con productos Cummins.**
 - **Generación de conocimiento experto en personal Cummins y de faena (cliente).**

10.3 Costos

De acuerdo a las políticas contables que establece el holding, como costo se considera todo aquello que esté relacionado al giro de la empresa, desde este punto de vista, el proyecto sólo considera un costo:

- **Costos de operación (extra):**

Debido a la menor cantidad de horas de mantenimiento, las horas de operación aumentan, por lo tanto, aumenta el ingreso por horas de operación y por ende los costos variables. Dentro de este costo se encuentra considerado el costo de operación de las máquinas junto con el costo de mantenimiento extra que será recibido a consecuencia de las propuestas de revisión que el sistema entregará.

10.4 Gastos

El proyecto al ser un desarrollo relacionado al área TI, la mayor parte de los gastos se encuentra asociado a la mantención del sistema desarrollado para esto, se considera lo siguiente:

- **Equipo TI:**

Se considera como parte del equipo TI a dos personas encargadas del mantenimiento del sistema desarrollado, uno enfocado en la mantención de la interfaz Web y algunos ETL's de mediana complejidad (Analista funcional) y un Consultor funcional encargado de la operación de la base de datos que mantendrá la información transaccional para análisis operativa. Serán estas dos personas quienes mantendrán operativo el sistema, donde se estima se utilizará el 15% de su tiempo mensual.

- **Equipo de analistas:**

Se considera toda el área de monitoreo, el cual se compone de 3 analistas y un jefe, los cuales serán los usuarios directos del proyecto. Se estima que su tiempo de utilización será de un 30% del tiempo total.

- **Mantención licencias software predictivo:**

Después de la inversión inicial en licencias, se debe pagar el 22% del precio total de la inversión por concepto de mantenimiento, el cual considera 6 licencias (4 visualizadores y 2 expertos).

- **Mantención Servidor BD:**

Por la magnitud de *data warehouse* que será construido, el área TI del *holding* exige el apoyo de una empresa externa que mantenga el sistema en el régimen 24x7.

10.5 Depreciación

De acuerdo a políticas contables, un desarrollo web y su software deben ser activados, por lo tanto, genera la siguiente depreciación:

- **Amortización de licencias:**

Considera el costo de las licencias del software para el desarrollo de la plataforma Web junto con el costo de las licencias del software predictivo.

- **Software Predictivo:**

La inversión del costo de adquisición del software que no corresponde al licenciamiento, también se considera para la activación, por lo que debe depreciar.

- **Servidor BD:**

La compra de 2 servidores durante el periodo de evaluación del proyecto debe ser considerado como depreciación, ya que se convierte en activo fijo de la empresa.

- **Plataforma Web:**

El desarrollo de la plataforma se convierte en un activo de la empresa como un software, el cual también consideramos como depreciación en el flujo de caja del proyecto.

10.6 Otras consideraciones

10.6.1 Horizonte de evaluación

Para definir este parámetro, se debió considerar:

- Proyección de inversiones en el área de minería.
- Proyección de ingresos en los próximos años.
- Estado actual de los contratos que existen en la empresa.

Se espera que los contratos actuales no varíen en los próximos 5 años de forma relevante, es por ello que se pretende ejecutar este proyecto en las faenas que actualmente tienen contrato, sin dejar de mencionar que el sistema se podrá ejecutar en cualquier faena, estableciendo el periodo de evaluación a 5 años, considerando el año fiscal correspondiente al año japonés de medición a todo el holding (por ejemplo, el FY 2015 inicia en Abril 2015 y termina en Marzo 2016).

10.6.2 Impuestos

Al ser una empresa con fines de lucro, dentro del flujo de caja deben ser considerados los impuestos a pagar por las utilidades que el proyecto traerá consigo, por lo tanto, se establece, considerando que la reforma tributaria indica que el impuesto a las utilidades de las empresas aumentará, de acuerdo a la siguiente tabla:

Año tributario	Año Comercial	Impuesto
2012 al 2014	2011 al 2013	20%
2015	2014	21%
2016	2015	22,5%
2017	2016	24%
2018	2017	25%
2018 (*)	2017	25,5%
2019 (*)	2018	27%

Tabla 10.1: Impuesto utilidades para empresas con fines de lucro. Fuente: Servicio de impuestos internos.2016.

Los asteriscos en la Tabla 10.1 indican que esos porcentajes deben ser aplicados a la utilidad de la empresa siempre y cuando la empresa se encuentre en régimen semi-integrado, por el contrario, si la empresa se encuentra en régimen renta atribuida, tendrá el beneficio de que el impuesto sólo aumentará hasta un 25%.

Komatsu Cummins Chile S.A. se rige por el régimen semi-integrado, por lo que se aplicarán los impuestos mencionados (con un máximo de 27%).

10.6.3 Tasa de descuento

La tasa de descuento representa el costo de oportunidad que una empresa considera al momento de asignar sus recursos en una inversión o en otra. Para Komatsu Cummins Chile Ltda., la tasa de descuento a emplear es de 10%, tasa entregada por el área de finanzas del holding quién administra todo este tipo de proyectos.

10.7 Construcción Flujo de Caja

- **Inversión:**

Las inversiones que serán necesarias para la implementación del proyecto se detallan a continuación:

- **Desarrollo externo modelo predictivo:**

Se considera la participación de una empresa consultora, ésta generará los modelos predictivos necesarios para el análisis de la información basado en los requerimientos del negocio. La inversión por este concepto es de USD 27.000, con una vida útil de 5 años (60 meses).

- **Software Predictivo:**

La implementación del software en equipos de Distribuidora Cummins junto con capacitación de uso y pequeña asesoría es valorizada en USD 23.000 con una vida útil de 5 años (60 meses).

- **Licencias Software Predictivo:**

Se considera 2 licencias experto y 4 licencias experto, las que permiten visualizar resultados en los servidores de desarrollo, junto con el desarrollo de modelos analíticos a medida, dentro del software ya implementado en el punto anterior. Este monto esta valorizado en USD 11.096 (2x USD 3.148 experto, 4x USD 1.200 visualizador) con una vida útil de 5 años (60 meses).

- **Desarrollo plataforma Web e integración:**

Se considera el analista funcional que se encuentra en el área de TI, full time, ya que se encontrará dedicado al desarrollo de este sistema junto con la tarea de integrar otros sistemas existentes en el área. Se estima que esta persona se dedicará 2 meses al desarrollo de la plataforma, cuyo costo será de USD 4.720 con una vida útil de 5 años (60 meses).

- **Capacitación software y modelo predictivo:**

Considerando 20 días mensuales hábiles, se considera que la capacitación del uso del software (software predictivo y plataforma web) y la explicación de desarrollo de los modelos utilizará al jefe de proyectos como profesor, junto con todo el equipo de analistas, por 3 días completos, lo cual se valoriza en USD 1.723 con una vida útil de 5 años (60 meses).

- **Servidor software predictivo y base de datos:**

Se estima la inversión de 2 servidores de base de datos y software predictivo debido a las políticas de activación que el holding indica. Para esta inversión se estiman USD 9.000 para la inversión en el año 0 y USD 9.000 para finales del año 3, con una vida útil de 3 años cada uno (36 meses).

- **Equipo interno proyecto:**

El equipo interno de proyecto considera a los recursos que se encuentran dentro de la empresa (que no son parte de la empresa consultora que generará modelos predictivos). A continuación se muestran los costos de cada uno de los integrantes:

RRHH	Cargo	Sueldo Mensual (CLP)	Sueldo Mensual (USD)	Costo x hora (USD)
Equipo Analistas	Jefe Monitoreo	\$ 2.000.000	\$ 3.147	\$ 20
Equipo Analistas	Analista Monitoreo	\$ 1.000.000	\$ 1.573	\$ 10
Equipo Analistas	Analista Monitoreo	\$ 1.000.000	\$ 1.573	\$ 10
Equipo Analistas	Analista Monitoreo	\$ 1.000.000	\$ 1.573	\$ 10
Desarrollo Interno	Jefe de Proyecto	\$ 2.300.000	\$ 3.619	\$ 23
Desarrollo Interno	Analista Proyecto	\$ 1.500.000	\$ 2.360	\$ 15

De los cuales, se considera que el equipo de analistas le dedicará un 30% de su tiempo total al desarrollo del proyecto por 4 meses, por otro lado, el equipo de desarrollo interno le dedicará el 20% de su tiempo total por 4 meses al proyecto, lo cual se encuentra valorizado en USD 14.223 cuya vida útil es de 5 años (60 meses).

- **Ingresos por venta:**

Se proyecta que por el concepto “Horas de operación (extra)”, cada año se obtendrá un ingreso extra por 0,05%, sobre la proyección anual normal para cada año.

Para el caso del concepto “Ahorro por detección de falla imprevista”, se considera que se logrará disminuir en al menos 1 falla catastrófica, donde la estadística indica que anualmente se generan 31 fallas imprevistas.

El monto asociado al ahorro por este concepto dependerá del valor total de reparación, y que se estableció de acuerdo a la siguiente proyección:

Tipo	Costo (USD)				
	FY 2016	FY 2017	FY 2018	FY 2019	FY 2020
Proyección Overhaul	315.000	330.750	347.288	364.652	382.884

- **Costos de venta:**

El costo asociado al concepto “Costo de operación (extra)”, fue estimado en base a los márgenes que el área tiene considerado para ese periodo, en donde se consideran los costos de operación más un delta correspondiente al uso en mayor medida de los técnicos. Los márgenes establecidos por año son los siguientes:

	FY 2016	FY 2017	FY 2018	FY 2019	FY 2020
Margen %	22%	25%	20%	28%	28%

- **Gastos de Administración y Ventas:**

Dentro de este concepto se encuentran 4 ítems:

- **Equipo TI:**

Consta de los siguientes integrantes con sus costos asociados:

RRHH	Cargo	Sueldo Mensual (CLP)	Sueldo Mensual (USD)	Costo x hora (USD)
Equipo TI	Analista Funcional	\$ 1.500.000	\$ 2.360	\$ 15
Equipo TI	Consultor Senior	\$ 2.900.000	\$ 4.563	\$ 29

Del costo total mensual de este equipo, se considera que el tiempo dedicado a la ejecución del proyecto de manera anual (y mensual) será de un 15% del total de su tiempo.

- **Equipo Analistas:**

Consta de los siguientes integrantes con sus costos asociados:

RRHH	Cargo	Sueldo Mensual (CLP)	Sueldo Mensual (USD)	Costo x hora (USD)
Equipo Analistas	Jefe Monitoreo	\$ 2.000.000	\$ 3.147	\$ 20
Equipo Analistas	Analista Monitoreo	\$ 1.000.000	\$ 1.573	\$ 10
Equipo Analistas	Analista Monitoreo	\$ 1.000.000	\$ 1.573	\$ 10
Equipo Analistas	Analista Monitoreo	\$ 1.000.000	\$ 1.573	\$ 10

Del costo total del equipo, se considera que el tiempo dedicado a la ejecución del proyecto de forma anual (y mensual) es el 30%.

- **Mantenición licencias:**

Por el tipo de licenciamiento, se debe pagar de forma anual el 22% de la inversión correspondiente a las licencias. De acuerdo a la cantidad

indicada en las inversiones, el total a pagar (considerando que la cantidad de licencias no variará) será de USD 2.441 de forma anual.

- **Mantenimiento servidor BD:**

Se considera el costo de servicio de mantenimiento externo del servidor de base de datos al considerarse como crítico por las definiciones del proyecto y por el área de TI del holding, por lo que se considera una mensualidad de USD 1.569 por los 5 años que dura el proyecto.

- **Depreciación:**

De los costos del proyecto que se deben activar, se presentan los siguientes ítems:

- **Amortización licencias:**

En este ítem se amortizan las licencias correspondientes al software predictivo más las licencias utilizadas para desarrollar el entorno web, junto con sus funcionalidades. El tiempo de depreciación considerado para el software se considera de 5 años.

- **Software predictivo:**

La inversión inicial del software predictivo obliga a que sea activado, y por ende genera depreciación, este activo también se deprecia a 5 años.

- **Depreciación servidor:**

De acuerdo a políticas del holding, todo hardware de equipos computacionales debe depreciar a 3 años, es por esto que se genera una nueva inversión a finales del año 3.

- **Implementación proyecto:**

La depreciación considerada en este ítem involucra la inversión por concepto de desarrollo del modelo predictivo, capacitación del software y modelo predictivo y los costos del equipo interno de proyecto.

Con todo lo mencionado anteriormente, se muestra el flujo de caja:

Flujo de caja Proyecto MBE

	Año 0	Año 1	Año 2	Año 3	Año 4	Año 5
(+) Ingresos por Ventas		352.441	374.667	397.681	421.522	446.231
Horas de operación (extra)		37.441	43.917	50.394	56.870	63.347
Ahorro por detección fallas imp.		315.000	330.750	347.288	364.652	382.884
(-) Costos de venta		-29.204	-32.938	-40.315	-40.947	-45.610
Costo operación (extra)		-29.204	-32.938	-40.315	-40.947	-45.610
(-) Gastos de Administración y Ventas		-62.054	-62.054	-62.054	-62.054	-62.054
Equipo TI		-12.460	-12.460	-12.460	-12.460	-12.460
Equipo Analistas		-28.319	-28.319	-28.319	-28.319	-28.319
Mantención Licencias		-2.441	-2.441	-2.441	-2.441	-2.441
Mantención Servidor BD		-18.833	-18.833	-18.833	-18.833	-18.833
(-) Depreciación		-19.352	-19.352	-19.352	-19.352	-19.352
Amortización Licencias		-3.163	-3.163	-3.163	-3.163	-3.163
Software Predictivo		-4.600	-4.600	-4.600	-4.600	-4.600
Depreciación Servidor		-3.000	-3.000	-3.000	-3.000	-3.000
Implementación proyecto		-8.589	-8.589	-8.589	-8.589	-8.589
(=) Resultado Operacional (A)		241.831	260.323	275.960	299.169	319.215
(=) Resultado No Operacional (B)		0	0	0	0	0
(=) Utilidad antes de Impuesto (A+B)		241.831	260.323	275.960	299.169	319.215
(-) Impuesto a las empresas		-54.412	-62.478	-70.370	-80.776	-86.188
(=) Utilidad después de Impuesto		187.419	197.846	205.590	218.394	233.027
(+) Depreciación		19.352	19.352	19.352	19.352	19.352
(=) Flujo Operacional (C)		206.771	217.198	224.943	237.746	252.380
(-) Inversión Fija	-90.761	0	0	-9.000	0	0
Desarrollo Externo Modelo Predictivo	-27.000					
Software Predictivo (SAS Miner)	-23.000					
Licencias Software SAS Miner	-11.096					
Desarrollo Plataforma Web e integración	-4.720					
Capacitación software y modelo	-1.723					
Servidor Software Predictivo/BD	-9.000			-9.000		
Equipo interno proyecto	-14.223					
(=) Flujo de Capitales (D)	-90.761	0	0	-9.000	0	0
(=) Flujo de Caja Privado (C+D)	-90.761	206.771	217.198	215.943	237.746	252.380
VAN				848.808		
TIR				231%		

* Montos en USD

Tabla 10.2: Flujo de caja para proyecto MBE: Desarrollo de modelo predictivo de mantenimiento. Fuente: Elaboración propia.

10.8 Análisis de sensibilidad

De acuerdo al análisis del flujo de caja anterior, y teniendo en cuenta las variables importantes del proyecto, se definieron las variables que son sensibles en cualquier caso (o las de mayor variabilidad) para el flujo de caja, siendo éstas el aumento de horas en la disponibilidad y las fallas imprevistas que se logran evitar.

Para el caso del flujo de caja, se consideró un aumento de un 0,05% para el aumento de los ingresos por disponibilidad sobre las proyecciones realizadas sin proyecto, por otro lado, se considera que al menos se reducirán las fallas imprevistas en 1 por cada año.

Para el análisis de sensibilidad se considera que 0,05% de aumento es el máximo que se puede lograr debido a que no existe evidencia de que se pueda aumentar en mayor cantidad, no así en el caso de las fallas imprevistas que se logren evitar.

Se presentan los análisis de sensibilidad para ambos casos:

- a. % Disponibilidad fija: 0% - Fallas imprevistas (*overhaul*) variable

Cantidad Overhaul	0	1	2	3	4	5
VAN (M USD)	-164	56	264	459	640	814
TIR	??	??	191%	215%	222%	224%
% Disponibilidad	0%	0%	0%	0%	0%	0%

Para este escenario, el peor de los casos obviamente es cuando no se detecta ningún *overhaul* y por lo tanto no se logra el ahorro esperado, en esta situación el VAN (en miles de USD) que entrega el flujo de caja para el proyecto a 5 años es de USD -164.000, lo que no permite recuperar la inversión de USD 90.761, por otro lado, si se detecta (o evita) 1 *overhaul*, el VAN es de USD 56.000 (aproximadamente), lo que no es del todo malo ya que sería en este paso el proceso de aprendizaje y el proyecto tiene un horizonte de 5 años. En el caso que se disminuya en 2 fallas imprevistas, se logran beneficios cuantificables en

el tiempo y desde ahí en adelante comienzan los retornos para la empresa. Se debe considerar para efectos de este análisis que es la variable más importante y que generó el desarrollo del presente proyecto.

b. % Disponibilidad variable – Fallas imprevistas (*overhaul*): 0

% aumento disp.	-0,05%	-0,03%	-0,01%	0%	0,01%	0,03%	0,05%
VAN (M USD)	-199	-185	-171	-164	-158	-144	-130
TIR	??	??	??	??	??	??	??
Cantidad Overhaul	0	0	0	0	0	0	0

En este escenario, para el caso de que la cantidad de *overhaul* detectados es 0 y el porcentaje de aumento de disponibilidad es 0%, se da el mismo efecto que el punto a, sin embargo, si se cumple el aumento estimado por los expertos del negocio de un 0,05% (como máximo), el proyecto no logra recuperar la inversión inicial, y eso se debe principalmente a que el aumento de la disponibilidad será muy bajo.

11 Generalización de la experiencia

11.1 Introducción

Los patrones de negocios expuestos por el profesor Oscar Barros, nos entrega un *framework* que permite ajustar realidades de distintas empresas a una estructura clara que facilita el análisis de forma ordenada de los procesos y procedimientos existentes, permitiendo identificar errores, falencias y fortalezas en cada una de ellas, no importando el rubro de la organización a analizar.

De forma análoga, será generado un *framework* o marco de trabajo, el cual estará compuesto de clases y lógicas de negocio que servirán de base para el desarrollo de un software con los requerimientos de cualquier empresa, que en este caso, será una abstracción de lo realizado en este proyecto de memoria, cuyo requerimiento se encuentre dentro del dominio de esta tesis.

El *framework* planteado por Oscar Barros en 2004, busca facilitar y acelerar el proceso de desarrollo promoviendo buenas prácticas mediante el uso de patrones de procesos de negocio.

11.2 Aplicación del *framework*

El proceso fundamental de este proyecto es el desarrollo del proceso de “Mantenimiento Predictivo”, donde el objetivo principal es detectar equipos que puedan fallar y equipos que no fallarán, además, el sistema debe ser capaz de clasificar la predicción de acuerdo a una probabilidad de ocurrencia preestablecida.

El dominio lo constituyen empresas que posean maquinaria con sensores de medición de sus variables críticas, ya sea recibiendo o entregando servicios por parte de estas máquinas. Se proponen un *framework* que disponibilizará los modelos matemáticos indicados en el marco teórico. Si bien la necesidad de la empresa es, además de la clasificación entre si el equipo fallará o no fallará, la información que puede entregar un árbol de decisión, este *framework* se encargará de entregar la mejor predicción, sin considerar si es árbol de decisión u otro modelo matemático.

11.3 Construcción del *framework*

Un *framework* se encuentra compuesto por clases de elementos que permiten realizar un diseño orientado a objetos del modelo, y que sea capaz de integrarse con los sistemas de la organización.

En primer lugar se muestra el diagrama de clases de control:

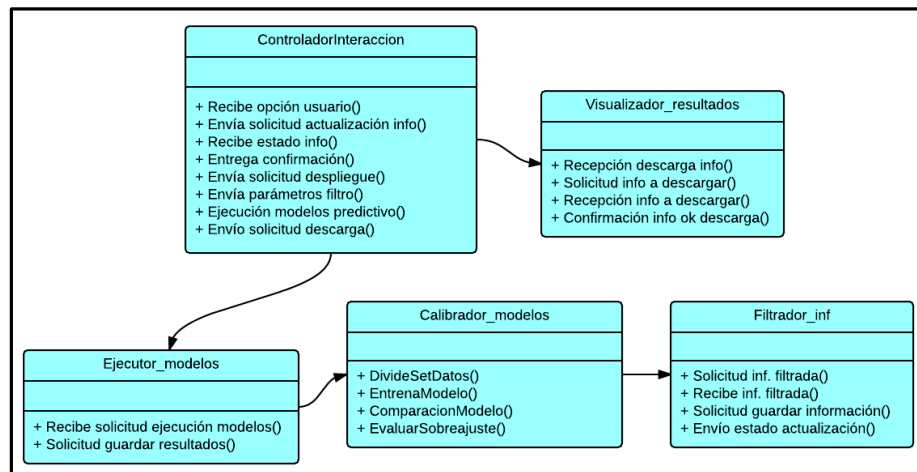


Figura 11.1: Diagrama de clases de control. Fuente: Elaboración propia.

- **Controlador_Interaccion:** Posee métodos que permiten controlar la interacción completa del sistema, recibe las respuestas del usuario desde la interfaz gráfica, hasta coordinar la secuencia de ejecución de las clases de control.
- **Filtrador_inf:** Contiene métodos que toman la información total del día para analizar la información y aplica los filtros correspondientes, de acuerdo a las preferencias del usuario accediendo a las bases de datos necesarias y guardando la información en tablas especiales.
- **Calibrador_modelo:** Contiene métodos que permiten tomar la información a analizar, separa las muestras entre *test* y *train* (de acuerdo a la necesidad), compara distintos modelos (utilizando los que se exponen en el marco teórico), para seleccionar el mejor resultado y evalúa el sobreajuste.

- Ejecutor_modelo: Contiene métodos que permiten ejecutar un modelo matemático predeterminado para obtener resultados de clasificación de la información a analizar, entregando los resultados para que sean guardados.
- Visualizador_resultados: Contiene métodos abstractos que permiten ajustar los resultados de los distintos modelos a la salida que el usuario requiere.

Para el caso de las clases que requieren acceso a información, deberán ser ajustados a las bases y modelos de datos de la organización definiendo los protocolos y mecanismos de conexión.

Las clases que cumplen con el procesamiento estadístico de la información, se requerirá que estos se adapten al software que la empresa determine y que es óptima de acuerdo a sus necesidades y capacidad instalada.

A continuación se muestra el diagrama *Entity*:

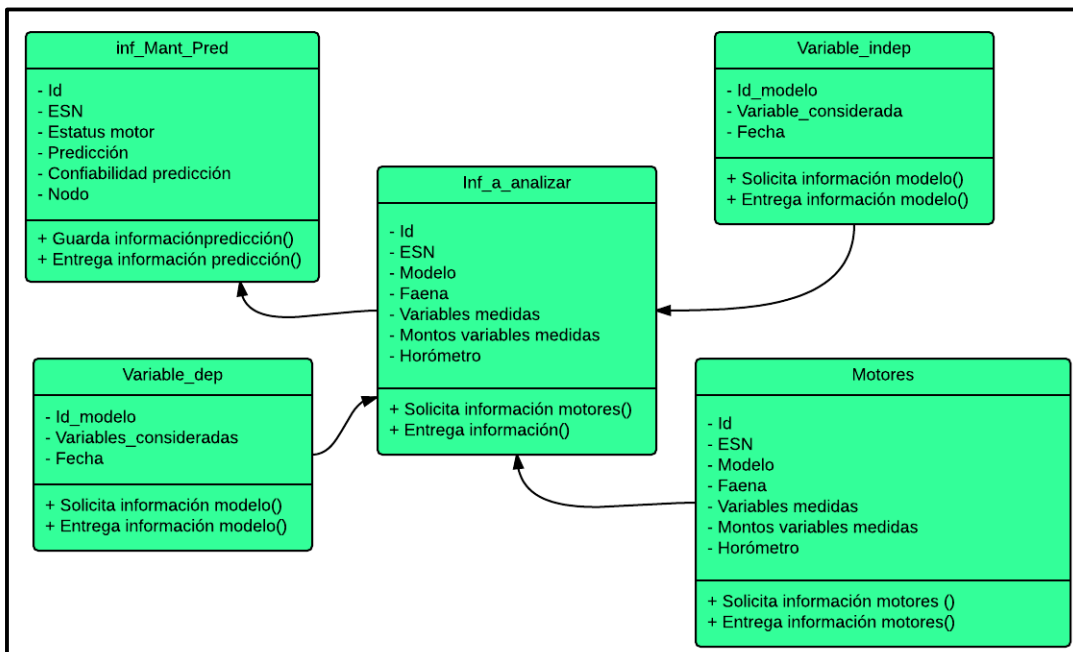


Figura 11.2: Diagrama Entity. Fuente: Elaboración propia.

- `Inf_mant_pred`: Esta clase guarda los resultados obtenidos por la clase control `Ejecutor_modelo`, guarda la predicción, probabilidad de ocurrencia, en el caso de los árboles de decisión, guarda el nodo por el cual fue clasificado, permitiendo recuperar la información pronosticada.
- `Motores`: Clase que contiene toda la información obtenida para análisis. Permite mantener un histórico con toda la información que en desarrollos posteriores permitirá realizar nuevos análisis. Es aquí donde no se realiza filtro por variable para análisis posteriores.
- `Inf_a_analizar`: Clase que almacena la información que será clasificada, la cual debe contener todos los campos definidos en la clase `Entity Variables_indep`.
- `Variable_indep`: Clase que almacena las variables independientes de acuerdo al modelo establecido en la clase `Calibrador_modelo`.
- `Variable_dep`: Clase que almacena las variables dependientes que la clase `Calibrador_modelo` determinará que es importante para el modelo de acuerdo al proceso de generación modelo matemático.

11.4 Aplicación generalización en otros dominios

Los dominios válidos para este *framework* son todos aquellos donde existan motores los cuales posean sensores capaces de alimentar una base de datos ya sea de forma online o de forma regular extrayendo la información cada cierto tiempo. Desde este punto de vista, existen muchos ámbitos donde esta aplicación puede ser implementada, empresas dedicadas al transporte terrestre y/o aéreo (logística, encomiendas, etc), mantenimiento de maquinaria para *packing*, aserraderos automáticos, entre otros procesos que dependan del rendimiento de maquinaria que

cumpla con las condiciones de monitoreo de datos y almacenamiento en bases de datos computacionales.

Desde este modo, las Macros, dependiendo del rubro de la empresa es donde deberá implementarse el diseño o rediseño del proceso, por ejemplo, para el caso de este proyecto, el proceso de mantenimiento se encuentra relacionado con la cadena de valor del principal producto entregado por esta empresa, por lo que se definió que el rediseño se implementaría en Macro 1 – Cadena de valor, sin embargo, dentro de las macros existentes, las empresas, de acuerdo a sus necesidades podrían intentar rediseñar dentro de otras macros tales como Macro 2 – Creación de nuevas capacidades o Macro 4 – Procesos de apoyo.

Para Distribuidora Cummins Chile, importante es la ejecución de este proceso y el rendimiento de las predicciones, donde existen el recurso humano para monitorear y determinar que efectivamente los modelos estén ajustados a su realidad, sin embargo, en caso de necesitar modificar los parámetros de los modelos o incluso cambiar los algoritmos, es necesario contratar a una empresa consultora para que realice el estudio y puedan ser asesorados, ya que por el momento no están interesados en generar la capacidad internamente.

En una empresa consultora, cuyo foco sea el mantenimiento de motores mineros, camiones, aviones y/o maquinaria industrial, por obligación tendrá que generar rediseño en Macro 2 – Creación de nuevas capacidades y posiblemente deba modificar una de sus variadas cadenas de valor en Macro 1 – Cadena de valor para lograr implementar su nuevo servicio.

Por otro lado, es interesante revisar cómo se podría aplicar este *framework* a una empresa de transporte terrestre de sustancia peligrosa por carretera, obviamente de forma muy genérica. Su principal servicio es la entrega oportuna de productos en los puntos y horas donde el cliente lo requiera, estos productos son transportados por

camiones donde esta empresa posee capacidad de acceso a los parámetros del motor de cada camión.

En este caso, suponiendo que la empresa de transporte tiene las mismas condiciones que Distribuidora Cummins Chile, y no está interesada en crear una nueva capacidad, solo genera el rediseño en Macro 1 – Cadena de valor, debiendo enfocarse también en Planificación y control producción. La diferencia principal existe por la naturaleza de los servicios que entrega cada empresa, por una parte Distribuidora Cummins Chile requiere entregar el servicio a los equipos que el cliente le indique, en el caso de la empresa de transporte terrestre de sustancia peligrosa, al cliente sólo le interesa que el producto sea entregado, no le interesa el camión que lo haga (esto principalmente por el producto que es transportado y las características del negocio). Es por lo anterior que en esta empresa cambian los supuestos y la forma de implementación, debiendo incluir en Planificar capacidad un proceso que permita discriminar entre los camiones que se encuentran en buenas condiciones para utilizarlos en la planificación de la producción y no considerar a los que tengan una mayor probabilidad de falla. El caso de una empresa de transporte como la que se describe, una componente como la posición geográfica es clave para determinar si debe ser considerado un vehículo como parte de la planificación de capacidad o no, ya que en este caso la Planificación de producción se debe realizar de acuerdo a los requerimientos del cliente y que no consideran el estado del motor.

El proceso que discrimina los equipos que son candidatos para los requerimientos del cliente debe convertirse en *input* para el proceso de mantenimiento de esta empresa, que según la estructura de macros, debiese encontrarse en Macro 1 – Cadena de Valor, específicamente para la administración relación con proveedores (los cuales pueden ser internos o externos).

Para el caso de la empresa de transporte, el objetivo que cumple la predicción de falla del motor es distinta ya que principalmente se utiliza para descartar equipos de forma previa, el segundo objetivo es muy parecido, ya que permite estimar cuando se

requerirán servicios de proveedores de mantenimiento, aunque, si bien este tipo de empresa tiene sus talleres propios, la mayoría de las veces reaccionan cuando los equipos ya han fallado (mantenimiento correctivo).

De acuerdo a lo ya mencionado, y el punto 11.3 Construcción del *framework*, es posible aplicar este *framework* ya que cumple con el objetivo de la necesidad final de la empresa de transporte terrestre de sustancias peligrosas que es no fallar en su servicio final que es la entrega de sus productos.

12 Conclusiones

12.1 Planteamiento estratégico

El análisis del planteamiento estratégico *Mining Group* está centrado en eficiencia operacional, basado en la necesidad de mejorar el desempeño de su servicio de mantenimiento preventivo a mantenimiento predictivo.

Esta necesidad surge de la gran cantidad de fallas imprevistas que por su naturaleza afectan los indicadores económicos, y que este trabajo busca disminuir.

- Se establece la búsqueda de integración de los procesos junto con tecnología (apuntando a tecnologías de información) como un punto relevante dentro de la estrategia, el cual tendrá un impacto en todas las perspectivas del mapa estratégico (clientes, procesos y uso del recurso humano).

El área del *Mining Group* cuenta con un mapa estratégico muy bien estructurado que ayuda a entender e identificar donde deben ser enfocados los esfuerzos de rediseño de acuerdo a la metodología utilizada en esta tesis, sin embargo, es importante que existan instancias de revisión y actualización de esta información cuando sea necesario, debido a que en un inicio, el análisis fundamental para detectar la oportunidad de mejora y establecer todo lo referente a la aplicación de estructura de Macroprocesos estuvo guiado por antiguos lineamientos estratégicos, que no tenían relación con los objetivos estratégicos que las altas jerarquías de Distribuidora Cummins Chile buscaban en ese momento, este proceso podría estar relacionado a una oportunidad de mejora en el rediseño de Macro 3: Planificación del Negocio, y como también el desarrollo de Macro 2 Desarrollo de nuevas capacidades, para mantener en el tiempo esta habilidad y otras que serán explicadas más adelante.

12.2 Arquitectura de macroprocesos

De acuerdo al análisis, el trabajo se enfoca en dos focos principales:

- Desarrollo de proceso de mantenimiento predictivo, el cual permite analizar la información existente, que además, es añadido a la planificación de mantenimiento que se utiliza a la fecha.
- Se desarrolla el proceso de control automático que básicamente utiliza la mayoría de los pasos del proceso de mantenimiento predictivo.

El rediseño se desarrolla en el Macroproceso 1, afectando directamente la cadena de valor que corresponde al servicio de mantenimiento. Dentro del macroproceso son intervenidas las tareas de planificación y gestión de los servicios de mantenimiento, específicamente lo que corresponde a Determinar motores para mantenimiento, siendo en este punto donde el mantenimiento predictivo debe aportar con su pronóstico, pero además, fue diseñado un proceso automático que funciona de manera proactiva (parámetros configurable del usuario) para que el sistema genere alertas de forma automática.

Una de las tareas más complicadas de realizar es la determinación de cuál es la oferta y cuál es la demanda para cada área dentro de las unidades de negocio, especialmente cuando es necesario adaptarlo a los macroprocesos, ya que es necesario mirar el área como una empresa en sí, dado que la estructura de este *framework* puede ser en algunas ocasiones confusa, pero que finalmente ayudan a determinar de mejor forma los límites para cada área/proceso de una empresa. Importante son los diagramas en notación BPMN, debido a que entregan claridad de los distintos procesos y sus distintas interacciones, junto con sus alcances. Esta notación permite identificar interacciones que no son sencillas visualizar, generando procesos eficientes, eliminando ambigüedades o duplicidad de tareas, ya sea en el diseño, rediseño o procesos antiguos.

El uso de estos diagramas y los macroprocesos permitió determinar que el nuevo proceso de Mantenimiento Predictivo para ser implementado, sólo afecta las tareas del Ingeniero de Planificación, y no afectará la cadena de valor en su forma actual, sino que viene a mejorar sus resultados, siendo perfecto para el negocio y sus trabajadores el comenzar a interactuar con este nuevo proceso de forma gradual. A medida que los usuarios comiencen a utilizar el sistema, verán los beneficios y comenzarán a generar nuevos requerimientos, los cuales siempre deberán evaluarse para verificar que son aplicables a este sistema o de otra forma, reevaluar el diseño actual, sin embargo, el trabajo de estandarización para su posterior análisis tiene su inicio en esta tesis.

En este proyecto queda pendiente extender el alcance del mismo, ya que aprovechando la capacidad de predicción del sistema, es natural pensar que identificando la demanda pronosticada por estos servicios, se puede realizar una conexión entre la demanda de servicio de mantenimiento por cada motor y su demanda de insumos y repuestos, lo que involucra el sistema de inventario (que de acuerdo a la estructura de este *Holding* es necesario buscar dentro de Macro 4 Procesos de Apoyo), por otro lado, cuando el sistema tenga madurez, y exista información relacionada a los tipos de fallas, será posible identificar los equipos que fallarán (insumos necesarios junto con la logística correspondiente) y el motivo de la falla, lo que permite identificar el tipo de mecánico (y sus conocimientos) que serán necesarios en ese momento (RRHH cuenta con ese tipo de registro en un sistema *legacy*), y esto, por mencionar 2 de muchas otras interesantes alternativas de rediseños,

12.3 Prueba de concepto

De las pruebas realizadas, se intenta establecer un modelo matemático que satisfaga la necesidad de Distribuidora Cummins Chile, por lo tanto, lo lógico sería que el sistema fuese capaz de dar soporte mediante distintos algoritmos a la vez, para que de esa

forma el usuario tome la mejor opción disponible, sin embargo, por la complejidad de este proyecto, en especial, la disponibilidad de información, se vuelve muy complejo por el volumen de datos que el sistema deberá manejar.

Al comparar los resultados de los modelos matemáticos, en todas las mediciones los resultados de *Árbol de decisión* fue superior a los demás algoritmos, superando incluso a *Support Vector Machine* con *kernel* sigmoideal. Existen diferentes tipos de árbol de decisión, sin embargo, el uso del algoritmo CHAID Exhaustivo tiene la finalidad de no limitar el árbol que se generará a resultados binarios (2 nodos) como lo hace CART o C4.5, esto, ya que uno de los principales objetivos para el futuro utilizando esta herramienta es generar información para las áreas de soporte técnico y de estadística.

La prueba de concepto arroja que el árbol de decisión entrega el mejor resultado en cuanto al acierto de las fallas, el que se traduce a un 80,59%, lo que de acuerdo al requerimiento del negocio es más que suficiente, en comparación al modelo que obtuvo el peor rendimiento, que fue *Support Vector Machine* con *kernel* Sigmoideal con un 50,36%. Lo anterior es ratificado por *F-measure* quién entrega la mejor relación *Precision* versus *Recall* en el árbol de decisión con un 78,93% en comparación al que obtuvo el menor rendimiento con un 50,12% que consecuentemente sigue siendo *Support Vector Machine* con *kernel* Sigmoideal.

También fue realizada una comparación con respecto a función de costo. El resultado de esta prueba indica que el algoritmo con mayor costo es el árbol de decisión versus *Support Vector Machine* con *kernel* Lineal, que es la que genera el menor costo, sin embargo, se debe realizar una interpretación muy cuidadosa debido a que el objetivo de este proyecto es detectar posibles fallas, y lo que indica esta comparación es quién tiene mayor costo en su clasificación, sin embargo, si vemos el detalle, árbol de decisión posee mayor costo debido a que es el algoritmo con más aciertos reales (TN), con respecto a los demás algoritmos, por lo que árbol de decisión CHAID Exhaustivo es la alternativa más conveniente.

Existen muchos algoritmos dentro de la literatura, donde se indican cuáles son los que poseen mejores o peores rendimientos de acuerdo a la aplicación, información y tiempos de procesos. Si bien, cuando se busca la mejor opción de análisis los primeros pasos deben darse desde la literatura, es necesaria la curiosidad e incredulidad de la persona que está realizando el análisis, ya que este tipo de pruebas permite identificar más detalles de la información y su comportamiento, que al final del análisis ayuda a seleccionar la mejor opción. Además, siempre será necesario conocer claramente las intenciones del negocio con respecto a la solución, debido a que dependiendo del requerimiento y la necesidad del negocio se puede discriminar con mayor fundamento una opción de la otra.

Por otro lado, en trabajos futuros, es necesario mejorar el sistema de almacenamiento de información, migrar la información hacia un *data warehouse* mucho más robusto, orientado al concepto de *bigdata* (por el volumen de información y velocidad de almacenamiento), esto permitirá que los tiempos y el manejo de información tengan menos demanda de recursos computacionales y permita mayor cantidad de servicios en este aspecto junto con la consolidación de información en bases de datos relacionales, además, el *holding* tiene como tarea pendiente la definición de un *software* estadístico que pueda explotar toda la información existente dentro del principal sistema ERP que tiene la empresa que es SAP R3, junto con su herramienta de inteligencia de negocios y más aún, explotar el potente servicio de bases de datos en tiempo real que es SAP HANA.

12.4 Evaluación económica

La inversión del proyecto con respecto al retorno puede considerarse bajo, esto, ya que bajo los supuestos establecidos, solamente conseguir evitar una falla catastrófica permitirá recuperar USD 57.000, lo anterior, bajo el supuesto que se evita sólo una falla catastrófica en un periodo de 5 años, lo cual no debería ocurrir, ya que a medida que transcurre el tiempo las personas se acostumbrarán al uso del sistema y aprenderán a obtener mucho más provecho que en el periodo de puesta en marcha,

logrando disminución de costos que serán marginales en el corto plazo, siendo fundamental el periodo de aprendizaje del sistema para aprovechar todo su potencial, por otro lado, se estima que el aumento en la flota por mayor productividad será de un 0,05%, siendo este aspecto un aumento de carácter marginal.

El proyecto debe al menos evitar una falla catastrófica al año, donde el promedio anual de todas las faenas es 31 fallas catastróficas, de esta forma, el proyecto recuperará la inversión en el corto plazo, por lo que la medición de los efectos del proyecto debe ser anual.

La implementación de este proyecto implica la compra de activo fijo que estará disponible para otras aplicaciones de similares características, siendo este el paso inicial para sacar mayor provecho de la inversión. Se deberá aprovechar de la mejor forma posible el conocimiento de los consultores y de las personas que se encuentren a cargo del proyecto ya que serán ellos los referentes para tipo de desarrollo en la empresa, convirtiéndose en este conocimiento en el activo de mayor valor y riesgo para la generación de ventaja competitiva.

12.5 Conclusión general

De acuerdo a los resultados, se establece que el proyecto es factible de realizar, si bien, la inversión inicial es alta, principalmente generado por la adquisición de infraestructura, software y desarrollo del proyecto, el área del *Mining Group* posee al personal competente capaz de continuar desarrollando el proyecto, extendiendo el alcance a los otros sistemas de información que no han sido abarcados en este trabajo de memoria, como fue mencionado en el punto 0.

Del punto 12.3 es necesario destacar la importancia del resultado. Se determinó que el mejor algoritmo para pronosticar que equipos podrían fallar era Árbol de Decisión CHAID Exhaustivo, y esto se da principalmente ya que este algoritmo, a diferencia de otros árboles de decisión no entrega un árbol binario, aumentando la complejidad y

ramificación del árbol de decisión, sin embargo, cumple 2 objetivos principales, posee el mejor rendimiento en su predicción y además, entrega información de acuerdo al motivo de la falla (nodos), lo que es muy importante para que el negocio logre identificar y adquirir, a medida que utilizan este sistema, conocimiento acerca de la falla de los motores que antes no poseía, sin embargo, esto no quiere decir que Árbol de Decisión sea el algoritmo que siempre deberá ser utilizado, más bien es el algoritmo base, el cual servirá de apoyo cuando en el futuro, este modelo matemático se vuelva obsoleto y sea necesario cambiarlo por un *Support Vector Machine* o *Red Neuronal* los cuales son mucho más ambiguos en cuanto a sus criterios de pronóstico.

Es importante mencionar que así como esta metodología fue utilizada en una faena y en un sistema del motor, de acuerdo a la literatura, es factible extrapolar la misma forma de análisis a las demás faenas y sistemas, sin embargo, para lograr el siguiente paso, se deben solucionar problemas actuales de carácter técnico relacionado al acceso a la *data*, ya que el sistema como se encuentra operando actualmente, no tiene la capacidad para abastecer de forma óptima y segura de información para automatizar el proceso de manera confiable, por lo que debe ser considerado como prioridad en Cummins este aspecto, y es por lo mismo, que este proyecto de memoria no se logró implementar, ya que básicamente el *Mining Group* no tiene capacidad de reacción mientras no sea solucionado el problema de gestión de información en su *data warehouse*.

Importante será el trabajo de las personas que se encuentren en constante uso de la herramienta, ya que serán ellos quienes podrán mejorar el sistema y adaptarlo a nuevas necesidades que por naturaleza tendrán que surgir. Con el uso del sistema, surgirán necesidades de información con mayor profundidad de información y exactitud, por lo que siempre se deberá mejorar el diseño del actual proyecto.

De la literatura revisada para efectos de este proyectos, todas coincidieron que la mejor forma de obtener resultados es en base a los parámetros de los sensores del motor, pero además, de la medición otro tipo de variable, como por ejemplo, análisis

de aceites. Si bien Distribuidora Cummins Chile posee análisis de aceite, este análisis no es estándar por lo que no pudo ser agregado a este modelo de predicción. Para el futuro es importante estandarizar este tipo de análisis ya que el aceite en un motor es clave cuando estos comienzan a fallar, que permite deducir mucha información relacionada al origen del mismo.

Relacionado a lo anterior, dentro del *holding* existe otra empresa que está desarrollando análisis de aceite *online* donde se encuentran en proceso de creación del *hardware* y *software* necesario para el cálculo y recepción de esta información vía *online*, tal como lo hace el sistema de monitoreo Specto.

Sin lugar a dudas este proyecto marcará precedente dentro de Distribuidora Cummins Chile y dentro del Holding Komatsu Cummins Chile, debido a que es un nuevo enfoque en el aspecto de entrega de servicios que puede generar, utilizando la misma lógica de análisis reducción de costos dentro de muchos procesos productivos relacionados entre sí, y donde actualmente la industria de la minería tiene puesto su foco: reducción de costos y eficiencia en sus procesos productivos.

13 Bibliografía

1. Anguita, P. (2011). *Planificación estratégica y gestión de cartera de proyectos en clínica alemana*. Santiago, Chile.
2. Aydogan, H. (2014). *Performance, emission and combustion characteristics of bioethanol-biodiesel-diesel fuel blends used in a common rail diesel engine*. Turquía.
3. Barrientos, F., & Ríos, S. (2013). Aplicación de Minería de Datos para Predecir Fuga de Clientes en la Industria de las Telecomunicaciones. *Revista Ingeniería de Sistemas*, 73-107.
4. Barros, O. (2014). *Ingeniería de Negocios Diseño Integrado de Servicios, sus procesos y apoyo TI. Master in Business Engineering*. Santiago, Chile.
5. Basurkoa, O., & Uriondo, Z. (2015). *Condition-Based Maintenance for medium speed diesel engines used in vessels in operation*. España.
6. BN Americas. (2014). *¿Cómo seguirá creciendo la industria minera después del súper ciclo?*. Santiago.
7. BN Americas. (2014). *Outlook 2015: Construyendo confianza en tiempos difíciles*. Santiago.
8. Bouza, C., & Santiago, A. (2014). *La minería de datos: Árboles de decisión y su aplicación en estudios médicos*. México.
9. Burgueño, J. (2013). *Mejora en eficiencia de campañas de migración de prepago a suscripción por medio del estudio del comportamiento de clientes en Entel*. Santiago.
10. Consejo Minero. (2014). *Reporte Anual*. Santiago.

11. Consejo Minero. (2015). *Minería en Cifras*. Santiago.
12. Corredor, L., Escobar, A., Portnoy, I., & Vélez, E. (2012). *Prediction of specific fuel consumption in turbocharged diesel engines under partial load performance*. España.
13. Domm, M. (2003). *Mathematical programming in data mining: Models for binary classification with application to collusion detection in online gambling*. Tucson.
14. Escobar, M. (1992). *El análisis de segmentación: concepto y aplicaciones*. Madrid.
15. Escobar, M. (1998). *Las aplicaciones del análisis de segmentación*. Salamanca.
16. Escobar, M. (1998). *Las aplicaciones del análisis de segmentación: El procedimiento CHAID*. Salamanca.
17. Espinosa, F. (2008). *Auditoria para la efectividad del mantenimiento*. Curicó.
18. Fayyad, U., & Stolorz, P. (1997). *Data Mining and KDD: Promise and Challenges*. Pasadena.
19. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *Association for the Advancement of Artificial Intelligence*, 17(3), 37-54.
20. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *Knowledge Discovery and Data Mining: Towards a Unifying Framework*. Estados Unidos.
21. Frawley, W., Piatetsky-Shapiro, G., & Matheus, C. (1992). Knowledge Discovery in Databases: An Overview. *Association for the Advancement of Artificial Intelligence*, 58-70.
22. Fygueroa, S. (2005). *Diesel Condition Monitoring Techniques*. Pamplona.

23. García Garrido, S. (2003). *Organización y Gestión Integral de mantenimiento*. España.
24. Garg, A., Diwan, P., & Saxena, M. (2014). *Artificial Neural Networks for Internal Combustion Engine Performance and Emission Analysis*. India.
25. Gibert, K., & Cortés, U. (1998). Clustering based on rules and Knowledge Discovery in ill-structured domains. *Computación y Sistemas*, 1(4), 213-227.
26. Hidalgo, J. (2013). Impacto de la tecnología en el mantenimiento. *Termogram*, 1-7.
27. Hung, A. (2008). *Mantenimiento Centrado en Confiabilidad como estrategia para apoyar los indicadores de disponibilidad y paradas forzadas en la Planta Oscar A. Machado*. La Havana.
28. Imamori, Y., Hiraoka, K., Endo, H., & Oda, Y. (2011). Combustion Simulations Contributing to the Development of Reliable Low-Emission Diesel Engines. *Mitsubishi Heavy Industries Technical Review*, 48(1), 65-69.
29. Jiang, F. (2006). *SVM-Based Negative Data Mining to Binary Classification*. Atlanta.
30. Johnson, M., Christensen, C., & Kagermann, H. (2008). Reinventing Your Business Model. *Harvard Business Review*, 57-68.
31. Kimmich, F., Schwarte, A., & Isermann, R. (2005). *Fault detection for modern Diesel engines using signal- and process model-based methods*. Alemania.
32. Kokkulunk, G., Akdogan, E., & Ayhan, V. (2013). *Prediction of emissions and exhaust temperature for direct injection diesel engine with emulsified fuel using ANN*. Turkey.

33. Kumar, V., & Verma, L. (2010). Binary Classifiers for Health Care Databases: A Comparative Study of Data Mining Classification Algorithms in the Diagnosis of Breast Cancer. *International Journal of Computer Science and Technology*, 124-129.
34. Lamarinis, V., & Hountalas, D. (2010). A general purpose diagnostic technique for marine diesel engines—Application on the main propulsion and auxiliary diesel units of a marine vessel. *Energy conversion and management*, 740-753.
35. Li, L. M., Wang, L. T., & Lup, L. W. (2000). *IntelliClean : A Knowledge-Based Intelligent Data Cleaner*. Singapur.
36. Lincovil, C., & Gutierrez, G. (2006). *Optimización Económica de la Disponibilidad*. Santiago, Chile.
37. Maass, B. D. (2011). *In-Cylinder Pressure Modelling with Artificial Neural Networks*.
38. Maji, S., Berg, A., & Malik, J. (2008). *Classification using Intersection Kernel Support Vector Machines is Efficient*. Berkeley.
39. Marrero, R. (2005). *Implementación de Support Vector Machines para reconocimiento de segmentos de la población*. Sartenejas.
40. Marrero, R. (2005). *Implementación de Support Vector Machines para reconocimientos de segmentos de la población*. Venezuela.
41. Mascarell, E. (2007). *Reliability Centered Maintenance*. Barcelona.
42. Mevlut, T., Fusun, T., & Imran, K. (2008). *Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients*. Turkey.
43. Molina, J. (2001). *Mantenimiento Industrial*. Argentina.

44. Monge, J., & Pérez, A. (2002). *Estadística no paramétrica: Prueba de Chi-cuadrado*. España.
45. Odei, S. (2006). *An Exploration of Classification Prediction Techniques in Data Mining: The insurance domain*. Inglaterra.
46. Olarte, W., Botero, M., & Cañon, B. (2010). *Técnicas de mantenimiento predictivo utilizadas en la industrial*. Colombia.
47. Oyarce, H., & Cifuentes, J. (2012). Factibilidad de Servicios de Optimización de Activos de Producción utilizando Tecnología de Diagnóstico Predictivo en Elementos Finales de Control. *Gestión de las personas y tecnología*, 118-131.
48. Parlak, A., Islamoglu, Y., Yasar, H., & Egrisogut, A. (2005). *Application of artificial neural network to predict specific fuel consumption and exhaust temperature for a Diesel engine*. Turkey.
49. Pérez, A., Kisys, R., & Manzanedo, L. (2002). *Regresión Logística Binaria*. España.
50. Pérez, C., & Moubray, J. (2004). *Mantenimiento Centrado en Confiabilidad*. Reino Unido.
51. Piñol, A., Ortega, J., & Romeral, J. (2004). *Mantenimiento Predictivo para motores de inducción*. Terrassa.
52. Prando, R. (1996). *Manual de Gestión de Mantenimiento a la medida*. Guatemala.
53. Quezada, A. (2013). *Diseño y construcción del proceso de priorización de pacientes*. Santiago, Chile.
54. Quiroz, E. (2011). *Diseño del proceso de análisis de pacientes para patologías crónicas en clínica Las Condes*. Santiago.

55. Reveco, C. (2011). *Pronóstico y análisis de demanda de la sala de urgencia del hospital Luis Calvo Mackenna y metodología para el cálculo de recursos críticos*. Santiago.
56. Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Inteligencia Artificial*, 11-18.
57. Roche, A. (2009). *Árboles de decisión y Series de tiempo*. Uruguay.
58. Ron, R., & Ron, M. (1999). La combinación del TPM y RCM. Michigan, Diesel Technology Company. *Gestión de Activos Industriales*, 49-56.
59. Salcedo, C. (2004). *Estimación de la ocurrencia de incidencias en declaraciones de pólizas de importación*. Lima.
60. Sanz Arazuri, E., & Ponce de León Elizondo, A. (2010). Claves en la aplicación del algoritmo CHAID. Un estudio del ocio físico deportivo universitario. *Psicología del Deporte*, 319-333.
61. Sorbie, J. (2014). *A Review of Marine Engine Fault Detection Techniques*. Inglaterra.
62. Srivastava, D. K., & Bhambhu, L. (2010). *Data classification using Support Vector Machine*. India.
63. Tan, P.-Q., Deng, K., & Lu, J.-X. (2004). *Analysis of particulate matter composition from a heavy-duty diesel engine*. China.
64. Tang, T.-I., Zheng, G., Huang, Y., Shu, G., & Wang, P. (2005). *A Comparative Study of Medical Data Classification Methods Based on Decision Tree and System Reconstruction Analysis*. Taiwan.
65. Valcárcel, V. (2004). Data mining y el descubrimiento del conocimiento. *Revista de la facultad de ingeniería industrial UNMSM*, 7, 83-86.

66. Zhixiong, L. (2012). Intelligent fault diagnosis method for marine diesel engines using instantaneous angular speed. *Journal of Mechanical Science and Technology*, 2413-2423.

14 Anexos

14.1 BPMN Determinar motores para mantenimiento

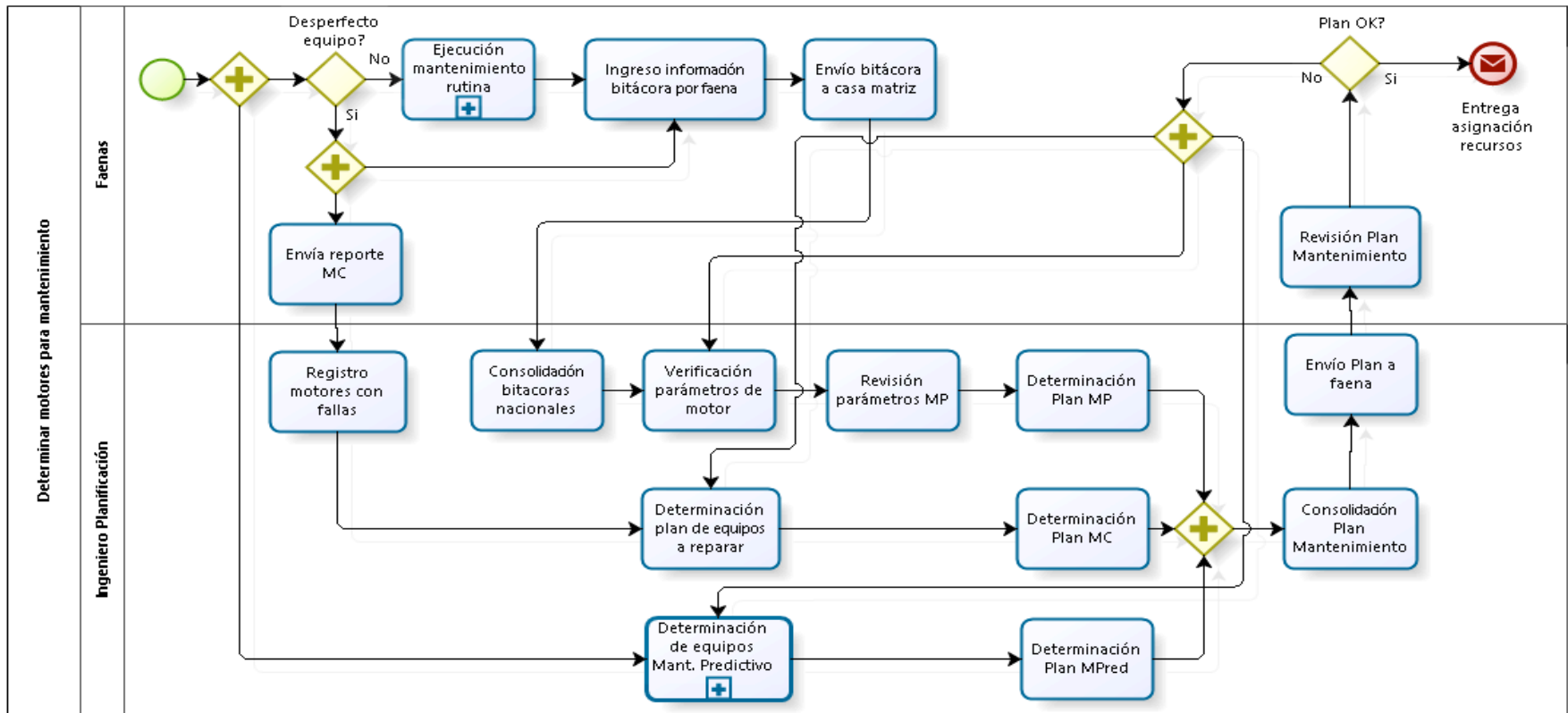


Figura 14.1: Proceso Determinar motores para mantenimiento. Fuente: Elaboración propia.

Para mayor información ver 14.2 BPMN Determinación de equipos Mantenimiento Pred y 14.4 BPMN Determinar motores para mantenimiento (actual).

14.2 BPMN Determinación de equipos Mantenimiento Predictivo

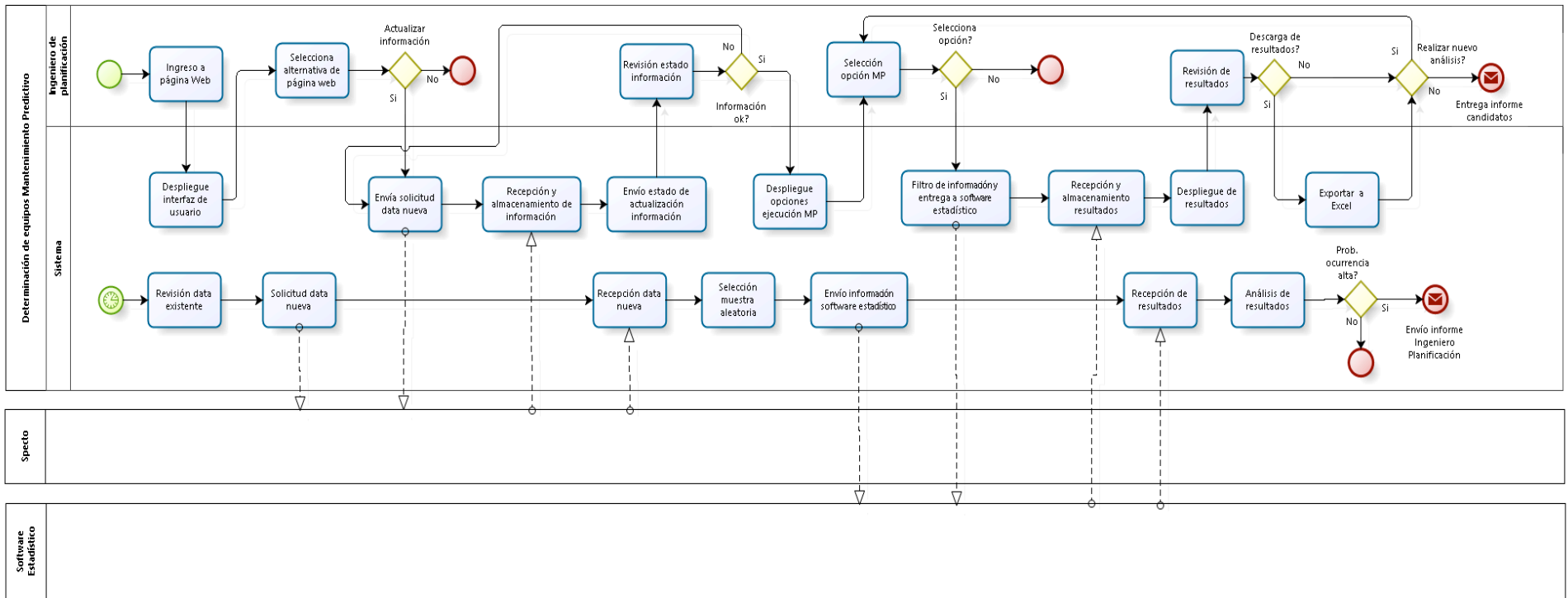


Figura 14.2: Proceso determinación de equipos Mantenimiento Predictivo. Fuente: Elaboración propia.

Para más información con respecto a este proceso puede ser revisado en página 106, 6.5.1.1 Determinación de equipos Mantenimiento Predictivo.

14.3 BPMN Controlar Mantenimiento

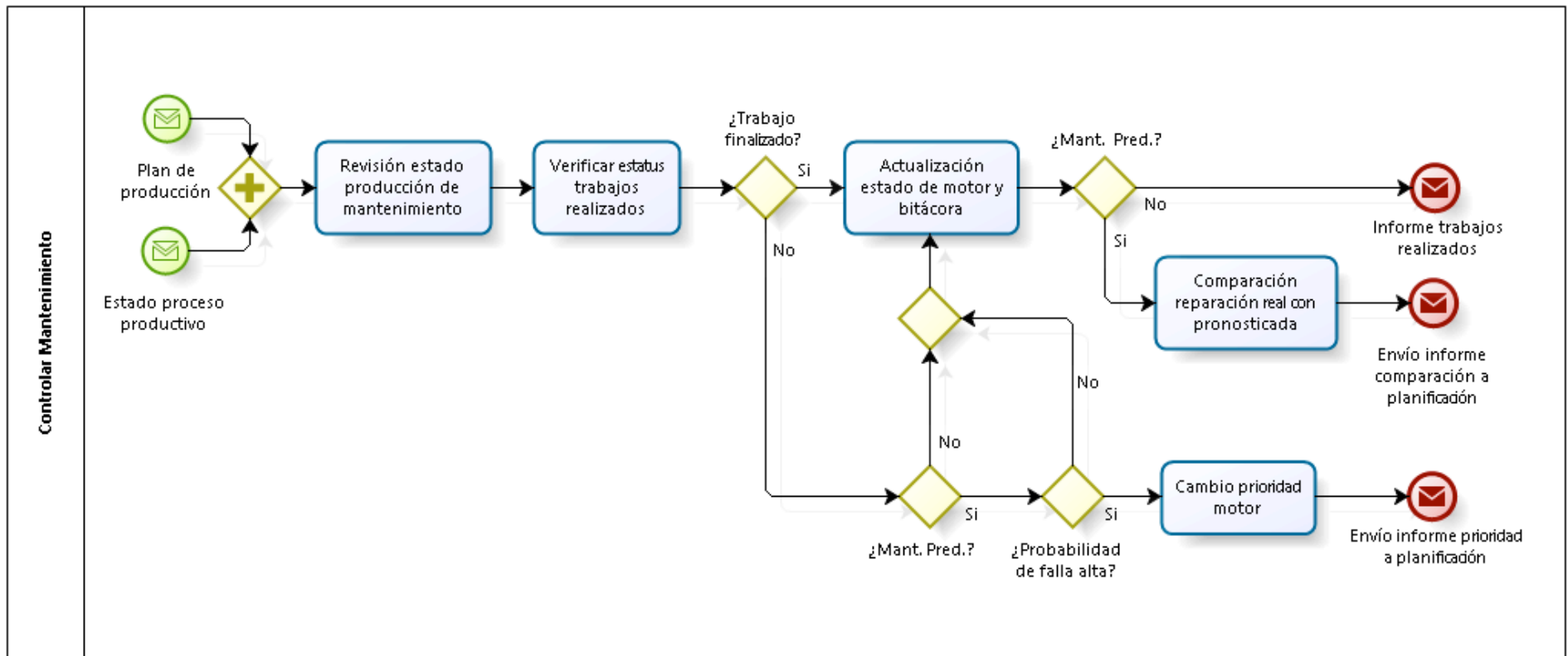


Figura 14.3: Proceso Controlar mantenimiento. Fuente: Elaboración propia.

Como prueba de concepto, este proyecto se enfoca en la exposición y definición de los distintos algoritmos de minería de datos que permiten analizar la información, sin embargo, se expone un proceso de control que permitirá a los distintos analistas medir la efectividad de los modelos matemáticos, y en caso de que sea requerido, ajustar dichos modelos.

- Revisión estado producción de mantenimiento

Se debe reunir la información del plan de mantenimiento con la producción real, esto debe ser centralizado para un resultado óptimo.

- Verificación de estatus trabajos realizados

La importancia de este punto es que son verificados los estatus iniciales donde se indica a priori que el motor debe ser intervenido versus la clasificación del mecánico en terreno de la real necesidad de mantenimiento del motor.

Si el trabajo ha sido finalizado:

- Actualización estado de motor y bitácora

Debe ser actualizado el estado de motor en la base de datos del sistema, junto con actualizar la bitácora del motor, ya sea electrónica o física.

Si es una reparación pronosticada por mantenimiento predictivo:

- Comparación reparación real con pronosticada

Es importante esta comparación ya que se busca establecer que las predicciones sean acertadas, de lo contrario el modelo matemático debe ser ajustado, donde esta información es enviada a planificación para que realice las acciones necesarias.

Si es una reparación correctiva:

- Se envía informe de trabajos realizados.

Si el trabajo no ha sido finalizado:

- Si es una reparación pronosticada por mantenimiento predictivo:

- Si la probabilidad de falla es alta:

- Cambio prioridad motor

Esta tarea es importante debido a que si las personas que planifican las reparaciones no conocen la probabilidad de falla, pueden no considerar equipos críticos, finalizando con el envío del informe a planificación.

- Si la probabilidad de falla es baja:

- Se actualiza estado de motor y bitácora

- Si no es una reparación pronosticada por mantenimiento predictivo:

- Se actualiza estado de motor y bitácora

14.4 BPMN Determinar motores para mantenimiento (actual)

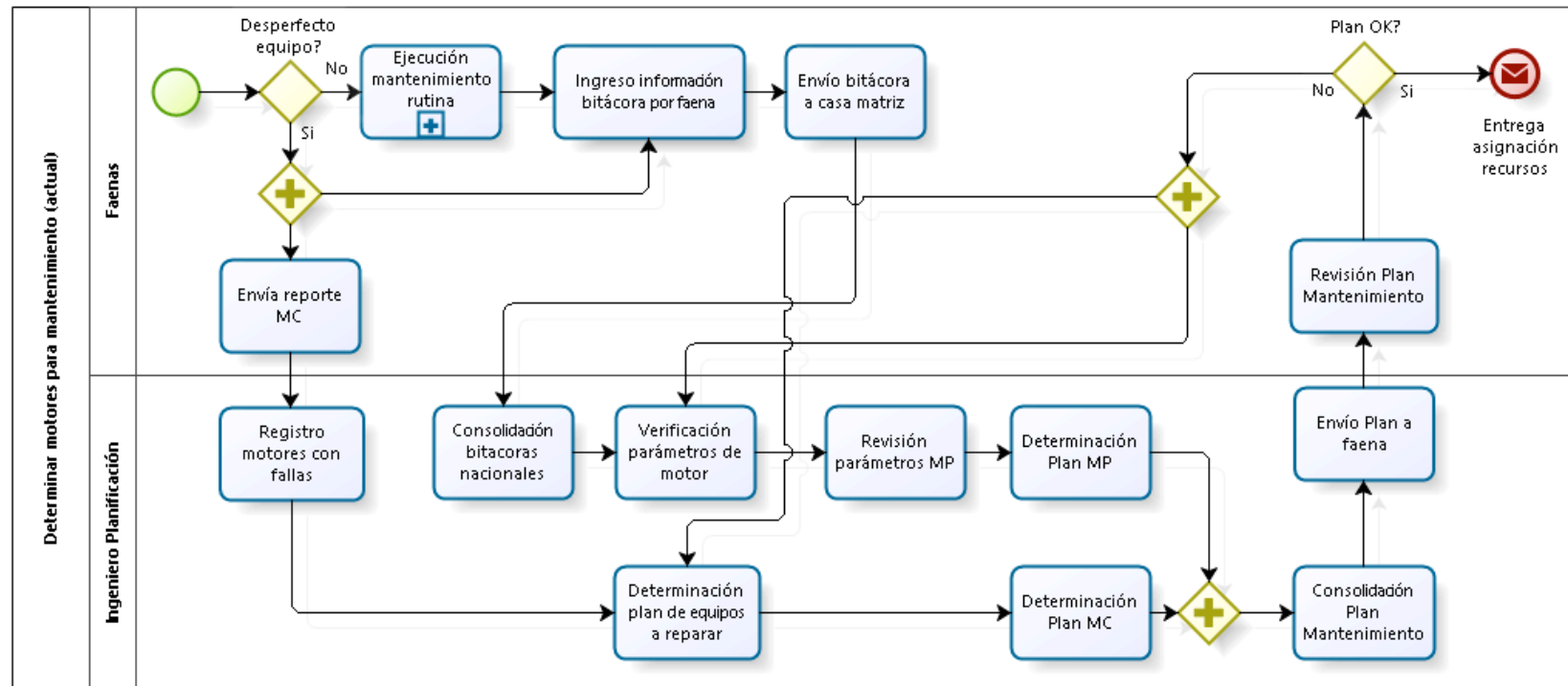


Figura 14.4: Proceso determinar motores para mantenimiento (actual). Fuente: Elaboración propia.

Para mayor información, dirirse a sección 6.4 Situación Actual en página 103.

14.5 Selección campos a considerar en base de datos

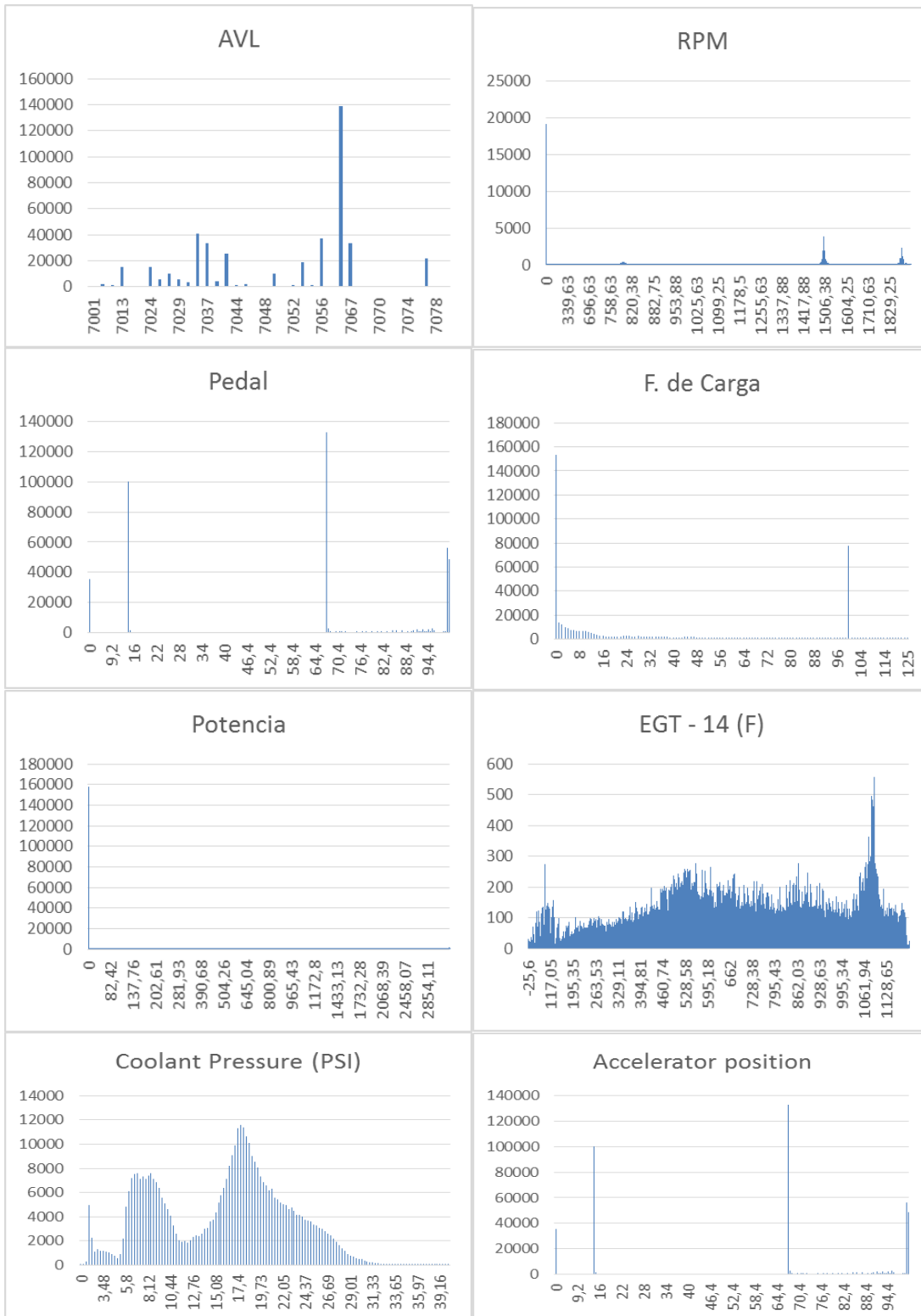
N°	Campo	Cuenta registros	%	Estado
1	Faena	437.042	100,00%	Mantener
2	Flota	437.042	100,00%	Mantener
3	ESN	437.042	100,00%	Mantener
4	AVL	437.042	100,00%	Mantener
5	Vehiculo	437.042	100,00%	Mantener
6	Fecha_Servidor	437.042	100,00%	Mantener
7	Fecha_AVL	437.042	100,00%	Mantener
8	RPM	437.042	100,00%	Mantener
9	Pedal	437.042	100,00%	Mantener
10	F_de_Carga	437.042	100,00%	Mantener
11	Potencia	437.042	100,00%	Mantener
12	Estado_Motor	431.232	98,67%	Eliminar
13	Engine_Torque_Mode	416.918	95,40%	Eliminar
14	Actual_Percent_Torque_	437.042	100,00%	Mantener
15	Actual_SpeedRPM	437.042	100,00%	Mantener
16	Accelerator_position	437.042	100,00%	Mantener
17	Percent_Load_At_Current_Speed	437.042	100,00%	Mantener
18	Fuel_rateLh	437.042	100,00%	Mantener
19	Barometric_PressurePSI	437.042	100,00%	Mantener
20	IMP-LBPSI	437.042	100,00%	Mantener
21	IMT-LBFF	437.042	100,00%	Mantener
22	EGT-AVF	437.042	100,00%	Mantener
23	Battery_potentialV	437.042	100,00%	Mantener
24	Electrical_potential_V	0	0,00%	Eliminar
25	Coolant_temperatureF	437.042	100,00%	Mantener
26	Fuel_TemperatureF	437.042	100,00%	Mantener
27	Engine_Oil_TemperatureF	437.042	100,00%	Mantener
28	Injector_MeteringPSI	437.042	100,00%	Mantener
29	Injector_TimingPSI	437.042	100,00%	Mantener
30	Pre-filter_Oil_PressurePSI	421.592	96,46%	Eliminar
31	Instantaneous_Estimated_Brake_PowerF	0	0,00%	Eliminar
32	IMT-LBRF	437.042	100,00%	Mantener
33	IMT-RBFF	437.042	100,00%	Mantener
34	IMT-RBRF	437.042	100,00%	Mantener
35	IMP-RBPSI	268.114	61,35%	Eliminar
36	IMP_PSI	0	0,00%	Eliminar
37	Turbocharger_2_Boost_PressurePSI	0	0,00%	Eliminar
38	Oil_Differential_PressurePSI	422.254	96,62%	Eliminar
39	Ecu_temperatureF	0	0,00%	Eliminar

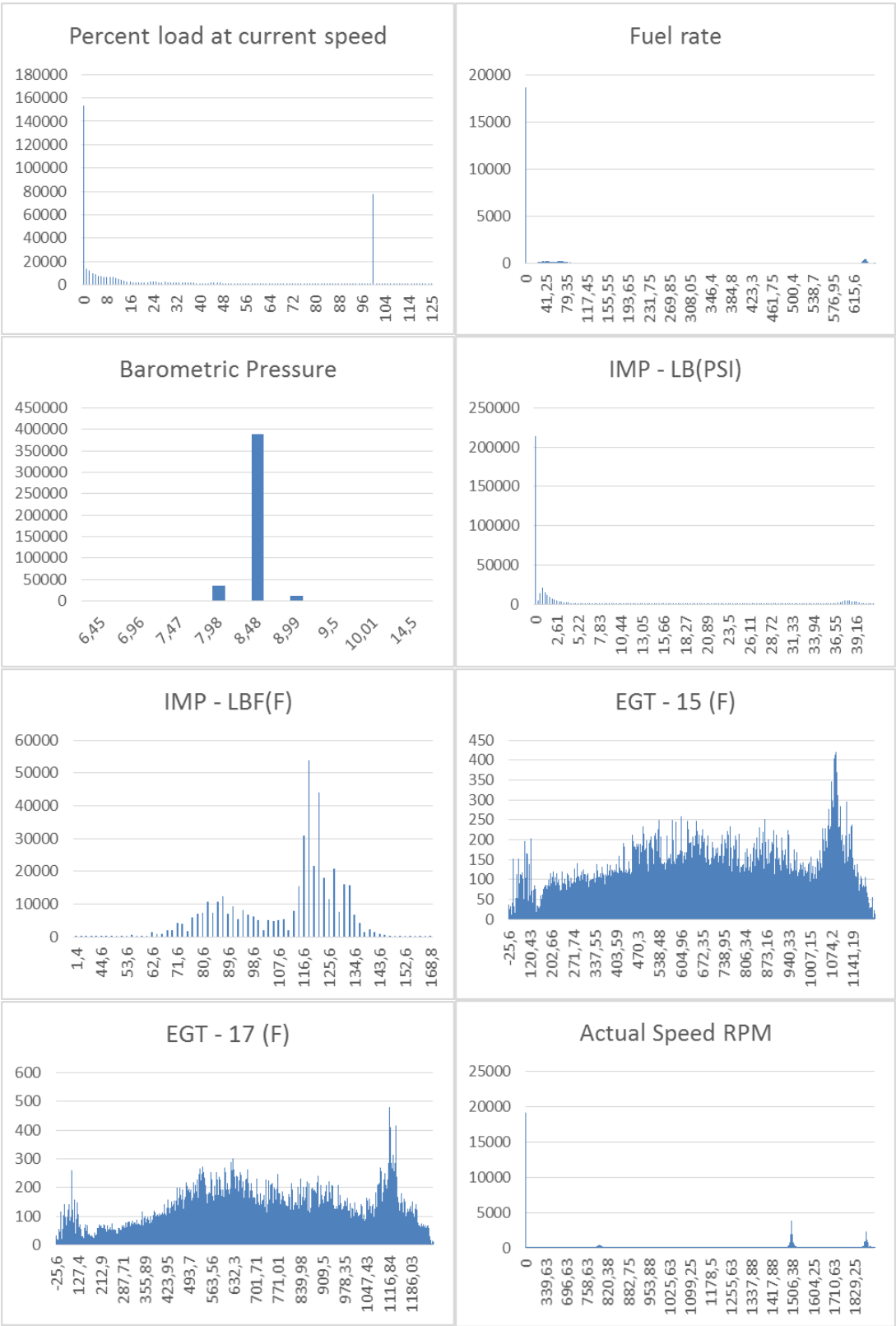
Nº	Campo	Cuenta registros	%	Estado
40	EGT-01F	437.042	100,00%	Mantener
41	EGT-02F	437.042	100,00%	Mantener
42	EGT-03F	437.042	100,00%	Mantener
43	EGT-04F	437.042	100,00%	Mantener
44	EGT-05F	427.788	97,88%	Eliminar
45	EGT-06F	437.042	100,00%	Mantener
46	EGT-07F	437.042	100,00%	Mantener
47	EGT-08F	437.042	100,00%	Mantener
48	EGT-09F	415.415	95,05%	Eliminar
49	EGT-10F	437.042	100,00%	Mantener
50	EGT-11F	437.042	100,00%	Mantener
51	EGT-12F	437.042	100,00%	Mantener
52	EGT-13F	437.042	100,00%	Mantener
53	EGT-14F	437.042	100,00%	Mantener
54	EGT-15F	437.042	100,00%	Mantener
55	EGT-16F	437.042	100,00%	Mantener
56	Crankcase_PressureHPIin-H2O	437.042	100,00%	Mantener
57	Engine_Oil_Level	437.042	100,00%	Mantener
58	Post_Oil_FilterPSI	437.042	100,00%	Mantener
59	Fuel_Delivery_PressurePSI	0	0,00%	Eliminar
60	Crankcase_PressureMCRSin-H2O	0	0,00%	Eliminar
61	Rifle_Oil_PressurePSI	437.042	100,00%	Mantener
62	Coolant_PressurePSI	437.042	100,00%	Mantener
63	Ambient_TemperatureF	388.123	88,81%	Eliminar
64	EGT-17F	437.042	100,00%	Mantener
65	EGT-18F	404.307	92,51%	Eliminar
66	Turbocharger_SpeedRPM	0	0,00%	Eliminar
67	EGT-RBMCRSF	0	0,00%	Eliminar
68	EGT-LBMCRSF	0	0,00%	Eliminar
69	Oil_Level-Reserve_Tank	0	0,00%	Eliminar
70	Fuel_Supply_Pump_Inlet_PressureMCRSPSI	0	0,00%	Eliminar
71	Post_Oil_FilterMCRSPSI	0	0,00%	Eliminar
72	IMT-LBMF	0	0,00%	Eliminar
73	IMT-RBMF	0	0,00%	Eliminar
74	PowerHP	437.042	100,00%	Mantener
75	Crankcase_PressureHPIin-H2O2	0	0,00%	Eliminar
76	EGT-AVF2	0	0,00%	Eliminar
77	IMP-LBPSI2	0	0,00%	Eliminar
78	IMP-RBPSI2	0	0,00%	Eliminar
79	Oil_Differential_PressurePSI2	0	0,00%	Eliminar
80	Rifle_Oil_PressurePSI2	0	0,00%	Eliminar
81	Target_1	437.042	100,00%	Mantener
82	Id_dato	437.042	100,00%	Mantener

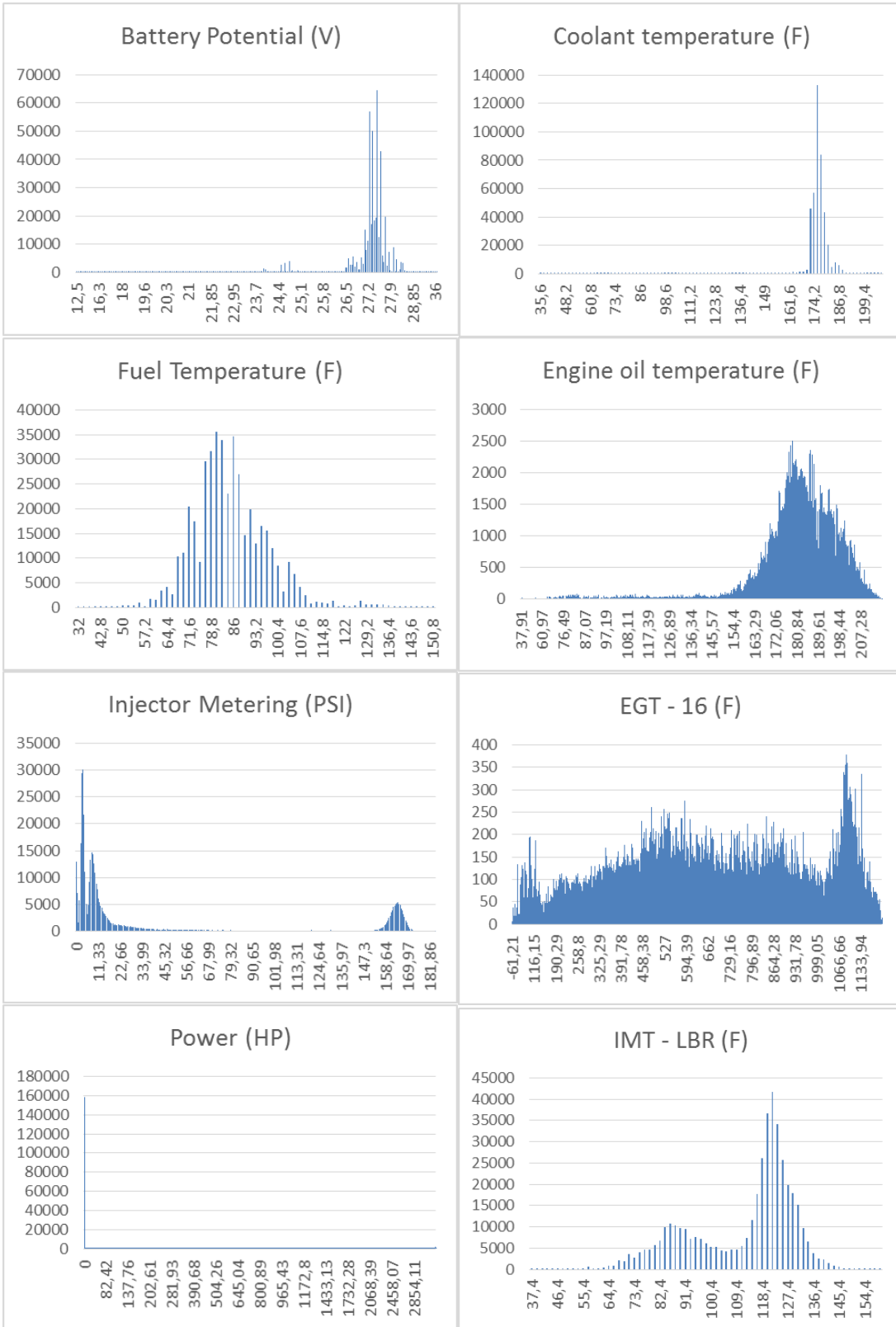
Tabla 14.1: Selección de variables a considerar para análisis. Fuente: Elaboración propia.

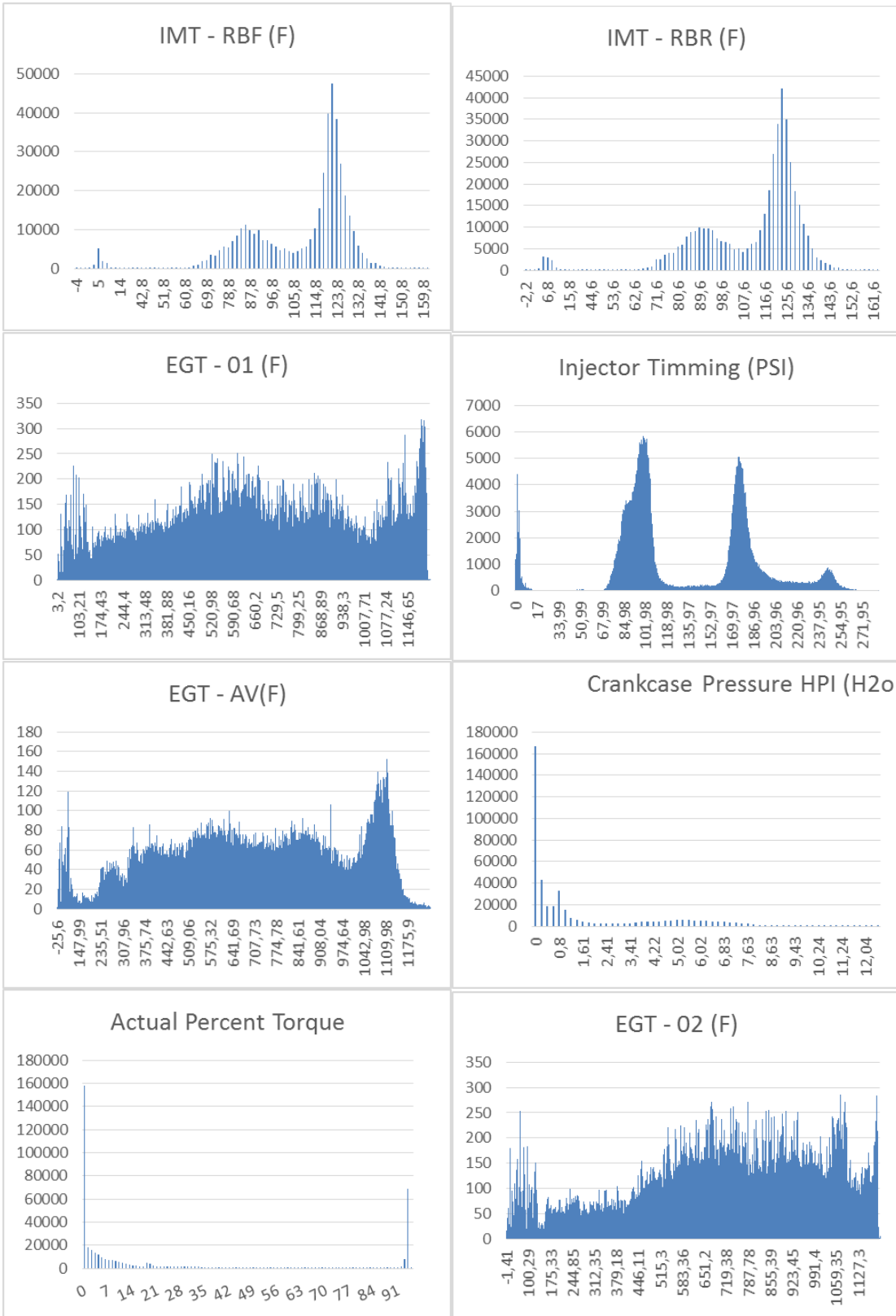
Para obtener mayor información acerca de los criterios de selección, revisar en sección 9.2.4 Análisis exploratorio de la base de datos en página 153.

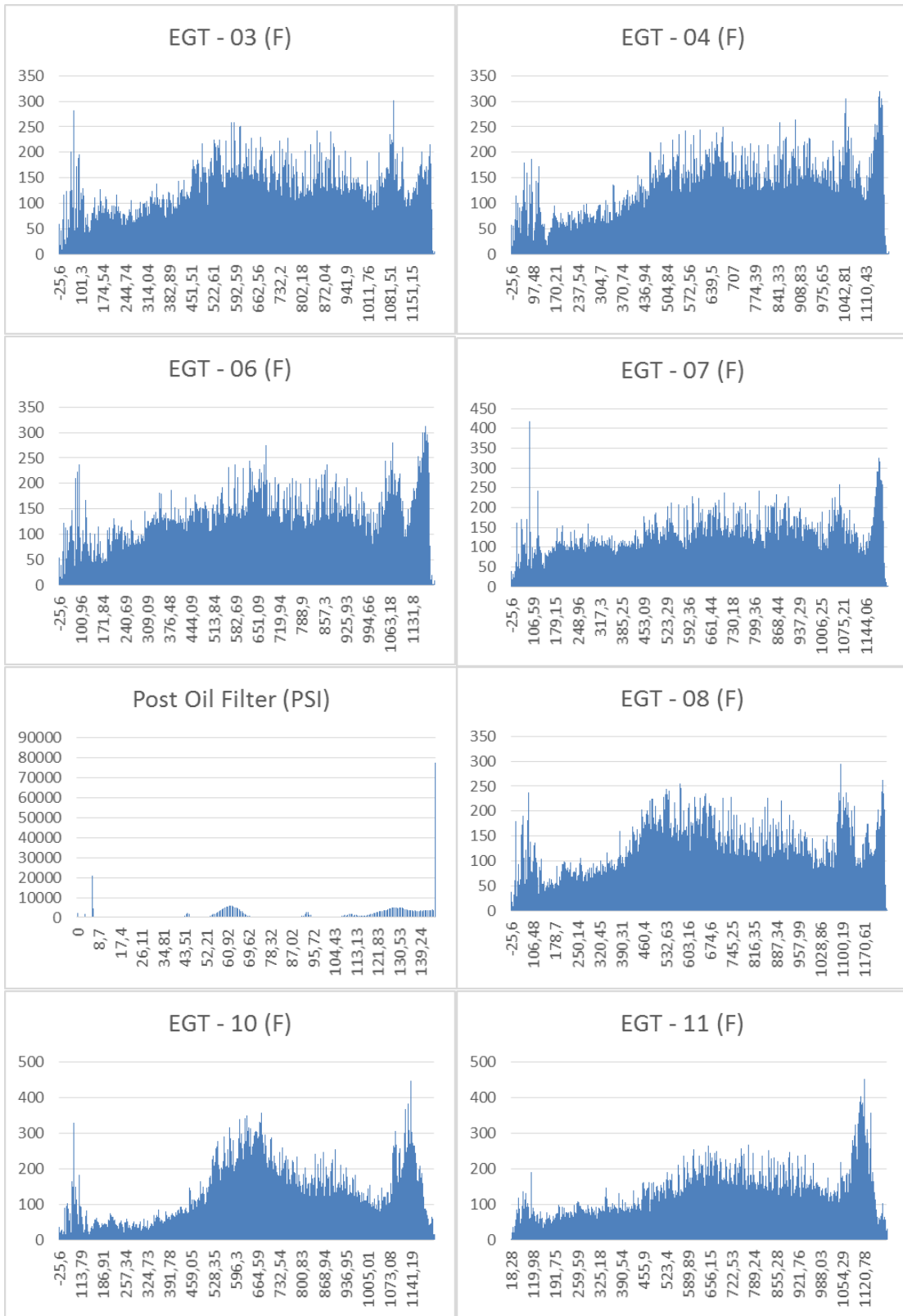
14.6 Análisis gráfico información (Frecuencia valores)

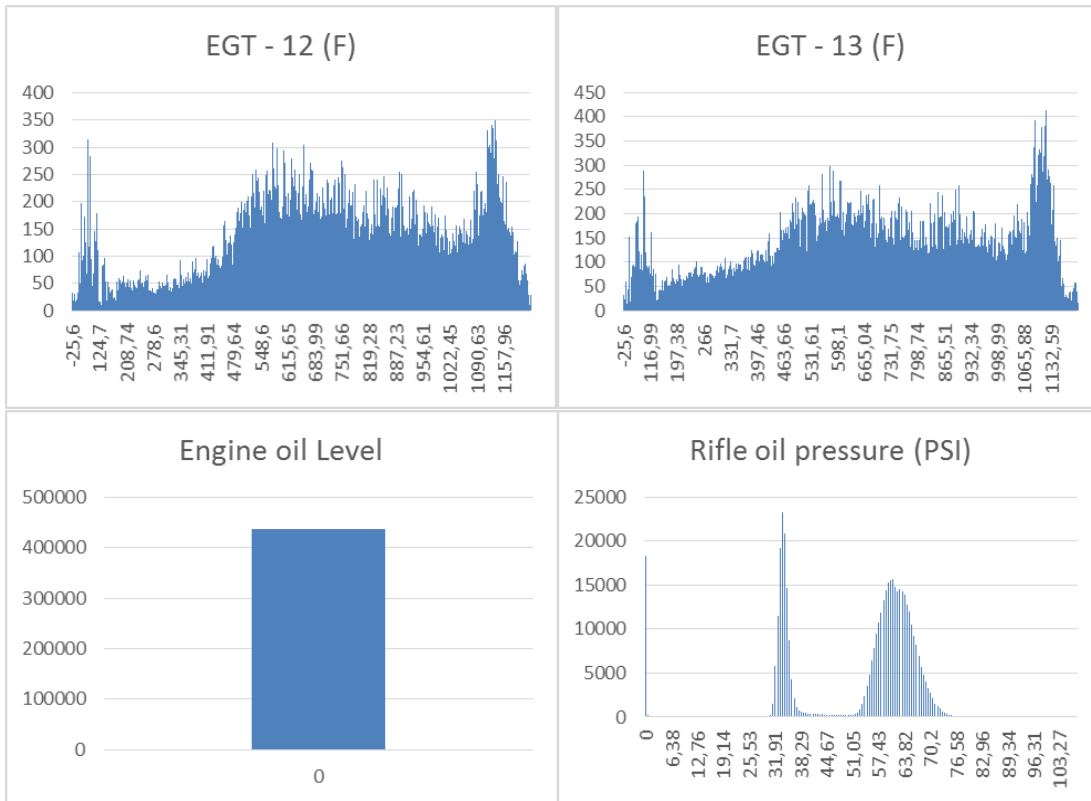












Estos gráficos son utilizados para revisar distribución de los datos dentro de una muestra, además permite discriminar de manera visual cuales son las variables que definitivamente no tienen información relevantes, como por ejemplo la variable “Engine oil Level”. Para mayor información ver sección 9.2.4 Análisis exploratorio de la base de datos en página 153.

14.7 SPSS Statistics input: Árbol de decisión CHAID Exhaustivo

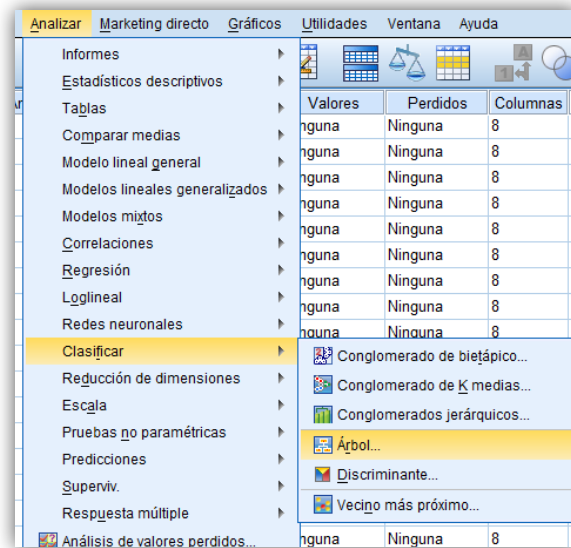
A continuación se muestran los parámetros de entrada que fueron ingresados para la ejecución del árbol de decisión en SPSS Statistics 19.

Paso posterior a la creación de *data set* que será utilizado para el análisis de información, se realizan los siguientes pasos:

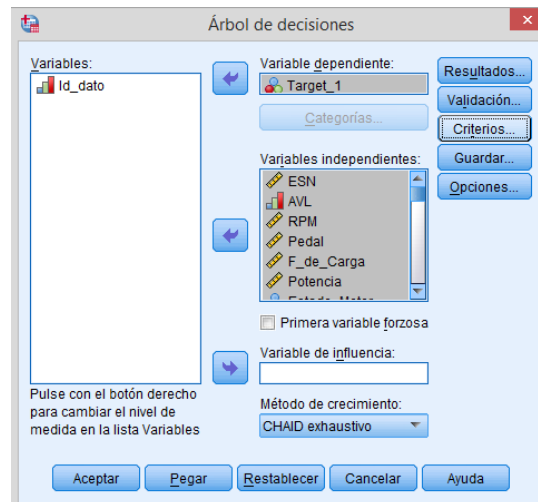
a. Configuración de variables:

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	ESN	Numérico	11	0		Ninguna	Ninguna	11	Derecha	Escala	Entrada
2	AVL	Numérico	11	0		Ninguna	Ninguna	11	Derecha	Ordinal	Entrada
3	RPM	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
4	Pedal	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
5	F_de_Carga	Numérico	11	0		Ninguna	Ninguna	11	Derecha	Escala	Entrada
6	Potencia	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
7	Estado_Motor	Cadena	255	0		Ninguna	Ninguna	50	Izquierda	Nominal	Entrada
8	Actual_Perc...	Numérico	11	0		Ninguna	Ninguna	11	Derecha	Escala	Entrada
9	Percent_Lo...	Numérico	11	0		Ninguna	Ninguna	11	Derecha	Escala	Entrada
10	Fuel_rateLh	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
11	Barometric_...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
12	IMPLBPSI	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
13	IMTLBFF	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
14	EGTAVF	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
15	Battery_pot...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
16	Coolant_te...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
17	Fuel_Tempe...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
18	Engine_Oil_...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
19	Injector_Met...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
20	Injector_Tim...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
21	IMTRBFF	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
22	IMTRBRF	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
23	EGT01F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
24	EGT02F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
25	EGT03F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
26	EGT04F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
27	EGT06F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
28	EGT07F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
29	EGT08F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
30	EGT10F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
31	EGT11F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
32	EGT13F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
33	EGT14F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
34	EGT15F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
35	EGT16F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
36	Crankcase_...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
37	Post_Oil_Fil...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
38	Rifle_Oil_Pr...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
39	Coolant_Pre...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
40	EGT17F	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
41	Target_1	Numérico	11	0		Ninguna	Ninguna	11	Derecha	Nominal	Objetivo
42	Id_dato	Numérico	11	0		Ninguna	Ninguna	11	Derecha	Ordinal	Ninguna

b. Selección de Árbol de decisión:

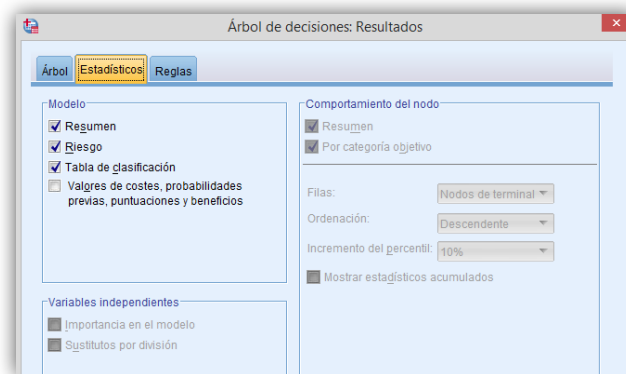
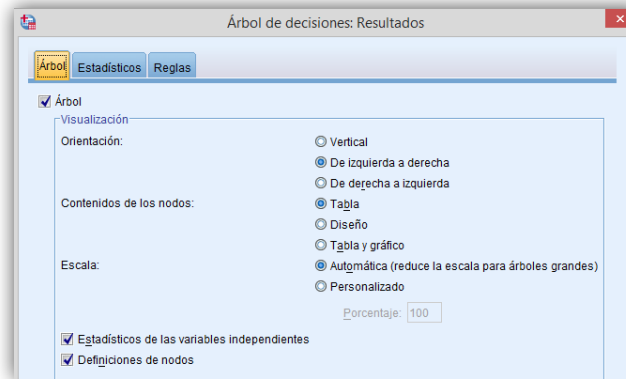


c. Selección de variables y tipo de árbol de decisión:



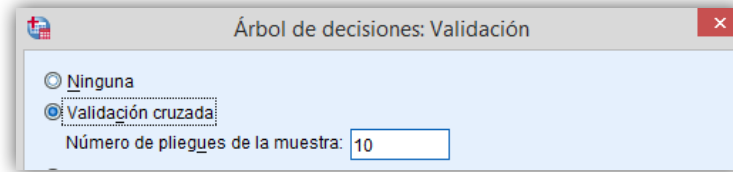
Especificación de variable objetivo (dependiente) y variables independientes.

d. Resultados:



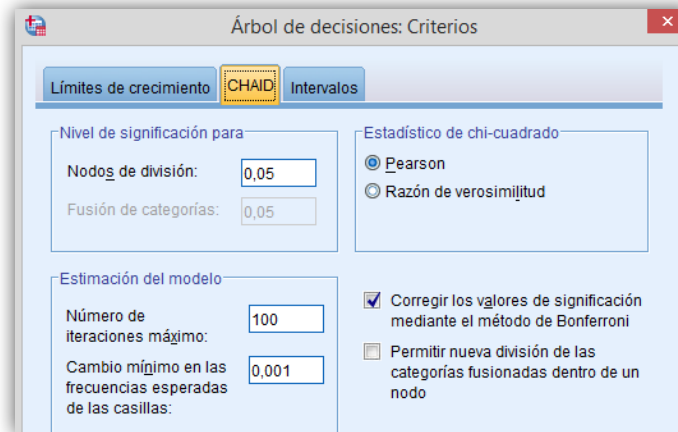
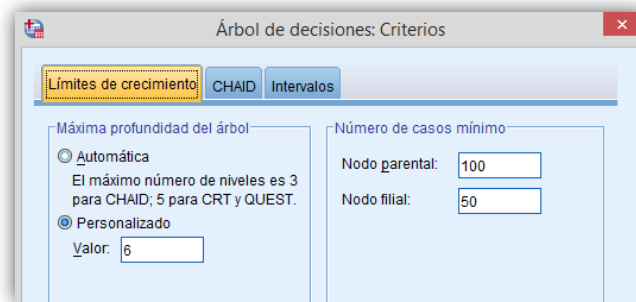
Importante es especificar en los resultados la orientación del árbol de decisión y la generación del archivo con las reglas para clasificación de la información.

e. Validación



Es importante especificar la cantidad de iteraciones para corroborar que no exista sobreajuste y corroborar el nivel de predicción del modelo.

f. Criterios



El criterio de profundidad de nodos está especificado en 6 debido a que el negocio lo solicitó de esa forma.

14.8 SPSS Statistics output: Árbol de decisión CHAID Exhaustivo

* Árbol de decisiones.

```

TREE Target_1 [n] BY Estado_Motor [n] RPM [s] Pedal [s] F_de_Carga [s] Potencia [s] Actual_Percent_Torque_ [s]
Percent_Load_At_Current_Speed [s] Fuel_rateLh [s] Barometric_PressurePSI [s] IMPLBPSI [s] IMILBFF [s] EGTAVF [s]
Battery_potentialV [s] Coolant_temperatureF [s] Fuel_TemperatureF [s] Engine_Oil_TemperatureF [s]
Injector_MeteringPSI [s] Injector_TimingPSI [s] IMTRBFF [s] IMTRBRF [s] EGT01F [s] EGT02F [s] EGT03F [s] EGT04F [s]
EGT06F [s] EGT07F [s] EGT08F [s] EGT10F [s] EGT11F [s] EGT13F[s] EGT14F [s] EGT15F [s] EGT16F [s]
Crankcase_PressureHPIinH2O [s] Engine_Oil_Level [s] Post_Oil_FilterPSI [s] Rifle_Oil_PressurePSI [s]
Coolant_PressurePSI [s] EGT17F [s] AVL [n]

/TREE DISPLAY=LEFTRIGHT NODES=BOTH BRANCHSTATISTICS=YES NODEDEFS=YES SCALE=AUTO
/DEPCATEGORIES USEVALUES={VALID}
/PRINT MODELSUMMARY CLASSIFICATION RISK
/METHOD TYPE=EXHAUSTIVECHAID
/GROWTHLIMIT MAXDEPTH=6 MINPARENTSIZE=100 MINCHILDSIZE=50
/VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
/CHAID ALPHASPLIT=0.05 SPLITMERGED=NO CHISQUARE=PEARSON CONVERGE=0.001 MAXITERATIONS=100 ADJUST=BONFERRONI
INTERVALS=10
/COSTS EQUAL
/MISSING NOMINALMISSING=MISSING.
    
```

Árbol de clasificación

Notas		26-feb-2016 19:26:54
Resultados creados		
Comentarios		
Entrada	Datos	C:\Users\Matías\Desktop\Data Tesis - MBE\Análisis exploratorio datos SPSS ultima base\Base datos random 38 variables con estado de motor.sav
	Conjunto de datos activo	Conjunto_de_datos1
	Filtro	<ninguno>
	Peso	<ninguno>
	Segmentar archivo	<ninguno>
	Núm. de filas del archivo de trabajo	421677
Tratamiento de los datos perdidos	Definición de perdidos	El tratamiento de los valores perdidos definidos por el usuario de las variables independientes nominales depende del método de crecimiento.
	Casos utilizados	Solamente se utilizarán los casos con valores válidos para la variable dependiente y todas o alguna variable independiente en el cálculo de los estadísticos.

Sintaxis	TREE Target_1 [n] BY Estado_Motor [n] RPM [s] Pedal [s] F_de_Carga [s] Potencia [s] Actual_Percent_Torque_ [s] Percent_Load_At_Current_Speed [s] Fuel_rateLh [s] Barometric_PressurePSI [s] IMPLBPSI [s] IMTLBFF [s] EGTAVF [s] Battery_potentialV [s] Coolant_temperatureF [s] Fuel_TemperatureF [s] Engine_Oil_TemperatureF [s] Injector_MeteringPSI [s] Injector_TimingPSI [s] IMTRBFF [s] IMTRBRF [s] EGT01F [s] EGT02F [s] EGT03F [s] EGT04F [s] EGT06F [s] EGT07F [s] EGT08F [s] EGT10F [s] EGT11F [s] EGT13F [s] EGT14F [s] EGT15F [s] EGT16F [s] Crankcase_PressureHPIinH2O [s] Engine_Oil_Level [s] Post_Oil_FilterPSI [s] Rifle_Oil_PressurePSI [s] Coolant_PressurePSI [s] EGT17F [s] AVL [n] /TREE DISPLAY=LEFTTORIGHT NODES=BOTH BRANCHSTATISTICS=YES NODEDEFS=YES SCALE=AUTO /DEPCATEGORIES USEVALUES=[VALID] /PRINT MODELSUMMARY CLASSIFICATION RISK /METHOD TYPE=EXHAUSTIVECHAID /GROWTHLIMIT MAXDEPTH=6 MINPARENTSIZE=100 MINCHILDSize=50 /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES /CHAID ALPHASPLIT=0.05 SPLITMERGED=NO CHISQUARE=PEARSON CONVERGE=0.001 MAXITERATIONS=100 ADJUST=BONFERRONI INTERVALS=10 /COSTS EQUAL /MISSING NOMINALMISSING=MISSING.
----------	---

Recursos	Tiempo de procesador	00 00:10:36,266
	Tiempo transcurrido	00 00:10:46,141
Archivos guardados	Archivo de reglas	

[Conjunto_de_datos1] C:\Users\Matias\Desktop\Data Tesis - MBE\Análisis exploratorio datos SPSS ultima base\Base datos random 38 variables con estado de motor.sav

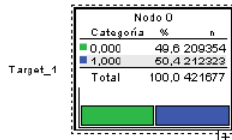
Advertencia

No se muestran las tablas de resúmenes de ganancias ya que no se han definido los beneficios.
No se muestran las tablas de ganancias de las categorías objetivo ya que no se han definido las categorías objetivo.

Resumen del modelo

Especificaciones	Método de crecimiento	CHAID exhaustivo
	Variable dependiente	Target_1
	Variables independientes	Estado_Motor, RPM, Pedal, F_de_Carga, Potencia, Actual_Percent_Torque_, Percent_Load_At_Current_Speed, Fuel_rateLh, Barometric_PressurePSI, IMPLBPSI, IMTLBFF, EGTAVF, Battery_potentialV, Coolant_temperatureF, Fuel_TemperatureF, Engine_Oil_TemperatureF, Injector_MeteringPSI, Injector_TimingPSI, IMTRBFF, IMTRBRF, EGT01F, EGT02F, EGT03F, EGT04F, EGT06F, EGT07F, EGT08F, EGT10F, EGT11F, EGT13F, EGT14F, EGT15F, EGT16F, Crankcase_PressureHPIinH2O, Engine_Oil_Level, Post_Oil_FilterPSI, Rifle_Oil_PressurePSI, Coolant_PressurePSI, EGT17F, AVL
	Validación	Validación cruzada
	Máxima profundidad de árbol	6
	Mínimo de casos en un nodo filial	100

Resultados	Mínimo de casos en un nodo parental	50
	Variables independientes incluidas	AVL, Fuel_TemperatureF, Coolant_PressurePSI, Pedal, EGT08F, EGT06F, IMTRBRF, EGT17F, Post_Oil_FilterPSI, Engine_Oil_TemperatureF, EGT14F, EGT04F, IMPLBPSI, RPM, Coolant_temperatureF, IMTRBFF, EGT10F, EGT15F, EGT01F, EGT11F, IMTLBFF, EGT16F, EGT03F, Fuel_rateLh, EGT02F, F_de_Carga, EGTAVF, EGT13F, Battery_potentialV, Estado_Motor, EGT07F, Injector_TimingPSI, Barometric_PressurePSI, Rifle_Oil_PressurePSI, Crankcase_PressureHPInH2O, Percent_Load_At_Current_Speed, Injector_MeteringPSI, Actual_Percent_Torque_, Potencia
	Número de nodos	4648
	Número de nodos terminales	2940
	Profundidad	6



Riesgo

Método	Estimación	Tip. Error
Resustitución	,194	,001
Validación cruzada	,208	,001

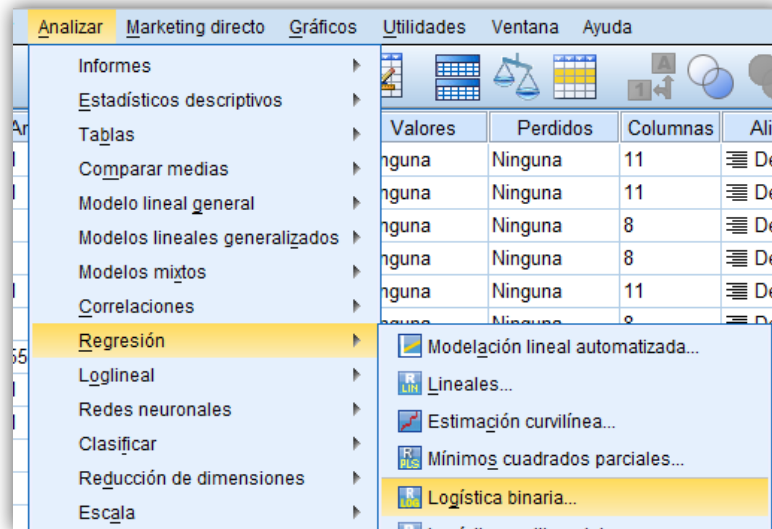
Clasificación

Observado	Pronosticado		Porcentaje correcto
	0	1	
0	38339	6357	,7
1	14109	46614	,9
Porcentaje global			,8

14.9 SPSS Statistics input: Regresión Logística

Considerando la misma configuración de variables del *dataset* del Anexo 14.7, los parámetros de ingreso son los siguientes:

a. Selección del algoritmo:

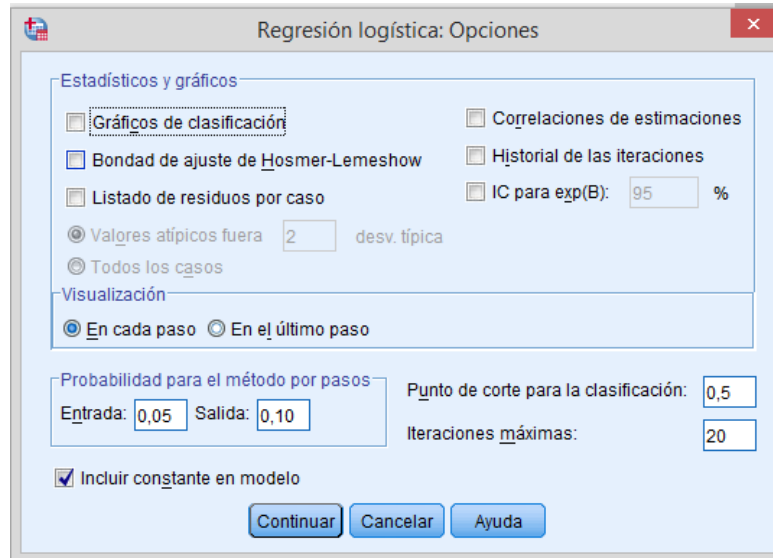


b. Especificación de variables:



Especificación de variable independiente y dependientes.

c. Selección de opciones:



En este caso se mantienen las opciones que SPSS entrega por defecto.

14.10 SPSS Statistics output: Regresión Logística

```
LOGISTIC REGRESSION VARIABLES Target_1
/METHOD=ENTER RPM Pedal F_de_Carga Potencia Estado_Motor Actual_Percent_Torque_Percent_Load_At_Current_Speed Fuel_rateLh
Barometric_PressurePSI IMPLBPSI IMTLBFF EGTAVF Battery_potentialV Coolant_temperatureF Fuel_TemperatureF Engine_Oil_TemperatureF
Injector_MeteringPSI Injector_TimingPSI IMTRBFF IMTRBRF EGT01F EGT02F EGT03F EGT04F EGT06F EGT07F EGT08F EGT10F EGT11F EGT13F EGT14F EGT15F
EGT16F Crankcase_PressureHPInH2O Engine_Oil_Level Post_Oil_FilterPSI Rifle_Oil_PressurePSI Coolant_PressurePSI
EGT17F
/CONTRAST (Estado_Motor)=Indicator
/CLASSPLOT
/PRINT=GOODFIT CI(95)
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Regresión logística

Notas

Resultados creados		29-feb-2016 22:35:12
Comentarios		
Entrada	Datos	C:\Users\Matías\Desktop\Data Tesis - MBE\Análisis exploratorio datos SPSS ultima base\Base datos random 38 variables con estado de motor.sav
	Conjunto de datos activo	Conjunto_de_datos1
	Filtro	<ninguno>
	Peso	<ninguno>
	Segmentar archivo	<ninguno>
	Núm. de filas del archivo de trabajo	421677
Tratamiento de los datos perdidos	Definición de perdidos	Los valores perdidos definidos por el usuario se consideran como perdidos

Sintaxis	<pre> LOGISTIC REGRESSION VARIABLES Target_1 /METHOD=ENTER RPM Pedal F_de_Carga Potencia Estado_Motor Actual_Percent_Torque_ Percent_Load_At_Current_Speed Fuel_rateLh Barometric_PressurePSI IMPLBPSI IMTLBFF EGTAVF Battery_potentialV Coolant_temperatureF Fuel_TemperatureF Engine_Oil_TemperatureF Injector_MeteringPSI Injector_TimingPSI IMTRBFF IMTRBRF EGT01F EGT02F EGT03F EGT04F EGT06F EGT07F EGT08F EGT10F EGT11F EGT13F EGT14F EGT15F EGT16F Crankcase_PressureHPIinH2O Engine_Oil_Level Post_Oil_FilterPSI Rifle_Oil_PressurePSI Coolant_PressurePSI EGT17F /CONTRAST (Estado_Motor)=Indicator /CLASSPLOT /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5). </pre>	
Recursos	Tiempo de procesador	00 00:00:10,094
	Tiempo transcurrido	00 00:00:10,101

[Conjunto_de_datos1] C:\Users\Matias\Desktop\Data Tesis - MBE\Análisis exploratorio datos SPSS ultima base\Base datos random 39 variables con estado de motor.sav

Resumen del procesamiento de los casos

		N	Porcentaje
Casos no ponderados			
Casos seleccionados	Incluidos en el análisis	421677	100,0
	Casos perdidos	0	,0
	Total	421677	100,0
Casos no seleccionados			
		0	,0
Total		421677	100,0

Codificación de la variable dependiente

Valor original	Valor interno
0	0
1	1

Codificaciones de variables categóricas

		Frecuencia	Codificación de parámetros			
			(1)	(2)	(3)	(4)
Estado_Motor	Alta en	394	1,000	,000	,000	,000
	Potencia	45487	,000	1,000	,000	,000
	Ralentá-	148	,000	,000	1,000	,000
	Retardo	30	,000	,000	,000	1,000
	Transici	375618	,000	,000	,000	,000

Bloque 0: Bloque inicial

Tabla de clasificación,a,b

			Pronosticado		Porcentaje correcto
			Target_1		
Observado			0	1	
Paso 0	Target_1	0	0	209354	,0
		1	0	212323	100,0
Porcentaje global					50,4

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 0	Constante	,014	,003	20,904	1	,000	1,014

Variables que no están en la ecuación

			Puntuación	gl	Sig.
Paso 0	Variables	RPM	76,441	1	,000
		Pedal	3,074	1	,080
		F_de_Carga	129,125	1	,000
		Potencia	58,776	1	,000
		Estado_Motor	262,844	4	,000
		Estado_Motor(1)	3,820	1	,051
		Estado_Motor(2)	190,831	1	,000
		Estado_Motor(3)	69,011	1	,000
		Estado_Motor(4)	1,117	1	,291
		Actual_Percent_Torque_	71,440	1	,000
		Percent_Load_At_Current_Speed	129,066	1	,000
		Fuel_rateLh	45,169	1	,000
		Barometric_PressurePSI	83,408	1	,000
		IMPLBPSI	,048	1	,827
		IMTLBFF	344,033	1	,000
		EGTAVF	214,788	1	,000
		Battery_potentialV	230,639	1	,000
		Coolant_temperatureF	692,866	1	,000
		Fuel_TemperatureF	1102,625	1	,000
		Engine_Oil_TemperatureF	38,452	1	,000

Injector_MeteringPSI	63,794	1	,000
Injector_TimingPSI	387,462	1	,000
IMTRBFF	582,083	1	,000
IMTRBRF	319,932	1	,000
EGT01F	63,362	1	,000
EGT02F	227,278	1	,000
EGT03F	265,126	1	,000
EGT04F	455,475	1	,000
EGT06F	130,016	1	,000
EGT07F	425,058	1	,000
EGT08F	266,173	1	,000
EGT10F	335,952	1	,000
EGT11F	378,864	1	,000
EGT13F	246,998	1	,000
EGT14F	7,842	1	,005
EGT15F	160,534	1	,000
EGT16F	67,365	1	,000
Crankcase_PressureHPInH2O	280,282	1	,000
Post_Oil_FilterPSI	14,773	1	,000
Rifle_Oil_PressurePSI	282,794	1	,000
Coolant_PressurePSI	745,362	1	,000
EGT17F	364,692	1	,000

Bloque 1: Método = Introducir

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	23348,936	41	,000
	Bloque	23348,936	41	,000
	Modelo	23348,936	41	,000

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	561198,607	,054	,072

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	4377,310	8	,000

Tabla de contingencias para la prueba de Hosmer y Lemeshow

		Target_1 = 0		Target_1 = 1		Total
		Observado	Esperado	Observado	Esperado	
Paso 1	1	31583	30151,879	10585	12016,121	42168
	2	28337	26204,927	13831	15963,073	42168
	3	25299	24054,016	16869	18113,984	42168
	4	21896	22385,852	20272	19782,148	42168
	5	19291	21047,941	22877	21120,059	42168
	6	17176	19897,352	24993	22271,648	42169
	7	16166	18825,163	26002	23342,837	42168
	8	16454	17653,980	25714	24514,020	42168
	9	16264	16142,312	25904	26025,688	42168
	10	16888	12990,579	25276	29173,421	42164

Tabla de clasificación

	Observado	Target_1	Pronosticado		Porcentaje correcto
			Target_1		
			0	1	
Paso 1	Target_1	0	29161	17958	55,6
		1	23287	35013	66,1
	Porcentaje global				60,9

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
								Inferior	Superior
Paso 1a	RPM	,000	,000	42,555	1	,000	1,000	1,000	1,000
	Pedal	,005	,000	275,954	1	,000	1,005	1,005	1,006
	F_de_Carga	,043	,032	1,798	1	,180	1,044	,980	1,112
	Potencia	,000	,000	2,266	1	,132	1,000	1,000	1,000
	Estado_Motor			135,228	4	,000			
	Estado_Motor(1)	,274	,105	6,857	1	,009	1,315	1,071	1,615
	Estado_Motor(2)	,110	,015	54,059	1	,000	1,116	1,084	1,149
	Estado_Motor(3)	-1,958	,228	73,958	1	,000	,141	,090	,221
	Estado_Motor(4)	,029	,383	,006	1	,940	1,029	,486	2,181
	Actual_Percent_Torque_	-,012	,003	13,940	1	,000	,989	,983	,995
	Percent_Load_At_Current_Sp	-,031	,032	,924	1	,336	,969	,910	1,033
	eed								
	Fuel_rateLh	,000	,000	3,404	1	,065	1,000	1,000	1,001
	Barometric_PressurePSI	,251	,021	143,498	1	,000	1,285	1,233	1,339
	IMPLBPSI	-,003	,001	38,335	1	,000	,997	,996	,998
	IMTLBFF	,013	,000	743,875	1	,000	1,013	1,012	1,014
	EGTAVF	,002	,000	82,047	1	,000	1,002	1,002	1,003
	Battery_potentialV	,087	,007	147,527	1	,000	1,091	1,076	1,107
	Coolant_temperatureF	-,012	,001	369,319	1	,000	,988	,987	,990
	Fuel_TemperatureF	,007	,000	329,070	1	,000	1,007	1,006	1,007

14.11 The R Project for Statistical Computing input / output: SVM Kernel Lineal

```
SVM_100_por ciento_Linear_C1.Rout x
1
2 R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
3 Copyright (C) 2015 The R Foundation for Statistical Computing
4 Platform: x86_64-unknown-linux-gnu (64-bit)
5
6 R is free software and comes with ABSOLUTELY NO WARRANTY.
7 You are welcome to redistribute it under certain conditions.
8 Type 'license()' or 'licence()' for distribution details.
9
10 Natural language support but running in an English locale
11
12 R is a collaborative project with many contributors.
13 Type 'contributors()' for more information and
14 'citation()' on how to cite R or R packages in publications.
15
16 Type 'demo()' for some demos, 'help()' for on-line help, or
17 'help.start()' for an HTML browser interface to help.
18 Type 'q()' to quit R.
19
20 Warning: namespace 'kernlab' is not available and has been replaced
21 by .GlobalEnv when processing object 'mod'
22 [Previously saved workspace restored]
23
24 > #Carga de librerias
25 > library("e1071")
26 > library(foreign)
27 >
28 > #Indica directorio propio para cargar librerias
29 > .libPaths('/home/asantibanez/librerias_r')
30 >
31 > #Importación de datos desde SPSS
32 > matias <- read.spss("/home/asantibanez/bases_de_datos/SPSS/Base datos
33 random 38 variables con estado de motor.sav", to.data.frame=TRUE)
34 re-encoding from CP1252
35 Warning message:
```

```

35 Warning message:
36 In read.spss("/home/asantibanez/bases_de_datos/SPSS/Base datos random 38
37 variables con estado de motor.sav", :
38 /home/asantibanez/bases_de_datos/SPSS/Base datos random 38 variables con
39 estado de motor.sav: Unrecognized record type 7, subtype 18 encountered in
40 system file
41 > dataset.1.porciento = subset(matias,select = c(ESN,AVL,RPM,Pedal,F_de_Carga,
42 Potencia,Actual_Percent_Torque_,Percent_Load_At_Current_Speed,Fuel_rateLh,
43 Barometric_PressurePSI,IMPLBPSI,IMTLBFF,EGTAVF,Battery_potentialV,Coolant_temperatureF,
44 Fuel_TemperatureF,Engine_Oil_TemperatureF,Injector_MeteringPSI,Injector_TimingPSI,
45 IMTRBFF,IMTRBRF,EGT01F,EGT02F,EGT03F,EGT04F,EGT06F,EGT07F,EGT08F,EGT10F,EGT11F,EGT13F,
46 EGT14F,EGT15F,EGT16F,Crankcase_PressureHPIinH2O,
47 Post_Oil_FilterPSI,Rifle_Oil_PressurePSI,Coolant_PressurePSI,EGT17F,Target_1,Id_dato))
48 >
49 > #Convirtiendo campo target en factor para c#C#culo de SVM
50 > attach(dataset.1.porciento)
51 > target <- as.factor(Target_1)
52 >
53 > #Generaci#n base de trabajo
54 > base2=data.frame(dataset.1.porciento,target)
55 >
56 > #Divisi#n en base Test y Train
57 > index <- sample(1:nrow(base2),round(0.75*nrow(base2)))
58 > entrena=base2[index,]
59 > muestra=base2[-index,]
60 >
61 > #Realizaci#n de modelo con la librer#e e1071
62 > modelo=svm(target~ESN+AVL+RPM+Pedal+F_de_Carga+Potencia+Actual_Percent_Torque_+
63 Percent_Load_At_Current_Speed+Fuel_rateLh+Barometric_PressurePSI+IMPLBPSI+IMTLBFF+EGTAVF+
64 Battery_potentialV+Coolant_temperatureF+Fuel_TemperatureF+Engine_Oil_TemperatureF+
65 Injector_MeteringPSI+Injector_TimingPSI+IMTRBFF+IMTRBRF+EGT01F+EGT02F+EGT03F+EGT04F+EGT06F
66 +EGT07F+EGT08F+EGT10F+EGT11F+EGT13F+EGT14F+EGT15F+EGT16F+Crankcase_PressureHPIinH2O+
67 Post_Oil_FilterPSI+Rifle_Oil_PressurePSI+Coolant_PressurePSI+EGT17F,data=entrena,
68 method="C-classification",kernel="linear",cost=1)
70 > #Analizamos el comportamiento
71 > predic=data.frame(predict(modelo,muestra))
72 > muestra=cbind(muestra,predic=predic)
73 >
74 > #Campo que guarda los resultados
75 > tabla.resultados <-table(muestra$target,muestra$predict.modelo..muestra.)
76 >
77 > as.data.frame(table(target))
78 target Freq
79 1 0 209354
80 2 1 212323
81 >
82 > summary(Target_1)
83 Min. 1st Qu. Median Mean 3rd Qu. Max.
84 0.0000 0.0000 1.0000 0.5035 1.0000 1.0000
85 > save.image("/home/asantibanez/Resultados_modelos/SVM_Base_100_porcentaje_Lineal_C1.RData")
86 >
87 > proc.time()
88 user system elapsed
89 61442.29 6.57 61407.47

```

Este proceso tiene como objetivo entregar la matriz de confusi#n que se guarda en la variable "tabla.resultados", que es el siguiente:

```

> tabla.resultados
      0      1
0 25539 26984
1 13861 39035

```

14.12 The R Project for Statistical Computing input / output: SVM Kernel Radial

```
SVM_100_porcentaje_Radial.Rout x
1
2 R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
3 Copyright (C) 2015 The R Foundation for Statistical Computing
4 Platform: x86_64-unknown-linux-gnu (64-bit)
5
6 R is free software and comes with ABSOLUTELY NO WARRANTY.
7 You are welcome to redistribute it under certain conditions.
8 Type 'license()' or 'licence()' for distribution details.
9
10 Natural language support but running in an English locale
11
12 R is a collaborative project with many contributors.
13 Type 'contributors()' for more information and
14 'citation()' on how to cite R or R packages in publications.
15
16 Type 'demo()' for some demos, 'help()' for on-line help, or
17 'help.start()' for an HTML browser interface to help.
18 Type 'q()' to quit R.
19
20 Warning: namespace 'kernlab' is not available and has been replaced
21 by .GlobalEnv when processing object 'mod'
22 [Previously saved workspace restored]
23
24 > #Carga de librerias
25 > library("e1071")
26 > library(foreign)
27 >
28 > #Indica directorio propio para cargar librerias
29 > .libPaths('/home/asantibanez/librerias_r')
30 >
31 > #Importación de datos desde SPSS
32 > matias <- read.spss("/home/asantibanez/bases_de_datos/SPSS/Base datos random 38
33 variables con estado de motor.sav", to.data.frame=TRUE)
34 re-encoding from CP1252
35 Warning message:
36 In read.spss("/home/asantibanez/bases_de_datos/SPSS/Base datos random 38 variables
```

```

37 con estado de motor.sav", :
38 /home/asantibanez/bases_de_datos/SPSS/Base datos random 38 variables con estado de
39 motor.sav: Unrecognized record type 7, subtype 18 encountered in system file
40 > dataset.1.porciento = subset(matias, select = c(ESN,AVL,RPM,Pedal,F_de_Carga,Potencia,
41 Actual_Percent_Torque,Percent_Load_At_Current_Speed,Fuel_rateLh,Barometric_PressurePSI,
42 IMPLBPSI,IMTLBFF,EGTAVF,Battery_potentialV,Coolant_temperatureF,Fuel_TemperatureF,
43 Engine_Oil_TemperatureF,Injector_MeteringPSI,Injector_TimingPSI,IMTRBFF,IMTRBRF,EGT01F,
44 EGT02F,EGT03F,EGT04F,EGT06F,EGT07F,EGT08F,EGT10F,EGT11F,EGT13F,EGT14F,EGT15F,EGT16F,
45 Crankcase_PressureHPIinH2O,Post_Oil_FilterPSI,Rifle_Oil_PressurePSI,Coolant_PressurePSI,
46 EGT17F,Target_1,Id_datos))
47 >
48 > #Convirtiendo campo target en factor para c#FF#culo de SVM
49 > attach(dataset.1.porciento)
50 > target <- as.factor(Target_1)
51 >
52 > #Generaci#FF#n base de trabajo
53 > base2=data.frame(dataset.1.porciento,target)
54 >
55 > #Divisi#FF#n en base Test y Train
56 > index <- sample(1:nrow(base2),round(0.75*nrow(base2)))
57 > entrena=base2[index,]
58 > muestra=base2[-index,]
59 >
60 > #Realizaci#FF#n de modelo con la librer#FF#a e1071
61 > modelo=svm(target~ESN+AVL+RPM+Pedal+F_de_Carga+Potencia+Actual_Percent_Torque +
62 Percent_Load_At_Current_Speed+Fuel_rateLh+Barometric_PressurePSI+IMPLBPSI+IMTLBFF+EGTAVF+
63 Battery_potentialV+Coolant_temperatureF+Fuel_TemperatureF+Engine_Oil_TemperatureF+
64 Injector_MeteringPSI+Injector_TimingPSI+IMTRBFF+IMTRBRF+EGT01F+EGT02F+EGT03F+EGT04F+EGT06F+
65 EGT07F+EGT08F+EGT10F+EGT11F+EGT13F+EGT14F+EGT15F+EGT16F+Crankcase_PressureHPIinH2O+
66 Post_Oil_FilterPSI+Rifle_Oil_PressurePSI+Coolant_PressurePSI+EGT17F,data=entrena,
67 method="C-classification",kernel="radial",cost=1,gamma=.03462159)

69 > #Analizamos el comportamiento
70 > predic=data.frame(predict(modelo,muestra))
71 > muestra=cbind(muestra,predic=predic)
72 >
73 > #Campo que guarda los resultados
74 > tabla.resultados <-table(muestra$target,muestra$predict.modelo..muestra.)
75 >
76 > as.data.frame(table(target))
77 target Freq
78 1 0 209354
79 2 1 212323
80 >
81 > summary(Target_1)
82 Min. 1st Qu. Median Mean 3rd Qu. Max.
83 0.0000 0.0000 1.0000 0.5035 1.0000 1.0000
84 > save.image("/home/asantibanez/Resultados_modelos/SVM_Base_100_porcentaje_Radial_1_core.RData")
85 >
86 > proc.time()
87 user system elapsed
88 27296.36 3.63 27280.67
89 .939

```

El resultado que entrega la matriz de confusi#FF#n en la variable "tabla.resultados" es la siguiente:

```
> tabla.resultados
```

```

      0      1
0 34701 10455
1 17680 42583

```

14.13 The R Project for Statistical Computing input / output: SVM Kernel Sigmoide

```
SVM_100_por ciento_Sigmoideal.Rout x
1
2 R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
3 Copyright (C) 2014 The R Foundation for Statistical Computing
4 Platform: x86_64-unknown-linux-gnu (64-bit)
5
6 R is free software and comes with ABSOLUTELY NO WARRANTY.
7 You are welcome to redistribute it under certain conditions.
8 Type 'license()' or 'licence()' for distribution details.
9
10 Natural language support but running in an English locale
11
12 R is a collaborative project with many contributors.
13 Type 'contributors()' for more information and
14 'citation()' on how to cite R or R packages in publications.
15
16 Type 'demo()' for some demos, 'help()' for on-line help, or
17 'help.start()' for an HTML browser interface to help.
18 Type 'q()' to quit R.
19
20 Warning: namespace 'caret' is not available and has been replaced
21 by .GlobalEnv when processing object 'ctrl'
22 [Previously saved workspace restored]
23
24 > #Carga de librerias
25 > library("e1071")
26 > library(foreign)
27 >
28 > #Indica directorio propio para cargar librerias
29 > .libPaths('/home/asantibanez/librerias_r')
30 >
31 > #Importación de datos desde SPSS
32 > matias <- read.spss("/home/asantibanez/bases_de_datos/SPSS/Base datos
33 random 38 variables con estado de motor.sav", to.data.frame=TRUE)
34 re-encoding from CP1252
35 Warning message:
36 In read.spss("/home/asantibanez/bases_de_datos/SPSS/Base datos random
```

```

37 38 variables con estado de motor.sav", " :
38 /home/asantibanez/bases_de_datos/SPSS/Base datos random 38 variables
39 con estado de motor.sav: Unrecognized record type 7, subtype 18
40 encountered in system file
41 > dataset.1.porciento = subset(matias,select = c(ESN,AVL,RPM,Pedal,F_de_Carga,
42 Potencia,Actual_Percent_Torque_,Percent_Load_At_Current_Speed,Fuel_rateLh,
43 Barometric_PressurePSI,IMPLBPSI,IMTLBFF,EGTAVF,Battery_potentialV,Coolant_temperatureF,
44 Fuel_TemperatureF,Engine_Oil_TemperatureF,Injector_MeteringPSI,Injector_TimingPSI,
45 IMTRBFF,IMTRBRF,EGT01F,EGT02F,EGT03F,EGT04F,EGT06F,EGT07F,EGT08F,EGT10F,EGT11F,
46 EGT13F,EGT14F,EGT15F,EGT16F,Crankcase_PressureHPIinH2O,Post_Oil_FilterPSI,
47 Rifle_Oil_PressurePSI,Coolant_PressurePSI,EGT17F,Target_1,Id_dato))
48 >
49 > #Convirtiendo campo target en factor para c(2) cálculo de SVM
50 > attach(dataset.1.porciento)
51 > target <- as.factor(Target_1)
52 >
53 > #Generación base de trabajo
54 > base2=data.frame(dataset.1.porciento,target)
55 >
56 > #División en base Test y Train
57 > index <- sample(1:nrow(base2),round(0.75*nrow(base2)))
58 > entrena=base2[index,]
59 > muestra=base2[-index,]
60 >
61 > #Realización de modelo con la librería e1071
62 > modelo=svm(target~ESN+AVL+RPM+Pedal+F_de_Carga+Potencia+Actual_Percent_Torque_+
63 Percent_Load_At_Current_Speed+Fuel_rateLh+Barometric_PressurePSI+IMPLBPSI+IMTLBFF+
64 EGTAVF+Battery_potentialV+Coolant_temperatureF+Fuel_TemperatureF+Engine_Oil_TemperatureF+
65 Injector_MeteringPSI+Injector_TimingPSI+IMTRBFF+IMTRBRF+EGT01F+EGT02F+EGT03F+EGT04F+
66 EGT06F+EGT07F+EGT08F+EGT10F+EGT11F+EGT13F+EGT14F+EGT15F+EGT16F+Crankcase_PressureHPIinH2O+
67 Post_Oil_FilterPSI+Rifle_Oil_PressurePSI+Coolant_PressurePSI+EGT17F,data=entrena,
68 method="C-classification",kernel="sigmoid",cost=1,gamma=.03462159)
69
70 > #Analizamos el comportamiento
71 > predic=data.frame(predict(modelo,muestra))
72 > muestra=cbind(muestra,predic=predic)
73 >
74 > #Campo que guarda los resultados
75 > tabla.resultados <-table(muestra$target,muestra$predict.modelo..muestra.)
76 >
77 > as.data.frame(table(target))
78 target Freq
79 1 0 209354
80 2 1 212323
81 >
82 > summary(Target_1)
83 Min. 1st Qu. Median Mean 3rd Qu. Max.
84 0.0000 0.0000 1.0000 0.5035 1.0000 1.0000
85 > save.image("/home/asantibanez/Resultados_modelos/SVM_Base_100_porcentaje_Sigmoide.RData")
86 >
87 > proc.time()
88 | user system elapsed
89 24944.033 3.213 24932.274

```

La matriz de confusión resultante en la variable “tabla.resultados” es:

```

> tabla.resultados
      0      1
0 26286 26163
1 26162 26808

```


14.14 Narrativa casos de uso: Interacción sistema e Ingeniero Planificación

Caso de Uso 1: Actualización y revisión de información

Primary Actor: Ingeniero_planificación

Caso de éxito o Flujo normal de Operación

#	Actor Action	System Response	Obs
1	INGENIERO_PLANIFICACION ingresa a la página de intranet en el navegador de su computador, en red de la compañía.		Puede ser Internet Explorer, Firefox o Chrome.
2		SISTEMA recibe solicitud de INGENIERO_PLANIFICACION y despliega la pantalla inicial en el navegador del usuario.	Pantalla de bienvenida e inicio del sistema.
3	INGENIERO_PLANIFICACION selecciona opción de iniciar el sistema o finalizar.		Si el usuario no inicia, el sistema se detiene.
4		SISTEMA solicita información al sistema SPECTO, quién contiene información de los motores Cummins.	SPECTO es un sistema externo, al cual se accede sólo mediante consultas de bases de datos.
5	SPECTO, mediante sus métodos de respuesta, envía la información solicitada a SISTEMA.		
6		SISTEMA almacena información dentro de su base de datos.	
7		SISTEMA despliega en pantalla un resumen de la información cargada en el navegador del usuario.	Información tal como Cantidad de filas cargadas, errores detectados, datos ya cargados y periodo de carga.

Caso de Uso 2: Selección tipo de análisis

Primary Actor: Ingeniero_planificación

Caso de éxito o Flujo normal de Operación

#	Actor Action	System Response	Obs
1	INGENIERO_PLANIFICACION escoge opción de mantenimiento predictivo.		El usuario puede escoger distintos tipos de modelo de predicción que son establecidos previamente.
2		SISTEMA filtra información desde sus bases de datos y envía a software estadístico.	En caso de ser necesario, el usuario podría realizar más filtros que los establecidos previamente.
3		SISTEMA ejecuta modelo matemático desde el software que sea necesario.	Hasta el momento se han utilizado 2 software: R y SPSS Statistics.
4		SISTEMA estandariza la respuesta y genera los resultados en el formato correspondiente.	Dependiendo de la fuente de los resultados, el sistema debe estandarizar la respuesta en los formatos correspondientes.
5		SISTEMA guarda los resultados dentro de sus bases de datos para uso posterior.	

Caso de Uso 3:

Revisión y descarga de documentos

Primary Actor

Ingeniero_planificación

Caso de éxito o Flujo normal de Operación

#	Actor Action	System Response	Obs
1		SISTEMA despliega resultados y opción de descarga en pantalla del usuario.	Los resultados se exponen con características tal como Faena, ESN, Probabilidad de falla, fecha y nodo (si corresponde).
2	INGENIERO_PLANIFICACION indica que el procesamiento de información está ok y selecciona descargar información		
3		SISTEMA prepara la información y la envía a través del navegador de INGENIERO_PLANIFICACION para su descarga.	El formato de salida común será planilla Excel.

14.15 Narrativa casos de uso: Sistema de control

Caso de Uso 1: Estimación demanda equipos

Primary Actor: Sistema_control

Caso de éxito o Flujo normal de Operación

#	Actor Action	System Response	Obs
1	SISTEMA_CONTROL solicita ejecución de algoritmo predeterminado para predicción estado de motores.		SISTEMA_CONTROL no tiene posibilidad de seleccionar alguno de los algoritmos disponibles para el análisis, ya que es un parámetro definido previamente.
2		SISTEMA solicita actualización de información de motores a SPECTO.	Esta información debe ser actualizada ya que esta revisión es realizada de forma automática durante el día.
3	SPECTO envía información actualizada de motores Cummins.		
4		SISTEMA recibe nueva información y es procesada con algoritmo predeterminado en el SISTEMA.	
5		SISTEMA procesa información y genera resultados.	
6		SISTEMA almacena los resultados dentro de su base de datos para análisis posteriores.	

Caso de Uso 2: Revisión y determinación equipos críticos

Primary Actor: Almacenamiento_planificación

Caso de éxito o Flujo normal de Operación

#	Actor Action	System Response	Obs
1		SISTEMA solicita información de la última planificación registrada para motores.	Esta información no se encuentra dentro de SISTEMA, es externo.
2	ALMACENAMIENTO_PLANIFICACION envía información solicitada por SISTEMA.		Generalmente esta información será entregada en Excel.
3		SISTEMA realiza comparación último plan versus pronóstico realizado.	
4		SISTEMA filtra (elimina) equipos ya intervenidos del listado final de equipos a analizar.	
5		SISTEMA revisa parámetros de equipos críticos en SISTEMA versus los establecidos en planificación.	Estos parámetros de criticidad se encuentran previamente definidos en SISTEMA.
6		Determinación y generación del listado por criticidad de equipo.	Este listado queda listo para cuando algún componente de SISTEMA lo requiera.

Caso de Uso 3: Generación alarmas
Primary Actor Sistema_correo_electrónico
Caso de éxito o Flujo normal de Operación

#	Actor Action	System Response	Obs
1		SISTEMA solicita listado criticidad de equipos.	Este listado corresponde a la información procesada y clasificada de acuerdo a la hora de procesamiento.
2		SISTEMA revisa información correspondiente a parámetros de acción ante criticidad de equipos para generación de alarmas.	Los parámetros de acción deben estar previamente establecidos en SISTEMA.
3		SISTEMA revisa información relacionada a los responsables para cada motor que se deba generar una alarma.	Información de los responsables debe estar previamente en SISTEMA.
4		SISTEMA genera listado de alarmas y contenido a enviar a los distintos responsables, solicitando la distribución de esta información.	
3	SISTEMA_CORREO_ELECTRONICO recibe información y comienza a enviar por correo electrónico la información descrita.		SISTEMA_CORREO_ELECTRONICO puede ser cualquier sistema automatizado de correo electrónico masivo.