



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

MODELO PREDICTIVO BASADO EN MACHINE LEARNING DE ORDENES DE
TRABAJO RIESGOSAS PARA MANTENIMIENTO DE EQUIPOS MINEROS

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

JAVIER ANTONIO BARROSO SALGADO

PROFESOR GUÍA:
IGNACIO ALEJANDRO CALISTO LEIVA

MIEMBROS DE LA COMISIÓN:
EDGARDO SANTIBÁÑEZ VIANI
FELIPE VILDOSO CASTILLO

Este trabajo ha sido parcialmente financiado por BHP

SANTIAGO DE CHILE
2018

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
POR: JAVIER ANTONIO BARROSO SALGADO
FECHA: 2018
PROF. GUÍA: IGNACIO ALEJANDRO CALISTO LEIVA

MODELO PREDICTIVO BASADO EN MACHINE LEARNING DE ORDENES DE TRABAJO RIESGOSAS PARA MANTENIMIENTO DE EQUIPOS MINEROS

Objetivo: Desarrollar un modelo predictivo que permita al área de Supply de BHP Minerals América hacer gestión proactiva sobre ordenes de trabajo de mantenimiento con requerimiento de materiales y pudiendo anticiparse así a posibles caídas en el indicador WOPIFOT de nivel de servicio de Supply.

Contexto: El área de Supply funciona como una balanza entre los requerimientos de materiales y servicios para la operación y los objetivos económicos del negocio, optimizando la relación costo/disponibilidad de materiales. WOPIFOT es un indicador binario de nivel de servicio de Supply que mide para cada orden de trabajo si esta fue completada satisfactoriamente y con todos los materiales requeridos. Durante el año fiscal 2017, el promedio de WOPIFOT fue de 85 %, o en otras palabras, 15 % de las órdenes de trabajo no se desarrollaron correctamente. Esto supone que existen oportunidades de mejora en el ámbito de disponibilidad de materiales y disciplina operacional que pueden ser apoyadas con la implementación de una herramienta de detección temprana de OTs con riesgo de fallas.

Hipótesis: Existen patrones históricos de fallas a nivel de materiales y su relación con otras variables como horizonte de planificación, impacto en el negocio del material, sistema de control de inventario, entre otras, que permite desarrollar un modelo predictivo basado en herramientas de machine learning que detecte estos patrones y también su evolución en el tiempo al modificarse la cultura operacional a medida que se haga gestión sobre las fallas.

Diseño: Se propone desarrollar una herramienta basada en el algoritmo Gradient Boosting Trees que cargue los datos de órdenes de trabajo abiertas en el periodo y las clasifique en riesgosas o no riesgosas de acuerdo al comportamiento histórico de sus componentes y otras características y asigne una posible causa de falla para apoyar la gestión.

Resultados: El algoritmo de GBT entrenado con 9 variables relativas a las órdenes de trabajos y materiales genera un modelo cuyas métricas de desempeño teóricas en un testeo con datos fuera de muestra, realizado en mayo/2018 con un modelo entrenado con datos históricos hasta abril/2018 son prometedores, con un accuracy de 82,2 %, Cohen's Kappa de 0,625 y AUC de 0,731, lo que permitiría en el caso ideal, aumentar el indicador WOPIFOT PM01 en Spence a sobre 80 % en caso de que las acciones y gestión proactiva sean eficaces.

Conclusión: Incluir nuevas tecnologías de analytics y data science permiten agregar valor en una industria que está constantemente enfocada en la mejora de la producción y donde existen oportunidades reales de mejorar los procesos. Automatizar los procesos de análisis de datos, mediante por ejemplo algoritmos de aprendizaje de máquinas permite a los profesionales dedicar menos tiempo en análisis y reportabilidad y enfocar recursos en gestión y decisión. La gestión del cambio es una etapa fundamental en la utilización de estas nuevas tecnologías.

Este trabajo es dedicado a mi papá, mamá y hermana.

Agradecimientos

Especial agradecimiento a mi familia, su constante apoyo y por ser un pilar fundamental en todo este proceso y por entregarme los valores, herramientas y motivación que me permiten hoy completar una etapa más de mi vida. Sin ellos, esto no hubiera sido posible.

A mis amigos, por su apoyo y compañía y los buenos momentos.

A Fabián Gutierrez y JP Larrondo, por apoyarme como tutores en el desarrollo de esta memoria.

Al equipo de Supply BHP Minerals Americas por la buena acogida y la ayuda para llevar a buen término este proyecto.

Tabla de Contenido

Introducción	1
0.1. Industria Minera del Cobre en Chile	1
0.2. Antecedentes Previos	1
0.3. Caracterización de la Industria	1
0.4. Principales Actores	1
0.5. Marco Legal y Regulatorio	2
0.6. BHP	2
0.7. Antecedentes Generales	2
0.8. Historia	3
0.9. Misión y Objetivos	3
0.10. Estructura Legal y Directiva	3
0.11. Estructura Organizacional	4
0.11.1. Organigrama General BHP Minerals América	4
0.12. Principales Productos	4
0.13. Producción, ventas y tamaño de la empresa	4
0.14. Justificación del Problema	6
0.14.1. Reducción en la disponibilidad de equipos	6
0.14.2. Surplus de Inventario	6
0.14.3. Precios de negociación de compras y contratos	6
0.14.4. Confianza y transparencia entre áreas	7
0.15. Área de Supply	8
0.15.1. Principales funciones	8
0.15.2. Organigrama	8
0.15.3. Clientes	9
0.15.4. Relación con el proyecto de memoria	9
0.16. Glosario	10
0.17. Work Orders (WO) de mantenimiento	11
0.18. Fallas en Work Orders	12
0.19. Impacto económico de fallas en WO	12
0.19.1. Disponibilidad de materiales	12
0.19.2.	12
0.20. Objetivos	13
0.20.1. Objetivo General	13
0.20.2. Objetivos Específicos	13
0.21. Resultados esperados y Alcances	13
0.21.1. Resultados esperados	13

0.21.2. Alcances	13
0.21.2.1. Proyecto enfocado solo a la operación minera Spence	13
0.21.2.2. Modelo enfocado en ordenes de trabajo tipo PM01	14
0.21.2.3. Modelo Predictivo, sin capacidad prescriptiva	14
0.21.2.4. Diseño e implementación del modelo	14
1. Capítulo 1: Marco Teórico y Metodología	15
1.1. Data Mining	15
1.1.1. Data Warehousing	15
1.1.1.0.1. Modelo Estrella	16
1.1.1.1. ETL	16
1.1.2. Modelos Predictivos	17
1.1.3. Árboles de decisión	17
1.1.3.1. Algoritmo ID3	18
1.1.3.2. Algoritmo C4.5	19
1.1.4. Algoritmos de ensamble	19
1.1.4.1. Bagging	19
1.1.4.1.1. Random Forest	20
1.1.4.2. Boosting	20
1.2. CRISP-DM	22
1.2.1. Entendimiento del negocio	23
1.2.2. Entendimiento de los datos	23
1.2.3. Preparación de los datos	23
1.2.4. Modelamiento	23
1.2.5. Evaluación	23
1.2.6. Despliegue	23
1.3. Metodología y plan de trabajo	24
1.3.1. Entendimiento del Negocio	24
1.3.2. Entendimiento de los datos	24
1.3.3. Preparación de los datos	25
1.3.4. Modelamiento	25
1.3.5. Evaluación	25
1.3.6. Despliegue	25
1.4. Levantamiento de procesos	26
1.5. Modelamiento y aplicación de algoritmos	26
2. Capítulo 2: Desarrollo del proyecto de memoria	27
2.1. Levantamiento del proceso de Supply	27
2.2. El proceso de Supply para mantenimiento	27
2.2.1. Metodología de levantamiento de procesos	27
2.2.2. Macro Proceso de supply	27
2.2.3. Proceso de generación de Work Orders	29
2.2.4. Proceso de Compras	30
2.2.5. Métricas de Supply	31
2.2.5.1. WOPIFOT	31
2.2.5.2. MIFOT	32
2.2.5.3. DIFOT	32

2.2.6.	Impacto del proyecto para Supply	32
2.2.7.	Impacto de posibles fallas	32
2.2.7.1.	Falsos Positivos	32
2.2.7.2.	Falsos Negativos	33
2.3.	Consolidación de la base de datos	33
2.4.	Caracterización de los datos históricos de WOPIFOT y MIFOT en PM01 . .	35
2.4.1.	Variables	35
2.4.1.1.	Variables Cualitativas	36
2.4.1.2.	Variables Temporales	37
2.4.1.3.	Transformación de variables	37
2.4.1.4.	Variables Cuantitativas	38
2.4.2.	Estadísticas descriptivas	38
2.4.2.1.	WOPIFOT	38
2.4.2.2.	MIFOT	40
2.4.3.	Desbalance de las clases en los datos	41
2.5.	Modelo predictivo como solución al problema planteado	42
2.6.	Modelo predictivo multi-etapa	42
2.6.1.	Selección del algoritmo a utilizar: Challenge de modelos machine learning	43
2.7.	Construcción del modelo final	47
2.7.1.	Etapa 1: Predicción de Hit o Miss	47
2.7.2.	Resultados teóricos	47
2.7.3.	Etapa 2: Predicción de posible causa de falla	53
2.7.4.	Implementación del modelo predictivo	54
2.8.	Diseño del dashboard para la herramienta WPT	56
2.8.1.	Diseño de experimento	58
2.8.1.1.	Experimento	58
2.8.1.2.	Resultados Esperados	58
2.9.	Visión estratégica	59
	Conclusión	60
2.10.	Cumplimiento de los objetivos propuestos	61
2.11.	Resultados Observados	61
2.12.	Trabajo Futuro	62
	Anexos	64
2.13.	Causas raices de fallas del indicador MIFOT separadas por rol responsable .	66
2.14.	Código WOPIFOT Predictive Tool en R	67
2.14.1.	Extracción, Transformación y Carga de data	67
2.14.2.	Estadística descriptiva	74
2.14.3.	Entrenar Gradient Boosting Trees	76
2.14.4.	Análisis y predicción de resultado de nuevas OT	81
2.14.5.	Incorporación y cruce de data de inventario	85

Índice de Tablas

1.	Producción 2015 y Participación de Mercado grandes empresas mineras . . .	2
1.1.	Algoritmo de Bagging	20
1.2.	Algoritmo de Boosting	21
2.1.	Volumen y variables de las bases de datos disponibles	33
2.2.	Variables de WOPIFOT	33
2.3.	Variables de MIFOT	34
2.4.	Variables de Oniqua	34
2.5.	Volumen y variables del Data Warehouse final	36
2.6.	Variables Cualitativas	37
2.7.	Variables temporales	37
2.8.	Variables Cuantitativas	38
2.9.	Comparación de resultados de modelos	43
2.10.	Hiparparámetros de GBT obtenidos mediante Caret para etapa 1.	47
2.11.	Importancia relativa de las variables	48
2.12.	Matriz de Confusión, Testeo fuera de muestra	52
2.13.	Accuracy del modelo	52
2.14.	Estadísticos del modelo	52
2.15.	Hiparparámetros de GBT obtenidos mediante Caret para etapa 2.	53
2.16.	Resultados GBT en predicción de RCA.	53
2.17.	Importancia relativa de variables en predicción de RCA	54
2.18.	Causas raíces de fallas en MIFOT por rol responsable 1	66
2.19.	Causas raíces de fallas en MIFOT por rol responsable 2	67

Índice de Ilustraciones

1.	Organigrama BHP Minerals América	4
2.	Grandes cifras BHP 2016	5
3.	Proceso de Work Management. Fuente: Taller en Cerro Colorado Marzo/2018	7
4.	Estructura de Supply	8
5.	Estructura de Supply	9
1.1.	Proceso iterativo CRISP-DM	22
2.1.	Macro proceso de supply para mantenimiento	28
2.2.	Proceso de generación de órdenes de trabajo	29
2.3.	Tipos de OT	30
2.4.	Proceso de compra de materiales	31
2.5.	Esquema del Data Warehouse a construir	35
2.6.	Evolución mensual WOPIFOT (6/17-4/18)	38
2.7.	Evolución mensual WOPIFOT por área (6/17-4/18)	39
2.8.	Evolución mensual MIFOT (6/17-4/18)	40
2.9.	Concepto de modelo predictivo multietapa	43
2.10.	Árbol entrenado con todas las variables disponibles	45
2.11.	Árbol entrenado con variables continuas (numéricas)	46
2.12.	Gráfico Importancia relativa de las variables	48
2.13.	Dependencia parcial de Time from Today (tft)	49
2.14.	Dependencia parcial de Time to Release (ttr)	50
2.15.	Dependencia parcial de Time to Execution (tte)	51
2.16.	Herramienta predictiva e interacción con el usuario	55
2.17.	Dashboard de la herramienta WOPIFOT Predictive Tool (WPT)	57
2.18.	Curva de evolución de Analytics. Complejidad vs Valor agregado	59
2.19.	Relación entre complejidad de modelos e intervención humana	60

Introducción

0.1. Industria Minera del Cobre en Chile

0.2. Antecedentes Previos

Chile es sin duda un país minero. La explotación del cobre provee entre el 25 % y el 30 % del presupuesto fiscal y concentra cerca del 15 % del empleo a nivel país¹.

0.3. Caracterización de la Industria

La gran minería del cobre (GMC) en Chile es entendida como el conjunto de empresas mineras que produzcan, dentro del país, en cantidades no inferiores a 75.000 toneladas métricas anuales mediante la explotación y beneficio de minerales de producción propia o de sus filiales o asociados.

En 1998 se crea en Chile el Consejo Minero, asociación que reúne a las empresas productoras de cobre, oro, plata y molibdeno de mayor tamaño que operan en Chile, tanto de capitales nacionales como extranjeros, sean de propiedad pública o privada.²

El año 2015, la producción total de cobre en Chile fue de 5,76 millones de toneladas métricas (TM), equivalente a aproximadamente 30 % de la producción mundial, poniendo a Chile en primer lugar en el ranking de productores de cobre a nivel mundial.

0.4. Principales Actores

Otros actores importantes dentro de la gran minería del cobre son Anglo American, Antofagasta Minerals, BHP, El Abra y Collahuasi, cuya participación de mercado se resume en la siguiente tabla:

¹<http://www.t13.cl/noticia/negocios/el-cobre-seguira-siendo-pilar-del-pib-chile-pero-menos-influen>

²<http://www.consejominero.cl/>

Empresa Minera	2015	Porcentaje del total
Total Codelco	1,731.7	30 %
BHP	1,152.5	21 %
Collahuasi	455.3	8 %
Los Pelambres (Antofagasta Minerals)	375.8	7 %
Anglo American	437.8	10 %
El Abra	147.2	3 %

Tabla 1: Producción 2015 y Participación de Mercado grandes empresas mineras

0.5. Marco Legal y Regulatorio

En relación al marco legal que regula la minería de cobre en Chile, toda actividad minera se rige bajo la ley 17.450, aprobada el 11 de julio de 1971. La ley expone que las empresas señaladas como de la gran minería, pueden ser nacionalizadas, en la oportunidad que determine el Presidente de la república.

*“El Estado tiene el dominio absoluto, exclusivo, inalienable e imprescriptible de todas las minas, comprendiéndose en éstas las covaderas, las arenas metalíferas, los salares, los depósitos de carbón e hidrocarburos y las demás sustancias fósiles, con excepción de las arcillas superficiales. . . ”*³

El decreto de ley 600 (DL600) de 1974 crea una institucionalidad favorable a la inversión extranjera. Esto implica la firma de un Contrato Ley por parte del Estado y del inversionista, situado por encima de la soberanía nacional y caracterizado por estar sujeto a largos periodos de invariabilidad, tanto en el plano tributario (37 %) como en otros aspectos, todos favorables a los intereses del inversionista.⁴

0.6. BHP

0.7. Antecedentes Generales

BHP es una compañía global de recursos naturales, organizada en tres unidades de negocios: Minerals Americas, Minerals Australia y Petroleum. Su casa matriz está ubicada en Melbourne, Australia. BHP Minerals Americas tiene su sede en Chile. BHP concentra la segunda participación de mercado más importante en Chile alcanzando 21 % el año 2015⁵

En cuanto a sus operaciones en Chile, BHP opera en tres yacimientos; Escondida, Cerro

³Código de Minería, http://www.sernageomin.cl/gobiernotransparente/marconormativo/pdf_mn/CODIGODEMINERIA.pdf

⁴DL 600, Estatuto de la inversión extranjera.

⁵<http://www.cochilco.cl/estadisticas/produccion.asp>

Colorado y Spence.

Minera Escondida es una mina de cobre a cielo abierto, ubicada en la Región de Antofagasta, con aproximadamente 10.000 trabajadores tanto propios como contratistas y una producción de 1.152.510 toneladas de cobre fino al año 2015.

BHP Pampa Norte cuenta con aproximadamente 5.500 trabajadores para la producción de cátodos de cobre y 250.000 toneladas de cobre fino producidas al año 2015. Las faenas pertenecientes a pampa norte son Cerro Colorado, una mina de cielo abierto en la región de Tarapacá y Spence, ubicada en Antofagasta.

0.8. Historia

BHP es la empresa de recursos naturales diversificados más grande del mundo, cuyo origen proviene de la fusión de la australiana Broken Hill Proprietary (BHP) y la inglesa Billiton, en 2001. Con presencia en 13 países, su casa matriz está ubicada en Melbourne, Australia.

BHP llegó a Chile en 1984 tras adquirir Utah Corporation, que entonces era el socio principal de Minera Escondida. En el año 2000, Billiton compró Rio Algom, que tenía entre sus activos Spence y Cerro Colorado. Desde mayo de 2013, el CEO de BHP es Andrew Mackenzie, mientras que el chileno Daniel Malchuk es el Presidente de Minerals Americas desde 2014.

0.9. Misión y Objetivos

El propósito de BHP es crear valor para los accionistas a largo plazo mediante el descubrimiento, adquisición, desarrollo y comercialización de recursos naturales, y su misión es poseer y operar grandes, de larga vida, de bajo costo y expansibles recursos diversificados por productos, geografía y mercado. Basándose en los valores fundamentales de sustentabilidad, integridad, respeto, desempeño, simplicidad y responsabilidad.⁶

0.10. Estructura Legal y Directiva

La empresa opera bajo una estructura de compañía de doble cotización con dos compañías matrices (BHP Limited y BHP Plc) operadas como si fuera una sola entidad económica. BHP está dirigida por una Junta y administración unificada.

⁶<http://www.bhp.com/our-approach/our-company/strategy>

0.11. Estructura Organizacional

0.11.1. Organigrama General BHP Minerals América

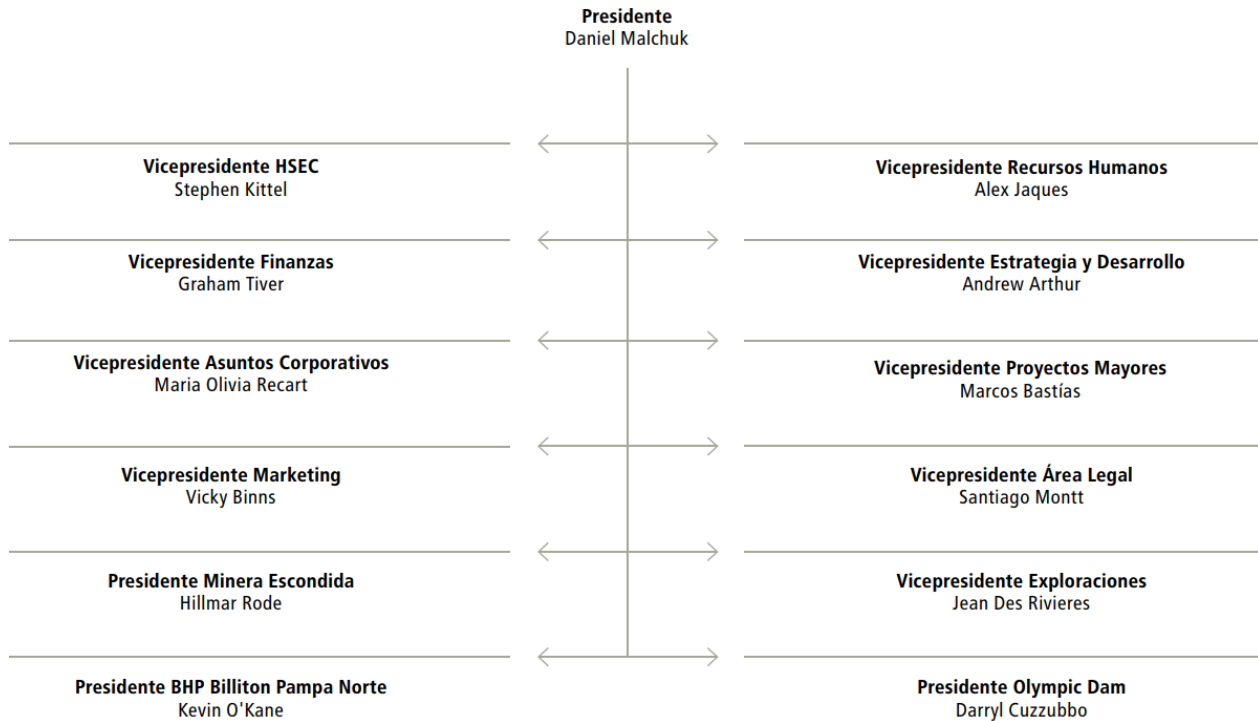


Figura 1: Organigrama BHP Minerals América

0.12. Principales Productos

BHP Chile opera sólo en la extracción de cobre para la producción de concentrado de cobre y cátodos de cobre, logrando al año 2016 una producción total de 1.243.524 t de cobre fino, correspondiente a un 22 % de la producción nacional y 6,2 % de la producción mundial.⁷

0.13. Producción, ventas y tamaño de la empresa

Al año 2016, BHP produce el 22 % de la producción nacional de cobre, según cifras de Cochilco. El siguiente cuadro resume las cifras de producción, venta y capital humano de la

⁷http://www.bhp.com/-/media/documents/media/reports-and-presentations/2017/170807_bhpinformedesustentabilidadatabhpchile.pdf

empresa durante el año fiscal 2016.⁸

produccion y ventas.png

2016					
Datos Operacionales	Métrica	Pampa Norte		Minera Escondida	Total BHP Chile
		Cerro Colorado	Spence		
Total cobre fino producido	tmf	74.002	167.491	1.002.031	1.243.524
Participación de producción de cobre en Chile	%	1,0	3,0	18,0	22
Ventas	MMUS\$	1.402		5.273	6.675
Empleados directos ⁽¹⁾	Personas	953	1.064	3.600	6.810
Contratistas ⁽²⁾	FTE ⁽³⁾	1.003	1.458	5.961	8.422
Inversión comunitaria	MMUS\$	13		25,4	38
Número de proveedores locales ⁽⁴⁾	Proveedores				1.815
Porcentaje de proveedores locales ⁽⁵⁾	%				94

(1) Incluye personal de oficina Santiago.

(2) No se incluyen valores de contratistas de Santiago.

(3) La sigla significa Full Time Equivalent.

(4) Se publica número de proveedores de las tres operaciones juntas.

(5) Porcentaje calculado sobre el total de proveedores nacionales.

Figura 2: Grandes cifras BHP 2016

⁸http://www.bhp.com/-/media/documents/media/reports-and-presentations/2017/170807_bhpinformedesustentabilidadabhpchile.pdf

0.14. Justificación del Problema

Se ha detectado que a nivel de WOPIFOT en PM01 de spence, existe un gap entre la performance real (de al rededor de un 75 % histórico) y el nivel de servicio comprometido como meta (traducido en un indicador de 88 %). Esto significa que aproximadamente un cuarto de las órdenes de trabajo emitidas por el área de mantenimiento no se llevan a cabo en la fecha comprometida (y de hecho esta cantidad puede ser mayor si se considera que otro porcentaje de las OTs son reprogramadas a fechas futuras cuando existen razones que impiden la ejecución de los trabajos de mantenimiento, pudiendo ser una de ellas la falta de materiales).

Esto genera diversos problemas a nivel organizacional y operacional que se traducen en un incremento de los costos de producción por las razones explicadas a continuación:

0.14.1. Reducción en la disponibilidad de equipos

La órdenes de trabajo reflejan una necesidad de materiales para un trabajo de mantenimiento de equipos. Es el deber del área de Supply proveer los materiales a tiempo dada la planificación por parte del área de Work Management. Un primer impacto que tienen las fallas en las órdenes de trabajo es la falta de materiales para mantenimiento lo que impacta directamente en la disponibilidad de equipos para la operación. Este tiempo no operativo de los equipos tiene un impacto negativo en la producción final, y por lo tanto un costo económico evidente pero que no ha sido aún medido por BHP y está también fuera del alcance de este trabajo de memoria.

0.14.2. Surplus de Inventario

Las fallas a nivel de Ordenes de trabajo son también reflejo de una falta de planificación por parte del área de WM. Cuando la planificación es inestable, los esfuerzos de predicción de demanda por parte del área de optimización de inventario tienen peores resultados y por lo tanto los parámetros de reposición y los stocks de materiales en bodega deben ser mayores para cubrir la incertidumbre en la demanda, lo que implica un capital de trabajo mayor en bodega y en muchos casos un surplus de inventario. Este capital de trabajo mayor es un costo monetario directo para la empresa y es también uno de los principales KPI del área de inventario.

0.14.3. Precios de negociación de compras y contratos

Considerando que las OTs son las que dictan la demanda por materiales, tener un proceso de planificación más riguroso permite también estimar de mejor forma la demanda futura, lo que implica que se expande la posibilidad de negociaciones con proveedores de forma de poder

lograr acuerdos más estables y mejores precios para la compañía, traducándose directamente en ahorros.

0.14.4. Confianza y transparencia entre áreas

Otro efecto negativo derivado de las fallas en órdenes de trabajo es la confianza entre áreas.

La siguiente imagen fue fotografiada en un taller entre Work Management y Supply realizado en la faena de Cerro Colorado en Iquique y muestra el proceso que se sigue para la generación de OTs. Esta imagen deja en evidencia que Supply resulta ser un agente transaccional en la visión de la planificación de mantenimientos, y por lo tanto no hay incentivos al trabajo colaborativo.

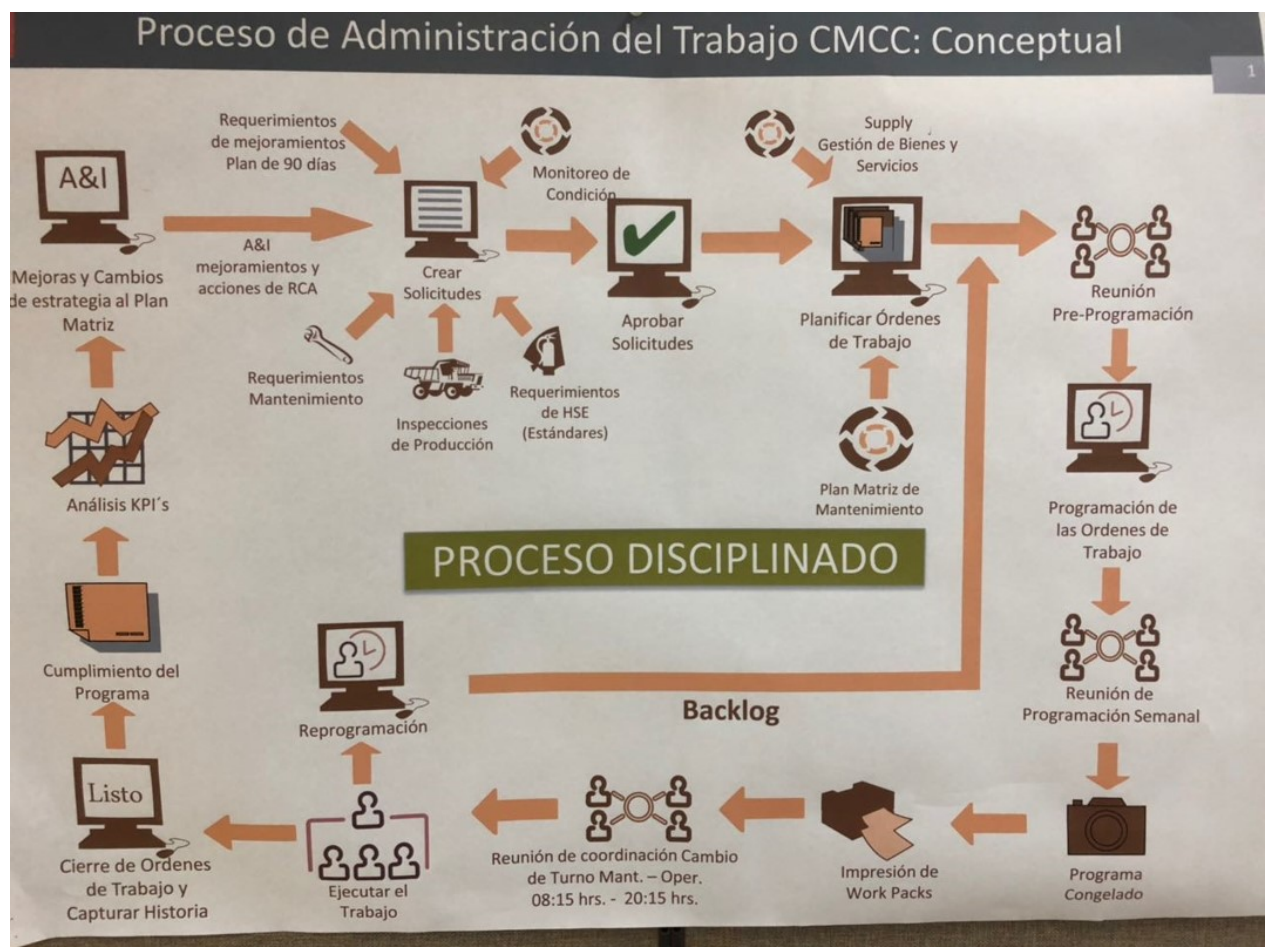


Figura 3: Proceso de Work Management. Fuente: Taller en Cerro Colorado Marzo/2018

WOPIFOT es un KPI compartido que fue concebido con la intención de medir tanto el desempeño de la planificación como del abastecimiento de materiales y con esto fomentar el trabajo colaborativo entre áreas.

0.15. Área de Supply

El área de Supply se divide en 6 sub unidades funcionales. Estas son Category management, Purchasing, Warehouse, Inventory & Logistics (WIL), Supply Innovation, Analytics & Improvement (A&I) y Business Partnership (BP).

0.15.1. Principales funciones

El siguiente cuadro detalla las principales propuestas de valor de cada sub unidad dentro de supply:

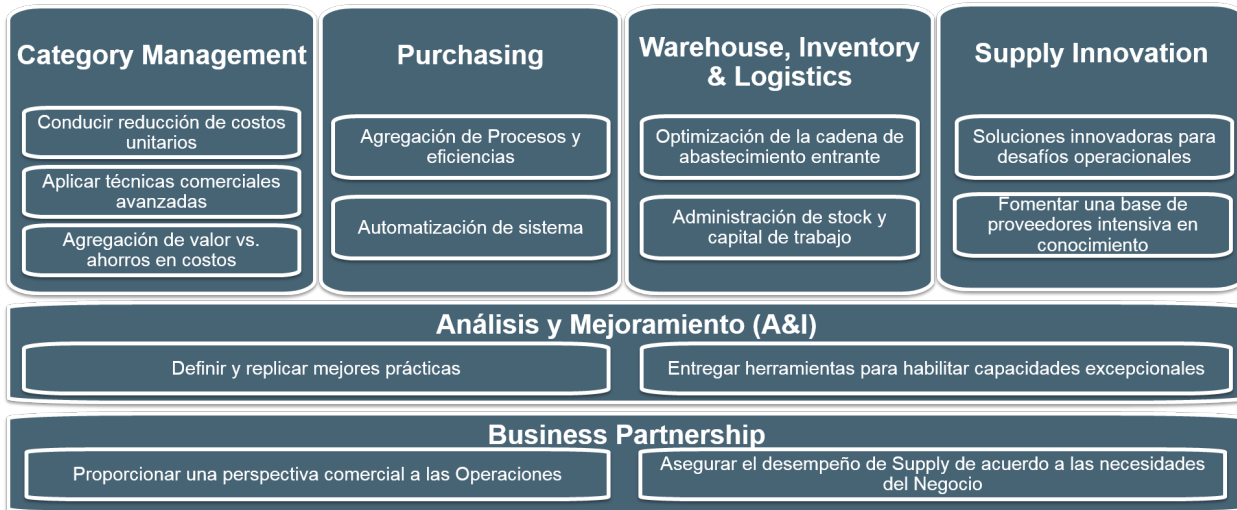


Figura 4: Estructura de Supply

0.15.2. Organigrama

La estructura organizacional de supply se basa en el principio de reporte global y ejecución local, es decir, existe un organismo centralizado a nivel internacional que precede la función de supply, pero cada asset posee un equipo propio de supply que ejecuta sus funciones a nivel local.

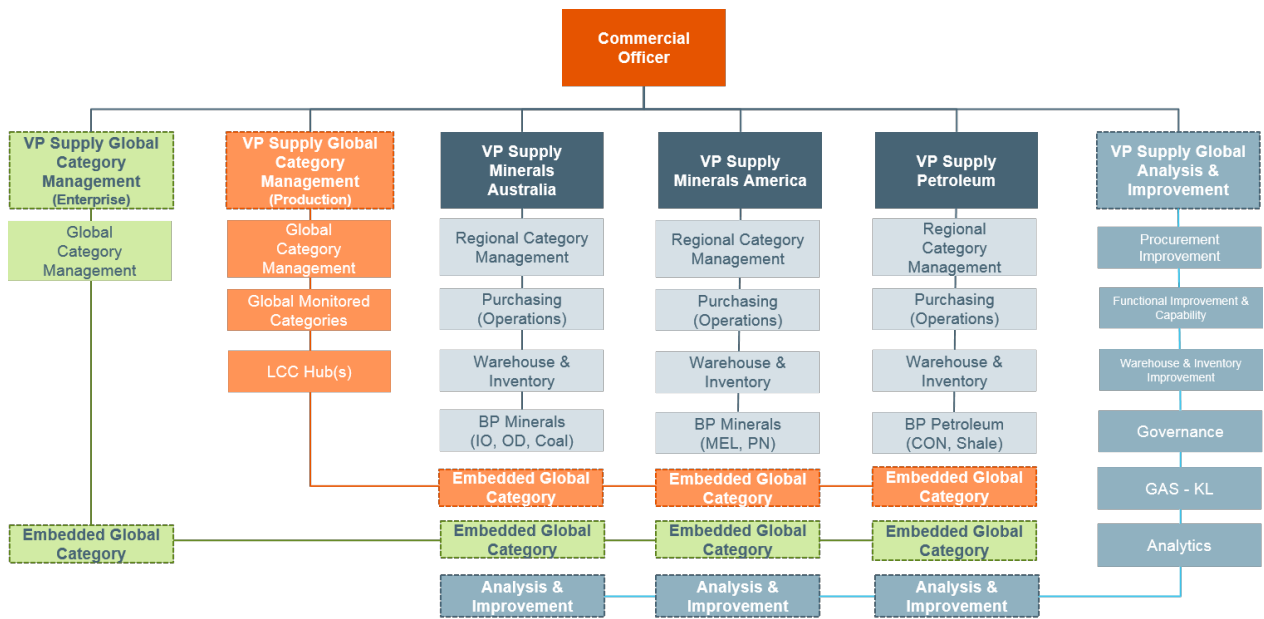


Figura 5: Estructura de Supply

0.15.3. Clientes

El área de supply cumple una función clave dentro del proceso productivo, tanto a la provisión de materiales como de servicios. El principal cliente en términos de transacciones es el área de mantenimiento (Reflejando esto la relevancia del indicador WOPIFOT) y en segundo lugar, el área operacional, recibiendo por parte de supply los insumos necesarios para las operaciones productivas. Junto con esto, supply se encarga también de contratos de servicios, siendo esto una labor transversal a las distintas áreas dentro de la empresa.

0.15.4. Relación con el proyecto de memoria

El proyecto de análisis predictivo del indicador WOPIFOT busca entregar una herramienta al área de planificación y mantenimiento de la mina Spence, de forma dar aviso a los planificadores de órdenes de trabajo que por sus características son difíciles de cumplir por el equipo de supply (ya sea por fechas que exceden el lead time de los proveedores, errores de los mantenedores, u otras muchas posibles causas). Así, las principales áreas beneficiadas por el proyecto son Planificación y mantenimiento, en cuanto tendrán una idea más clara de órdenes de trabajos factibles y Supply, ya que como consecuencia de lo anterior, les permitirá desarrollar mejor la gestión y planificación de materiales.

0.16. Glosario

Work Order (WO): También Orden de Trabajo (OT), registro electrónico de la necesidad de realizar un trabajo de mantenimiento o reparación. Incorpora datos como HH requeridas, Componentes necesarios y fecha de ejecución tentativa.

Basic Start Date (BSD): Fecha tentativa indicada para realizar el trabajo de mantenimiento

Basis Finish Date (BFD): Fecha de término del trabajo de mantenimiento. Corresponde a BSD + Tiempo necesario para realizar el trabajo

Release: Acción de liberar la WO hacia supply. Luego de esto el sistema revisa stock y genera PRs correspondientes. El status de la WO cambia a REL Actual Release Date: Fecha en la que se hace el release de una work order

Scheduled: Calendarizar WO y asignar trabajo a los operadores. El status de la WO cambia a SCH

TECO: Cierre técnico de la WO. Lo realiza un usuario de mantenimiento luego de la ejecución

Status de OTs:

- s/CNF: Sin confirmar realización de mantenimiento en el sistema
- CNF: Confirmado pero sin TECO
- PCNF: Confirmado y cerrado
- NMAT: OT que no requiere materiales
- PRI: OT programada pero que no se pudo hacer

Work Order Type:

- Pm01: WO Programada para al menos 7 días a futuro
- Pm02: WO planificada. Se ejecuta por sistema de acuerdo a una planificación anual
- Pm03: WO de emergencia

Materiales:

- Materiales Críticos: Materiales catalogados en 1SAP y clasificados como V1. Estos materiales son clasificados como críticos, debido a que el no contar con ellos se produciría un impacto material en la producción (Definición de Risk Management).
- Material PD: Materiales catalogados en 1SAP, los cuales no tienen stock y son solo adquiridos cuando el usuario genera un requerimiento para actividades específicas.
- Materiales V1: Materiales catalogados en 1SAP, los cuales mantienen stock dado el alto volumen de consume y rotación. Estos materiales tienen definido un punto de re-orden (ROP) y un máximo (ROQ) los cuales están configurados en 1SAP.

- Materiales ZD: Materiales catalogados en 1SAP, los cuales no tienen stock y son solo adquiridos cuando el usuario genera un requerimiento. Una vez que se recibe el material por bodega (recepción en 1SAP), el costo del material se carga directamente al centro de costo del usuario y el material es entregado físicamente.

Purchasing:

- Purchase Requisition (PR): Requerimiento de compra de materiales. Puede ser generado manualmente o vía sistema.
- Purchase Order (PO): Orden de compra. Documento electrónico de concreción de una PR. Refleja un acuerdo de compra entre la empresa y un proveedor.

Métricas:

- WOPIFOT: Work Order Provisioned In Full On Time. Mide el número de reservas planeadas/no planeadas incluyendo compras directas que provienen de una Orden de Trabajo y que fueron satisfechas a tiempo en todos sus ítems y en donde las Órdenes de Trabajo hayan sido cerradas técnicamente durante el período que se emite el reporte.
- MIFOT: Materials In Full On Time. Se calcula como el ratio entre cantidad de materiales entregado y materiales requeridos en la reserva a la fecha de requerimiento.
- DIFOT: Delivery In Full On Time. Mide si el proveedor entrega los materiales dentro del plazo acordado (cumple el lead time)

0.17. Work Orders (WO) de mantenimiento

Los planes de mantenimiento de equipos mina y planta son ingresados por los planificadores de mantenimiento a módulo Work Management de SAP. Estos ingresos son llamados Work Orders (WO - órdenes de trabajo) Estas WOs contienen el detalles de las actividades a realizar en cada uno de los puntos de trabajo:

- Tipo de Tarea
- Equipo a intervenir
- Tiempo para ejecución
- HH necesarias
- Servicios a utilizar
- Materiales/componentes a utilizar

En lo que respecta a este último punto, la emisión de una WO gatilla una solicitud (vía sistema) de materiales o componentes a utilizar para la mantención. Esta solicitud es recepcionada por área de Bodega - del área de Supply, quien debe asegurar la disponibilidad del material para la WO en la fecha comprometida.

Supply comprende toda la cadena logística: Contratos, Compras, Logística, Inventario, Bodega. La disponibilidad física de materiales en Bodega involucra que toda la cadena funcione.

0.18. Fallas en Work Orders

En un caso ideal, cada orden de trabajo debiese ser cumplida a tiempo, lo que involucra una disponibilidad de materiales completa en el tiempo indicado. Esto en la realidad no se cumple por diversas razones:

- Material stockout
- Tiempo de solicitud para mantenimiento excede lead time de entrega de proveedor
- Proveedor no entrega material en tiempo comprometido
- Contrato para material de reposición automática se ha cancelado/expirado
- Mantenedores no colectan los materiales en la fecha comprometida.

entre otras.

Cada una de esas fallas genera un “miss” en la WO y por lo tanto el mantenimiento no puede ser ejecutado como lo estipula el plan, generando re-trabajo, otras mantenciones no planificadas y consecuencias en disponibilidad de equipos. Un índice que mide disponibilidad de materiales para WO es “WOPIFOT” (Work Order Provisioned In Full On Time). Basta que exista un “miss” para que la WO sea catalogada como fallada en WOPIFOT, es decir, es índice corresponde a un variable binaria que toma el valor de 1 cuando la orden de trabajo se cumple satisfactoriamente en el plazo indicado (es decir, no hubo ninguna falla en todo el proceso) y 0 cuando no. Durante Año Fiscal 2017, las fallas por “misses” tuvo un efecto de aproximadamente 15 % en WOPIFOT. (o dicho de otra forma: “la disponibilidad de materiales para órdenes de trabajo sólo se cumplió en un 85 %)

0.19. Impacto económico de fallas en WO

0.19.1. Disponibilidad de materiales

Actualmente BHP Minerals America no cuenta con un estudio de impacto marginal de WOPIFOT, sin embargo los esfuerzos de mejorar disponibilidad de materiales para mantenimiento se sustentan en la idea del costo asociado a posponer mantenciones o fallas de equipamiento a causa de negligencias en las mismas. Por la dificultad técnica de este análisis y el horizonte de tiempo reducido, la cuantificación del impacto económico queda fuera del alcance de esta memoria.

0.19.2.

0.20. Objetivos

0.20.1. Objetivo General

Desarrollar en base a un modelo de minería de datos y machine learning un sistema de alerta temprana de ordenes de trabajo riesgosas, que permita a la gerencia de suministro de BHP Minerals América articular de mejor forma la cadena de abastecimiento, planificación y ejecución y con esto mejorar el índice WOPIFOT para alcanzar de manera sostenida el 88 % definida como meta a nivel gerencial, lo que también implica mejorar la disponibilidad de materiales para mantenimiento y el nivel de servicio de supply.

0.20.2. Objetivos Específicos

- Desarrollar el levantamiento del proceso de planificación de mantenimiento, diseñando para esto un diagrama BPMN.
- Ordenar, limpiar y seleccionar los datos relevantes para el proyecto, generando también estadísticas descriptivas que permitan un mejor entendimiento del problema y validación del proyecto.
- Determinar el modelo de minería de datos más apropiado para abordar el problema, en base a una validación cruzada utilizando los datos históricos
- Utilizar el modelo seleccionado para realizar predicciones sobre las órdenes de trabajo y asignar una posible causa de falla
- Utilizar una herramienta de visualización como TIBCO Spotfire, para desarrollar un dashboard de uso intuitivo que permita a los planificadores de inventario usar la herramienta predictiva en sus rutinas de planificación semanal

0.21. Resultados esperados y Alcances

0.21.1. Resultados esperados

El proyecto busca desarrollar un modelo predictivo de fallas en las órdenes de trabajo, siendo el entregable un modelo teórico de minería de datos junto con métricas evaluativas del desempeño del mismo, junto con las bases de la incorporación del modelo a un dashboard de visualización mediante el software TIBCO Spotfire.

0.21.2. Alcances

0.21.2.1. Proyecto enfocado solo a la operación minera Spence

Dentro de los dos assets de BHP Minerals América (Escondida y Pampa norte) el proyecto se enfocará en Pampa Norte, en particular en la mina Spence. Esto a razón

de que dado el horizonte de tiempo provisto por la empresa (5 meses), no será posible analizar las características necesarias para replicar el proyecto a nivel de otras faenas.

0.21.2.2. Modelo enfocado en ordenes de trabajo tipo PM01

Este proyecto no considera otros tipos de ordenes de trabajo, dado que el gap detectado en particular en las PM01 es el problema que se quiere abordar.

0.21.2.3. Modelo Predictivo, sin capacidad prescriptiva

El modelo a desarrollar no involucra líneas de acción para resolver las posibles causas de fallas de las Work Orders de mantenimiento. Esto a causa de que tales modelos no son factibles técnicamente con el conocimiento que se tiene actualmente de las causas raíces y los datos disponibles.

0.21.2.4. Diseño e implementación del modelo

El alcance de esta memoria es hasta la implementación del modelo, entregando resultados teóricos (Tales como accuracy, AUC, Kappa), y el diseño del experimento para ser puesto a prueba en las operaciones de planificación de inventario. No se considera dentro del alcance de esta memoria la gestión del cambio necesaria para la incorporación de la herramienta predictiva dentro de las rutinas de Work Management.

Capítulo 1

Capítulo 1: Marco Teórico y Metodología

1.1. Data Mining

El data mining (minería de datos), es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

La minería de datos es una técnica emergente que permite agregar valor a un negocio de distintas maneras: por un lado, resulta un buen punto de encuentro entre los investigadores y objetivos de negocio; por otro, permite el ahorro tanto monetario como en trabajo manual (traducido en tiempo) de análisis estadístico estándar, y puede ayudar a identificar nuevas oportunidades dentro del negocio. Es necesario tener siempre en cuenta que el vínculo con el negocio es crucial, puesto que la data y los descubrimientos que provienen de esta, deben ser validados por expertos del negocio y considerados como un apoyo a la hora de la toma de decisiones más que una forma de automatización de estas.

A continuación se describen los modelos utilizados en el desarrollo de este proyecto de memoria, partiendo desde los modelos básicos de árboles de decisión, que dan paso a la construcción de algoritmos más complejos como son los de bagging y boosting, siendo este último el que finalmente se implementa en el modelo predictivo.

1.1.1. Data Warehousing

El término Data Warehousing introducido por primera vez por Bill Inmon, quien lo define como “una colección de datos orientados a temas, integrados, no-volátiles y variables en el tiempo, organizados para soportar necesidades empresariales” [15].

Ralph Kimball, define un Data Warehouse como “una colección de datos en forma de una base de datos, que guarda y ordena información que se extrae directamente de los sistemas

operacionales y datos externos” [17]

Ambas definiciones ven en común un Data Warehouse como una colección de data organizada y orientada a un problema o negocio en particular y proveniente de distintos sistemas o fuentes de datos.

En relación a los procesos de data mining, es necesaria la previa construcción de un consolidado de data que permita trabajar de forma ordenada el modelo. Esta forma ordenada puede ser guiada mediante la construcción de un modelo estrella. Luego de este diseño teórico, es necesario poblar las tablas mediante un proceso conocido como ETL (Extraction, Transformation, Loading).

1.1.1.0.1 Modelo Estrella

Consiste en una representación multidimensional de datos utilizando la nomenclatura del modelo Entidad-Relación, pero que no involucra los estándares de normalización asociados a la metodología estándar de creación de bases de datos. Está compuesto por una tabla central (Fact) y un conjunto de tablas dimensionales. Cada tupla que compone la tabla Fact posee como identificador el conjunto de llaves primarias de las dimensiones. A su vez, cada tupla posee atributos asociados a las distintas combinaciones posibles de llaves dimensionales. Si alguna de las dimensiones posee entidades asociadas, el modelo posee el nombre de *snowflake*.

1.1.1.1. ETL

Como fue descrito anteriormente, las fuentes de datos necesarias para un proceso de minería de datos suelen provenir de distintos sistemas o fuentes de datos, cada una con un formato propio y distintos entre ellos, además, se deben procesar los datos de manera de ser transformados en información útil para ser cargada en un repositorio. Todo este proceso, llamado ETL [19]

- **Extracción:** Esta etapa consiste en obtener de las distintas fuentes de datos la data necesaria para posteriormente trabajar sobre esta. La etapa de extracción de la data va muy de la mano de las etapas de entendimiento del negocio y entendimiento de la data descritas más adelante en la metodología Crisp-DM [1.2], puesto que depende de los objetivos del negocio y de la interpretación de la data cuales serán los sistemas consultados y la data extraída.

Debido a esto, es necesario una estrategia bien definida de extracción de data, considerando la posibilidad de incorporar iterativamente nuevas fuentes de data y una calendarización adecuada.

- **Transformación:** Esta etapa resulta ser la más compleja del proceso ETL [19]. Esto a causa de diferentes razones, como son la inconsistencia de unidades de medidas entre distintas fuentes de datos, definición de jerarquías entre datos y generación de valores calculados para almacenar en el DW.

Estas dificultades pueden ser abordadas mediante un DSA (Data Staging Area) que se

considera una etapa intermedia donde los datos se cargan unicamente para ser transformados, pero sin ser consolidados aún en el DW definitivo. Esto permite mantener intacta la fuente original de datos en caso de cometer algún error.

- **Carga:** La carga de datos al DW resulta ser la etapa más sencilla [19] dado que generalmente luego de la etapa de transformación, los datos se encuentran en un formato amigable. Consiste en términos prácticos en la consolidación de todas las fuentes (ya incorporadas en el DSA) y es la base de datos con que se trabajará finalmente.

1.1.2. Modelos Predictivos

Cuando se utilizan métodos y tecnologías para detectar tendencias y patrones en los datos recopilados de diversos recursos, las empresas pueden responder a ellos e identificar nuevas oportunidades. Tanto los datos estructurados como los no estructurados se utilizan en el análisis predictivo.

El análisis predictivo es el nuevo mantra para empresas y expertos en data science porque les brinda una ventaja decisiva al momento de la toma de decisiones, ya que permite en base a patrones repetitivos encontrados en data histórica poder actuar proactivamente y abordar los problemas u oportunidades con un mayor sustento teórico.

La construcción de un modelo predictivo permite también automatizar el análisis de data transaccional, y es esta una de las razones por las que resulta ser una solución pertinente al problema que se trabaja en este trabajo de memoria.

1.1.3. Árboles de decisión

Los árboles de decisión son un modelo de minería de datos utilizado como técnica predictiva o de selección. Consisten en una división jerárquica y secuencial de los datos en la que cada división (o nodo) describe gráficamente las posibles decisiones y por lo tanto los resultados de las distintas combinaciones de decisiones y eventos.

Un árbol de decisión se representa gráficamente como un diagrama de flujo donde cada nodo interior (no hoja) denota una prueba en un atributo, cada rama representa el resultado de una prueba y cada hoja (nodo terminal) representa una clase.

La mayoría de los algoritmos utilizados para construir un árbol son variaciones de uno genérico llamado “Greedy algorithm” que básicamente va desde la raíz hacia abajo (Top-Down) buscando de manera recursiva los atributos que generan el mejor árbol hasta encontrar el óptimo global con una estructura de árbol lo más simple posible. Los algoritmos más conocidos son el ID3 el C4.5 [3, 4]

1.1.3.1. Algoritmo ID3

ID3 es un algoritmo no incremental, lo que significa que deriva sus clases de un conjunto fijo de datos de entrenamiento. Un algoritmo incremental revisa la definición del concepto actual, si es necesario, con una nueva muestra o conjunto de datos. Las clases creadas por ID3 son inductivas, es decir, dado un pequeño conjunto de datos de entrenamiento, se espera que las clases específicas creadas por ID3 funcionen para todas las realizaciones futuras. La distribución de las incógnitas debe ser la misma que en los casos de prueba. No se puede probar que las clases de inducción funcionen en todos los casos, ya que pueden clasificar un número infinito de instancias. Tenga en cuenta que ID3 (o cualquier algoritmo inductivo) puede clasificar erróneamente los datos.[4]

Los datos muestrales utilizados por ID3 tienen ciertos requisitos:

- Los mismos atributos deben describir cada ejemplo y tener un número fijo de valores.
- Clases predefinidas: los atributos de un ejemplo ya deben estar definidos, es decir, ID3 no los aprende
- Clases discretas: las clases deben estar claramente delineadas.
- Suficientes registros: dado que se utiliza la generalización inductiva (es decir, no demostrable), debe haber suficientes casos de prueba para distinguir patrones válidos de ocurrencias fortuitas.

¿Cómo ID3 decide qué atributo es el mejor? Se usa una propiedad estadística, llamada ganancia de información. La ganancia mide qué tan bien un atributo determinado separa los datos de entrenamiento en clases específicas. Se selecciona el que tiene la información más alta (la información es la más útil para la clasificación). Para definir la ganancia, primero tomamos una idea de la teoría de la información llamada entropía. La entropía mide la cantidad de información en un atributo.

Dada una colección S de c clases

$$Entropía(S) = \sum_c -p(I) * \log_2 p(I) \quad (1.1)$$

Donde $p(I)$ es la proporción de S que pertenece a la clase I

La ganancia de información sobre el set de datos S en un atributo A se define como

$$Ganancia(S, A) = Entropía(S) - \sum ((|S_v|/|S|) * Entropía(S_v)) \quad (1.2)$$

Donde

\sum es cada valor de v de cada posible valor del atributo A
 S_v es el subconjunto de S para cada atributo de A que tenga un valor v
 $|C|$ es la cardinalidad de un conjunto C

1.1.3.2. Algoritmo C4.5

Este algoritmo fue desarrollado por Ross Quinlan en 1993 es una versión mejorada del algoritmo ID3 antes descrito. Esta versión incorpora las siguientes mejoras

- Permite la utilización de valores continuos y discretos. Para el manejo de valores continuos, el algoritmo crea un umbral para después dividir el atributo en secciones discretas sobre y bajo el umbral.
- Permite trabajar con valores de atributos faltantes. Cuando un atributo corresponde a un null (vacío) el algoritmo utiliza una ponderación de valores y probabilidades que se obtienen directamente de las frecuencias observadas para esa instancia, lo que resulta en un poder predictor mayor. Esta característica es relevante para este estudio dado que los datos presentan en varias ocasiones atributos faltantes.
- El algoritmo construye un árbol de gran tamaño pero concluye con una “poda” de las ramas de forma de generar una visualización más simple. Esta característica es de vital importancia para este estudio, puesto que gran parte del valor entregado a la empresa corresponde a una herramienta de interpretación visual.

1.1.4. Algoritmos de ensamble

Los algoritmos de ensamble (o aditivos) son un método de combinar predictores débiles (con exactitud no muy superior a lanzar una moneda) para crear un predictor más robusto. Dos formas de algoritmos de ensamble son el Bagging y Boosting.

1.1.4.1. Bagging

Bagging es un método para generar múltiples versiones de un predictor. La agregación promedia las predicciones de cada predictor generado en el caso de regresiones y toma el "voto más común." en el caso de clasificación. Estas múltiples versiones de algoritmos débiles se crean en base a un bootstrapping del data set de entrenamiento. [8] En el caso de algoritmos de clasificación, supongamos $h(d, x)$ es un clasificador, por ejemplo un árbol, basado en el set de data D , produciendo una predicción dado el input x . Para hacer bagging sobre d , se generan submuestras mediante bootstrapping de D (d_1, \dots, d_M), cada una de tamaño N , entonces

$$\hat{H}_{bag}(x) = \text{VotoMayoritario}\{h(d_m, x)\}_{m=1}^M \quad (1.3)$$

Este método puede reducir la varianza de predictores inestables como los árboles, pero se pierden las propiedades de interpretabilidad de estos.

Algoritmo Bagging	
Input	M=Número de clasificadores débiles
	$h = \{h_1 \dots h_M\}$ Conjunto de clasificadores débiles.
	N Número de registros del set de datos utilizados para ajuste (entrenamiento)
	$N' < N$, Número de ejemplos generados por bootstrapping
	$d \subset D$ Conjunto de entrenamiento
	$g()$ Función de ensamble
Output	Modelo de ensamble $H_{bag}(X)$.
	para $m = 1 \dots M$
	Generar conjunto d_m seleccionando N' muestras por reemplazo desde d
	Entrenar clasificador h_m sobre el subset d_m
	Agregar h_m al ensamble
	fin
	devolver $\hat{H}_{bag}(x) = g(h_1(x), \dots, h_M(x))$

Tabla 1.1: Algoritmo de Bagging

1.1.4.1.1 Random Forest

Random Forests es un algoritmo de ensamble creado por Leo Breiman en 1994, que mejora la exactitud en la clasificación mediante la incorporación de aleatoriedad en la construcción de cada clasificador débil. Donde a diferencia de lo que sucede con el Bagging en cada split el algoritmo escoge sólo un subconjunto de los predictores, esto es, la aleatoriedad está presente tanto en las muestras de entrenamiento como en el conjunto de atributos.[9]

1.1.4.2. Boosting

El problema del aprendizaje predictivo puede ser caracterizado por un vector de variables explicativas o predictoras de un fenómeno $x = \{x_1, \dots, x_m\}$ y un vector de variable dependiente y . Dada una colección de M instancias $\{y_i, x_i\}$ con $i = 1 \dots M$ observadas, el objetivo es usar esta data para lograr estimar una función de mapeo del vector x en y , de forma de utilizar esta función para predecir instancias donde solo valores de x son observados. Formalmente, se intenta estimar la función $\hat{f}(x) : x \rightarrow y$ que minimiza una función de pérdida $L(y, f)$ sobre la distribución conjunta W de (y, x)

$$\hat{f}(x) = \operatorname{argmin}_{f(x)} E_{y,x} L(y, f(x)) \quad (1.4)$$

Algoritmo Boosting	
Input	M = Número de clasificadores débiles $h = \{h_1 \dots h_M\}$ Conjunto de clasificadores débiles. N Número de registros del set de datos utilizados para entrenamiento $d \subset D$ Conjunto de entrenamiento $g()$ Función de ensamble
Output	Modelo de ensamble $H_{boost}(X)$. inicializar la distribución $W = \{w_1, \dots, w_N\}$ de las muestras para $m = 1 \dots M$ Entrenar modelo h_m sobre d_m y su distribución w_m Evaluar el error $\varepsilon_m \leftarrow P_{x \sim W}(\mathbb{I}(h_m(x) \neq y_i))$ Actualizar la distribución W a partir del error ε_m . fin devolver $H_{boost}(x) = g(h_1(x), \dots, h_M(x))$

Tabla 1.2: Algoritmo de Boosting

Intuitivamente, la idea detrás de los métodos de boosting es que la combinación de varios predictores débiles (weak learners) produce un mejor modelo en términos de capacidad predictiva. Un predictor débil es un algoritmo cuya capacidad predictiva es levemente mejor que un resultado aleatorio, y con baja capacidad de generalización.

Cada predictor débil permite caracterizar reglas locales relativa a pocas variables, que si bien no es suficiente para lograr una buena capacidad predictiva, la combinación de muchos de estos predictores permite generalizar y construir un modelo más robusto. Esta idea es propuesta por Schapire, 1990 y es conocida como "the strength of weak learnability".[10]

1.2. CRISP-DM

CRISP-DM (Cross Industry Standard Process for Data Mining) propone una secuencia de pasos estándar para un proyecto de minería de datos orientado a la empresa. El modelo CRISP-DM cubre las fases de un proyecto, sus respectivas tareas y las relaciones entre estas tareas. En este nivel de descripción, no es posible identificar todas las relaciones; Las relaciones pueden existir entre cualquier tarea de acuerdo con los objetivos, el contexto y el interés del usuario en los datos.

La metodología CRISP-DM contempla el proceso de análisis de datos como un proyecto profesional, estableciendo así un contexto mucho más rico que influye en el desarrollo de los modelos. Este contexto tiene en cuenta la existencia de un cliente que no forma parte del equipo de desarrollo, así como el hecho de que el proyecto no solo no termina una vez que se encuentra el modelo ideal (ya que más adelante se requiere una implementación y mantenimiento), pero está relacionado con otros proyectos, y es necesario documentarlo exhaustivamente para que otros equipos de desarrollo utilicen el conocimiento adquirido y trabajen a partir de él.

El ciclo iterativo de tareas que conforman un proyecto de data mining bajo una metodología CRISP-DM se muestra en la siguiente figura

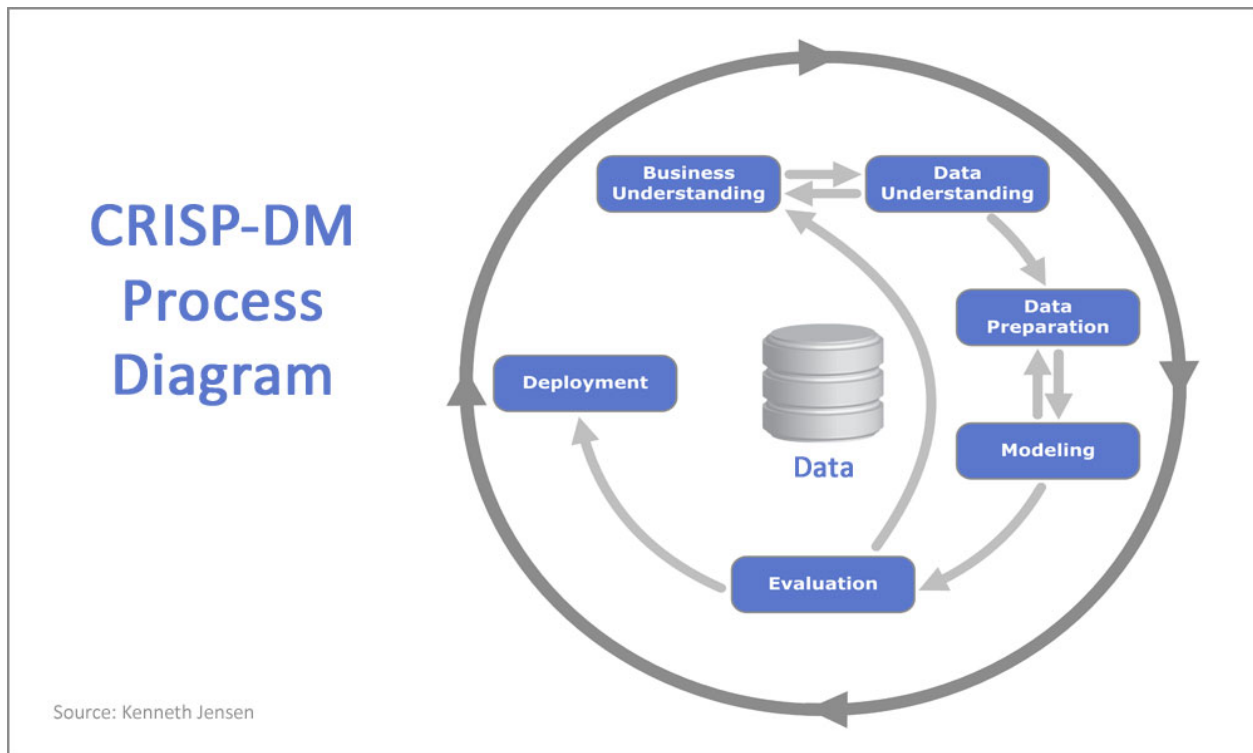


Figura 1.1: Proceso iterativo CRISP-DM

a continuación se describen en forma simple cada una de las 6 fases de un proyecto CRISP-DM [2]

1.2.1. Entendimiento del negocio

Esta fase destaca la importancia del conocimiento del negocio y de los procesos productivos que el proyecto impactará, permitiendo la orientación del proyecto a agregar valor al negocio. Esta fase tiene como output los objetivos de negocio buscados, una lista de los recursos disponibles y un plan preliminar para alcanzarlos.

1.2.2. Entendimiento de los datos

Esta fase tiene como inicio la recolección de los datos relevantes para el proyecto y su posterior análisis, tanto en términos de calidad como descriptivos y exploratorios. En esta fase se seleccionan las variables relevantes para el proyecto.

1.2.3. Preparación de los datos

Se abordan los problemas de calidad de los datos, se consolidan las distintas bases de datos y se crean y transforman variables de forma de poder ser utilizadas posteriormente en el modelamiento. Las tareas incluyen la selección de tablas, registros y atributos, así como la transformación y la limpieza de datos para las herramientas que modelan.

1.2.4. Modelamiento

En esta fase se aplican las técnicas de modelamiento que se consideren pertinentes para el problema detectado y cuyas propiedades aporten el mayor valor al negocio.

1.2.5. Evaluación

Esta etapa considera el testeado del modelo en relación a los datos, utilizando alguna de las diversas técnicas disponible, como Cross-Validation. Se evalúa también el proyecto en términos del negocio. Un objetivo clave es determinar si hay alguna cuestión importante de negocio que no haya sido considerada suficientemente. Al final de esta fase, se debería obtener una decisión sobre la aplicación de los resultados del proceso de análisis de datos.

1.2.6. Despliegue

Luego de la construcción y validación del modelo y del proyecto en general, es importante diseñar un plan de despliegue que incorpore métricas de seguimiento del avance del proyecto implementado, direcciones de mantenimiento y de actualización en caso de ser necesarias.

1.3. Metodología y plan de trabajo

En esta sección se hace una descripción breve de las metodologías planteada para lograr los objetivos antes mencionados.

1.3.1. Entendimiento del Negocio

Tarea	Entregable
Definición del proceso de planificación de materiales para mantenimiento, identificando variables, actores y oportunidades de mejora.	Diagrama BPMN del proceso
Definición del objetivo a alcanzar o meta que se busca con el proyecto, identificando métricas clave.	Documento con objetivos y criterios de falla/éxito
Definir terminología relevante al proyecto, con miras a documentar el proyecto para futura replicabilidad	Glosario
Análisis de impacto, costo/beneficio del proyecto (¿por qué es importante el proyecto?)	Documento de análisis de impacto
Análisis de fallas: Falsos positivos y falsos negativos	Análisis FMEA

1.3.2. Entendimiento de los datos

Tarea	Entregable
Recolección de bases de datos relevantes	Lista y locación de los datos
Descripción de los datos	Forma y volumen de datos
Exploración de los datos	Informe de estadística descriptiva
Análisis de calidad de los datos: identificar outliers, errores, vacíos. Proponer soluciones posibles.	Informe de calidad de datos

1.3.3. Preparación de los datos

Tarea	Entregable
Selección de variables relevantes. Criterios de selección apuntando al objetivo de negocios.	Listado de variables relevantes
Limpieza de datos: rectificación de valores dudosos, estandarización de formatos, manejo de outliers, evaluar implementar propuestas de soluciones definidas anteriormente.	Bases de datos limpia y estandarizada
Integración de bases de datos: Cruce de tablas de datos relevantes	Base de datos consolidada
Construcción y transformación de datos: Construcción de nuevas variables relevantes en base a los datos utilizados, transformación de variables para incorporar en el modelo.	Base de datos preparada para implementación del modelo predictivo

1.3.4. Modelamiento

Tarea	Entregable
Documentación del modelo elegido, argumentar en base a criterios de negocio	Documento argumentativo del modelo escogido.
Describir metodología de testeo del modelo, utilización de partición de los datos en entrenamiento y prueba y nuevos datos para validar el funcionamiento en términos del negocio. Definir criterios de éxito	Metodología de testeo.
Construcción del modelo, ajuste de parámetros	R-File del modelo de data-mining completo.
Evaluación del modelo en base a un testeo simple en datos de testeo y K-Folds Cross Validation	Matriz de confusión, precisión.

1.3.5. Evaluación

Tarea	Entregable
Evaluación de los resultados obtenidos con respecto al objetivo de negocios. Impacto de los resultados en el negocio	Documento, Presentación para stakeholders

1.3.6. Despliegue

Tarea	Entregable
Proyecto de interacción entre modelo analítico-predictivo en R y dashboard TIBCO Spotfire y el ERP SAP para predicción en tiempo real de WO	Líneas de acción para implementación del dashboard

1.4. Levantamiento de procesos

Previo al proceso CRISP-DM es necesario realizar un levantamiento del proceso de supply para la entrega de materiales requeridos mediante una orden de trabajo de mantenimiento. Para esto, se utilizan herramientas de observación y entrevistas con las distintas partes involucradas en el proceso, y así obtener la estructura general del proceso que se utiliza. Luego se utiliza la notación estándar BPMN para efectos de visualización del proceso y como herramienta útil para identificar las causas raíces de la falla en las WO.

1.5. Modelamiento y aplicación de algoritmos

En cuanto al proceso de minería de datos como tal, en primera instancia se obtienen los datos históricos de las órdenes de trabajo para mantenimiento de equipos del sistema SAP con el que cuenta BHP. Mediante un proceso de ETL (Extracción, Transformación y Carga), de forma de obtener los datos necesarios y en un formato correcto para la posterior aplicación de los modelos de minería de datos.

Luego de esto, se aplican distintos modelos de data mining, utilizando un segmento de los datos históricos como datos de entrenamiento de los modelos y validando estos mediante un proceso de validación cruzada, obteniendo así las respectivas matrices de confusión generadas por cada modelo como métrica de comparación.

Finalmente, la validación de los modelos se hace en conjunto con las personas pertinentes dentro de la empresa (Mantenimiento y abastecimiento).

Para llevar a cabo esta metodología se plantea utilizar el software de análisis estadístico R por diversas razones. i) ya que cuenta con diversidad de paquetes de minería de datos que permiten aplicar y testear diferentes algoritmos de machine learning sin la necesidad de construirlos desde cero, y además cuentan con funcionalidades de testeo (como K-Folds Cross Validation) que permiten tanto testear el modelo y su precisión y también ajustar los hiperparámetros sin recurrir necesariamente a prueba y error u otras heurísticas. ii) R es un lenguaje popular entre entendidos de estadística y data-scientists, por lo que la transferencia del conocimiento posterior a la implementación del modelo resulta más sencilla. iii) R es suficientemente robusto como para trabajar con grandes volúmenes de data y procesar algoritmos iterativos, y en caso de ser necesario, es compatible con clusters de computadores. iv) R cuenta con soporte masivo tanto por parte de la propia comunidad de R como por otras comunidades de ayuda como son Stack Overflow o RBloggers. v) R es gratis y masivamente disponible.

Para efectos de la visualización y la interfaz entre la herramienta y el usuario, se utiliza TIBCO Spotfire dado que es el software oficial de visualización en BHP, por lo que se cuenta con las capacidades por parte de los empleados de interactuar con los reportes generados a través de este software.

Capítulo 2

Capítulo 2: Desarrollo del proyecto de memoria

2.1. Levantamiento del proceso de Supply

2.2. El proceso de Supply para mantenimiento

2.2.1. Metodología de levantamiento de procesos

Para diagramar y subsecuentemente entender mejor el proceso que rige a supply se utiliza la metodología de Business Process Model and Notation (BPMN).

BPMN es una notación estándar de procesos de negocios basada en diagramas de flujo, orientada a dar soporte a la gestión de procesos de negocios. BPMN otorga una notación gráfica intuitiva tanto para usuarios técnicos como usuarios de negocio.

En cuanto a la plataforma utilizada para el desarrollo de los diagramas BPMN, se utiliza MS Visio 2013.

2.2.2. Macro Proceso de supply

El proceso de Supply se gatilla con la generación de una orden de trabajo (OT, o WO en inglés) por parte de Work Management.

Una vez definidos los materiales necesarios, dependiendo del tipo de material (Materiales V1 con alta rotación de inventario, o materiales PD que no mantienen inventario) el proceso continúa hacia distintas áreas. En ambos casos, una vez es aprobada la orden de trabajo por los planificadores de inventario, se genera una PR (purchase requisition) que a su vez genera

una reserva de material en el sistema.

En caso de materiales PD, es necesario que la función de purchasing gestione las cotizaciones con distintos proveedores y genere una orden de compra que garantiza que el proveedor entregará los materiales o directamente a las bodegas en faena, o alternativamente al operador logístico (Ferrovial).

En caso de materiales en stock, una vez generada la reserva, se descuentan los materiales del stock guardado en el sistema y estos quedan listos para entregar a Work Management en la fecha requerida.

Finalmente, una vez que el proveedor o el operador logístico entrega los materiales a bodega, estos se registran en sistema y quedan listos para ser entregados para la ejecución de las tareas definidas en la orden de trabajo.

A continuación se presenta el macro proceso de supply en formato BPMN:

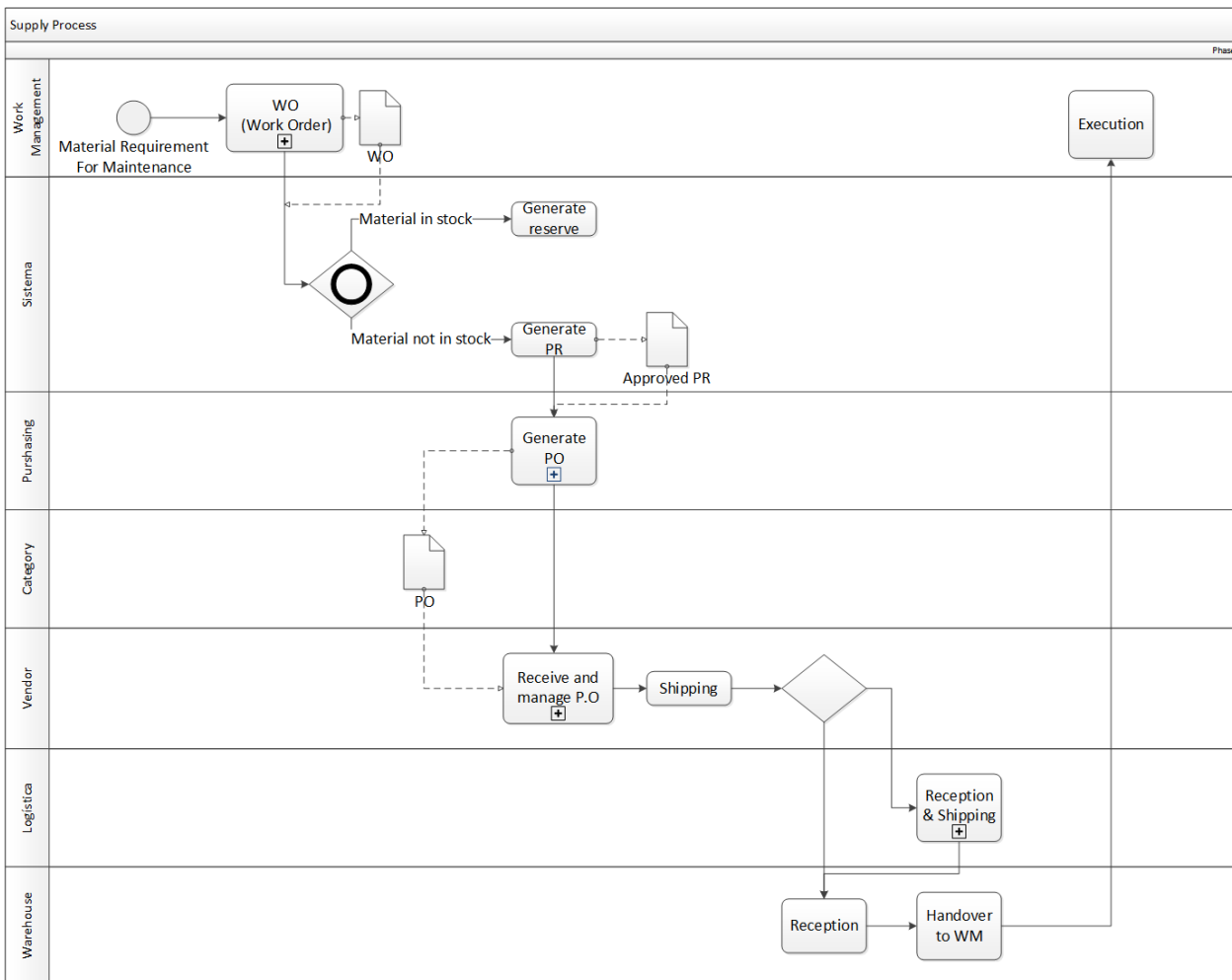


Figura 2.1: Macro proceso de supply para mantenimiento

2.2.3. Proceso de generación de Work Orders

Los planes de mantenimiento de equipos mina y planta son ingresados por los planificadores de mantenimiento a módulo Work Management de SAP. Estos ingresos son llamados Work Orders (WO - órdenes de trabajo) Estas WOs contienen el detalles de las actividades a realizar en cada uno de los puntos de trabajo:

- Tipo de Tarea
- Equipo a intervenir
- Tiempo para ejecución
- HH necesarias
- Servicios a utilizar
- Materiales/componentes a utilizar

En lo que respecta a este último punto, la emisión de una WO gatilla una solicitud (vía sistema) de materiales o componentes a utilizar para la mantención. Esta solicitud es recepcionada por área de Bodega - del área de Supply, quien debe asegurar la disponibilidad del material para la WO en la fecha comprometida.

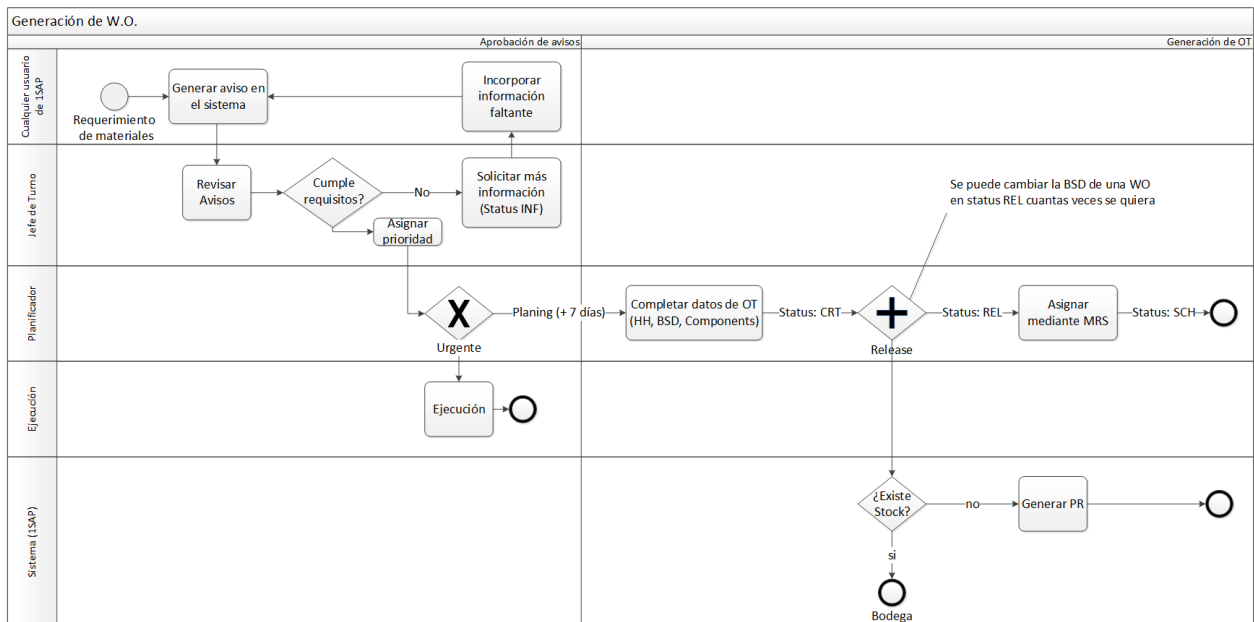


Figura 2.2: Proceso de generación de órdenes de trabajo

Existen 3 tipos de OT, PM01 son OTs programadas con al menos una semana de anticipación, PM02 son programadas por sistema de acuerdo a una estrategia de mantenimiento de equipos anual, y PM03 son OTs de emergencia.



Figura 2.3: Tipos de OT

2.2.4. Proceso de Compras

El proceso de compras comienza una vez la PR es aprobada por el área de inventario (o bien automáticamente en caso de materiales de alta rotación o catálogo electrónico).

Una vez sucede esto, dependiendo del monto asociado a la compra, el proceso es tomado por el área global (en caso de ser menor a 50.000USD) o por el área local en caso contrario.

Luego de esto, nuevamente dependiendo del monto se trabajan distintos tipos de licitaciones. Si el monto es superior a 250.000USD, se procede a la creación de un FPL (que funciona de forma similar a un contrato), y en caso contrario se procede a la creación de una orden de compra (PO) común.

Por acuerdo de nivel de servicio, el área de compra cuenta con 3 días para concretar la creación de una orden de compra luego de recibida la PR.

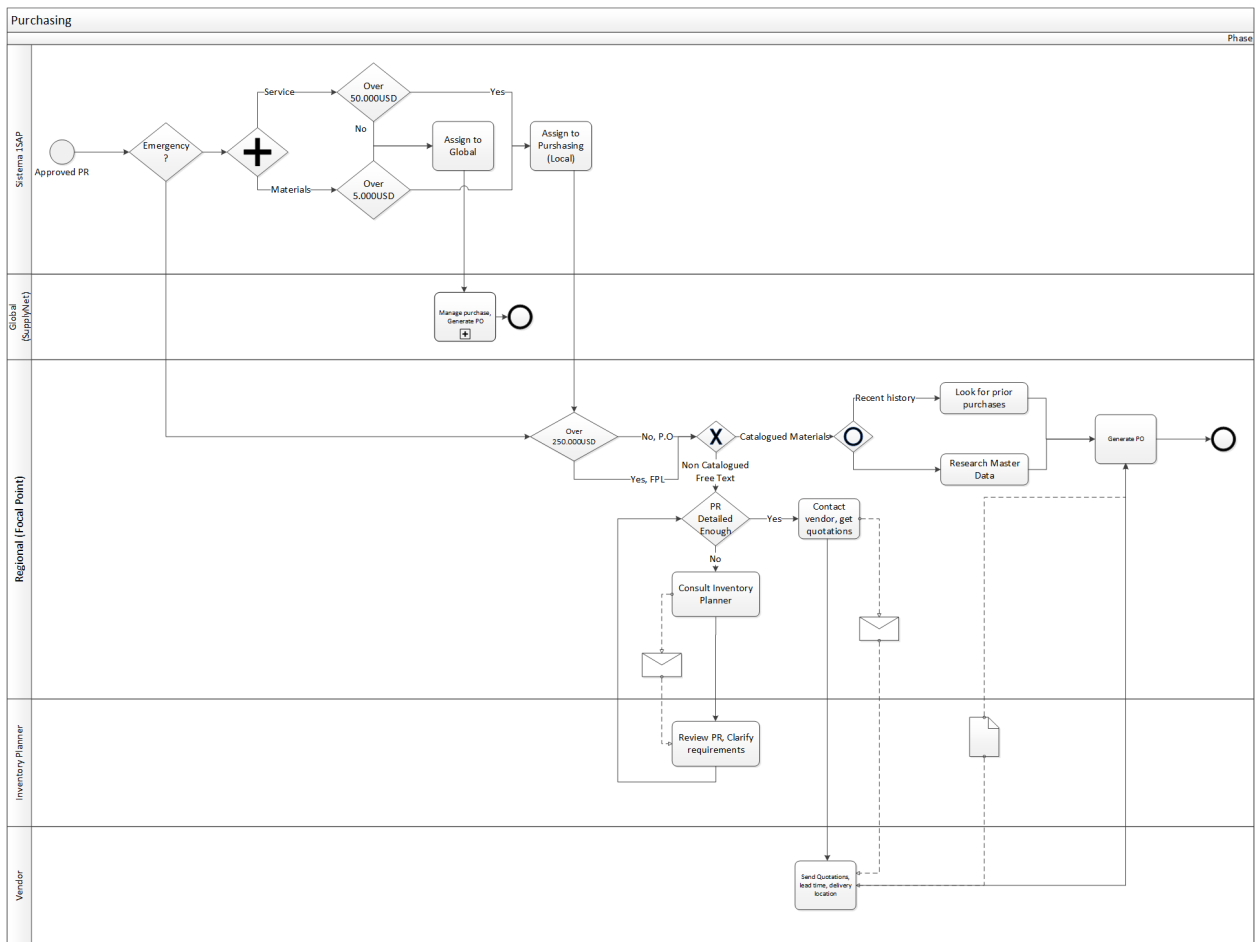


Figura 2.4: Proceso de compra de materiales

2.2.5. Métricas de Supply

2.2.5.1. WOPIFOT

WOPIFOT mide el número de reservas planeadas/no planeadas incluyendo compras directas que provienen de una Orden de Trabajo y que fueron satisfechas a tiempo en todos sus ítems y en donde las Ordenes de Trabajo hayan sido cerradas técnicamente durante el período que se emite el reporte. El indicador WOPIFOT mide el número total de reservas con stock/no-stock entregadas a tiempo, es decir rebajar las reservas antes de cumplir con la fecha y hora de requerimiento que se indica en la reserva que proviene de la WO y en todos sus ítems para la OT. EL indicador cuenta positivo cuando la reserva ha sido rebajada previo a la fecha de calculated requirement date. Esto es representado en forma binario con 0 si para algún material no se cumple el total o 1 si toda la OT es cumplida a tiempo.

2.2.5.2. MIFOT

MIFOT es una métrica similar a WOPIFOT pero que evalúa a nivel de cada material solicitado mediante la orden de trabajo si este fue rebajado de bodega en la fecha de requerimiento definida inicialmente al crear la OT. Para cada reserva, se calcula el ratio entre la cantidad rebajada de bodega y la cantidad total requerida.

2.2.5.3. DIFOT

Este indicador mide la eficacia de los proveedores simplemente contrastando la fecha de compromiso del proveedor vs la fecha real de entrega. Si la fecha real es menor o igual a la fecha de compromiso, el indicador marca 1, en caso contrario, 0.

2.2.6. Impacto del proyecto para Supply

Si bien el impacto económico de una mejora en el indicador WOPIFOT no está cuantificado, y hacerlo está fuera del alcance de esta memoria, es posible asumir que existe un costo infinito de una detención de un equipo (en esta caso por falta de materiales para mantenimiento), ya que un equipo detenido no genera producción.

2.2.7. Impacto de posibles fallas

Dado que se trata de un modelo predictivo, si consideramos las fallas relacionadas con la matriz de confusión, como falsos positivos y falsos negativos, existen dos posibles situaciones que podrían impactar en el trabajo y gestión asociada a la herramienta predictiva:

2.2.7.1. Falsos Positivos

Este caso significa que el modelo clasifica como OT propensa a fallar, una orden de trabajo que por su construcción no debería tener problemas a lo largo de su desarrollo, lo que significaría HH de trabajo injustificadas para el planificador de inventario en términos de revisión y gestión sobre materiales que no deberían presentar problemas. Este tipo de error no debería generar mayor impacto en trabajo extra dado que el proceso de revisión en el sistema generaría un falso positivo si es que el material clasificado como riesgoso ya está en bodega disponible para su rebaja, y esta revisión no toma más que minutos mediante el sistema SAP.

La cantidad de falsos positivos generados por el modelo se puede reducir significativamente al incluir un cruce de datos con los registros de stock, eliminando la alerta a aquellos materiales que se clasifican como fallas pero que sin embargo se encuentran disponibles en bodegas.

2.2.7.2. Falsos Negativos

En este caso el modelo clasifica como no riesgosa una OT que si tiene implícitamente propensión a fallas. Esto no genera trabajo adicional y no presenta impacto en relación al caso alternativo de no utilizar el modelo predictivo. Sin embargo, es deseable que no existan este tipo de error ya que mientras menos falsos negativos, mayor cantidad de OT riesgosas se pueden gestionar para mejorar el nivel de servicio.

2.3. Consolidación de la base de datos

Para desarrollar el modelo predictivo es necesario primero construir un Data Warehouse con todas las bases de datos necesarias consolidadas. Las bases de datos disponibles son:

Database	Volumen	Cantidad de Variables
WOPIFOT	18928	13
MIFOT	330553	5
ONQUA	340080	7
Reporte RCA	2090	147

Tabla 2.1: Volumen y variables de las bases de datos disponibles

La base de datos WOPIFOT contiene información de las OTs entre los meses de Junio 2017 y Abril 2018 y el resultado del indicador WOPIFOT. A continuación se detallan las variables consideradas en este reporte

Variable	Descripción
Región	Región donde se creó la OT. En este caso es siempre Minerals Americas
Operación	Spence, Cerro Colorado o Escondida
Work Order Number	Numero único de la OT
Work Order created By	Usuario quien creó la OT
Work Order Type	Tipo de OT (PM01, PM02, PM03, PM04 o PM06)
Planner Group	Grupo de planificación que creó la OT
Main Work Centre	Centro de trabajo para quien se crea la OT
Release Date	Fecha de Release de la OT
Priority	Prioridad asignada a las operaciones contenidas en la OT
Creation Day	Fecha de creación de la OT
Basic Start Date	Fecha en que se estima se realizaran las operaciones de mantenimiento
TECO	Fecha de cierre técnico
WOPIFOT	Indicador de resultados {0,1}

Tabla 2.2: Variables de WOPIFOT

Para el caso del reporte del indicador MIFOT, este contiene la información de cada OT a nivel de material, es decir, la lista de materiales requeridos en cada OT y el resultado del

indicador MIFOT para cada material. Las variables contenidas en esta base de datos son las siguientes:

Variable	Descripción
Work Order Number	Numero único de la OT
Material	Código único de material
Material Name	Nombre del material en formato estándar
Calculated Reservation Requirement Date	Fecha para la cual se hace la reserva del material
MIFOT	Indicador de resultados {0-100 %}

Tabla 2.3: Variables de MIFOT

Oniqua es un software de gestión de inventario utilizado en BHP para la gestión de los activos necesarios para mantenimiento de los equipos mineros. Este software define ciertas variables sobre cada material de forma de llevar registro del plan de optimización de inventario. una de las variables más relevantes es el MRP Type, que define el tipo de material (PD, V1, ZD, X0, ZD) y con esto su utilización dentro de los planes de inventario. Esta información se encuentra detallada en el glosario (0.16).

Variable	Descripción
Material	Código único de material
Business Impact	Impacto en el negocio de un material específico
MRP Type	Tipo de material de acuerdo a la planificación de inventario
Planta	Planta en la que el material tiene las características anteriores
Mes	Mes de planificación
Año	Año de planificación
KEY	llave única de oniqua para cada material

Tabla 2.4: Variables de Oniqua

El área de A&I desarrolló una herramienta de análisis de causa raíces (RCA) que permite asignar una causa de falla a cada material que presenta un miss a nivel de MIFOT, y de esta forma gestionar reactivamente la cobertura de materiales y la planificación de inventario.

La base de datos originada por el reporte de RCA contiene registros desde Junio de 2017 hasta Abril de 2018, y dispone de todas las variables consideradas en las bases de datos anteriores, un set de variables relacionadas con el proceso de purchasing (Relativas al Vendor, características de Ordenes de compras (PO), Requisitos de compra (PR)) y variables dummy asociadas a cada una de las causas raíz identificadas (Ver Anexos)

La base de datos de RCA no se utilizará en la consolidación de la base de datos principal utilizada como data histórica del modelo, sin embargo, si se usará para una segunda etapa de la herramienta predictiva, intentando detectar patrones en las causas raíces que permitan proyectarse a futuro y con esto sugerir una posible causa de falla en las órdenes de trabajo futuras.

La base de datos principal, y que se utilizará más adelante para entrenar un modelo, se construye a partir del siguiente esquema:

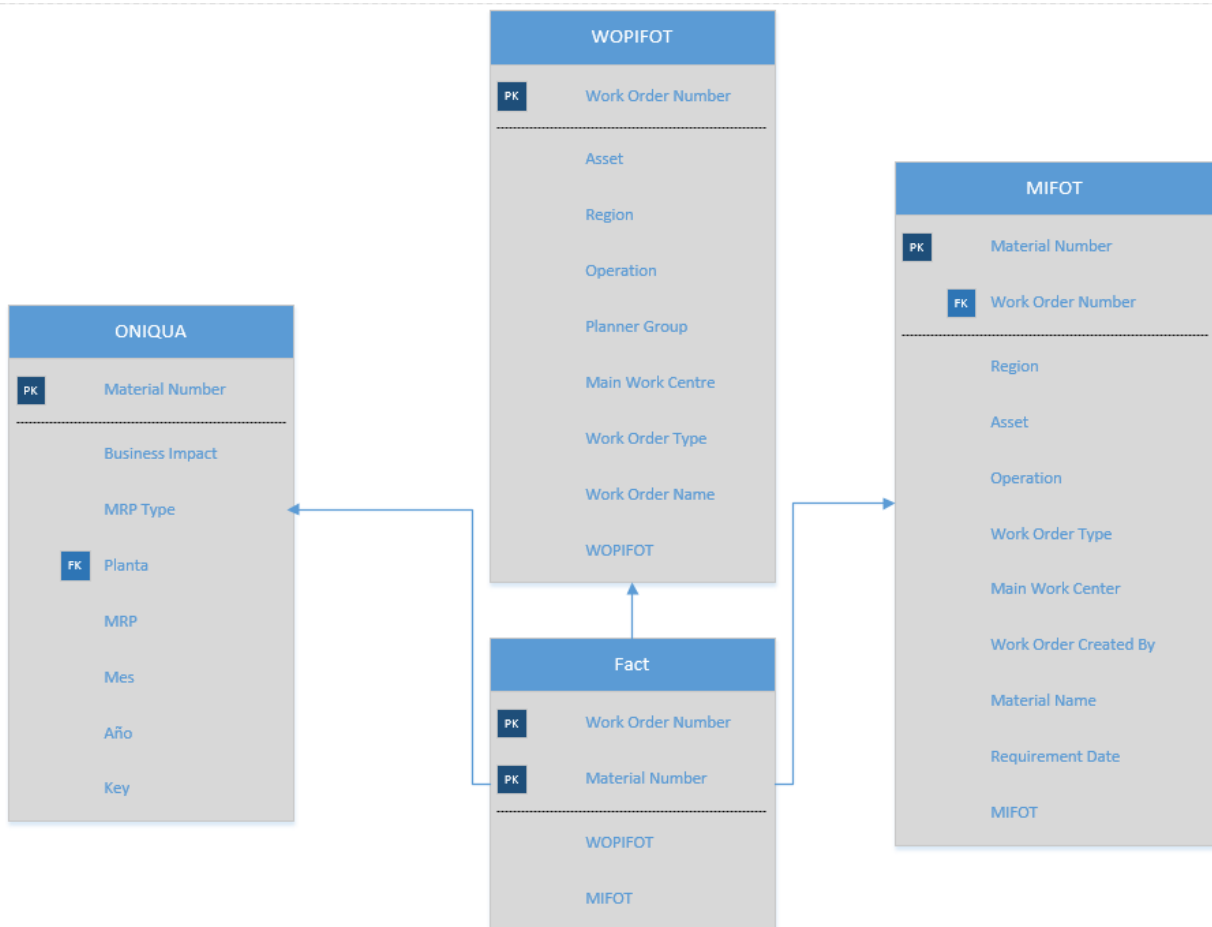


Figura 2.5: Esquema del Data Warehouse a construir

2.4. Caracterización de los datos históricos de WOPIFOT y MIFOT en PM01

En lo que sigue, se hace un análisis estadístico básico de las métricas de resultados WOPIFOT y MIFOT de manera de apoyar el entendimiento de la data y con esto dar las bases para el posterior modelamiento.

2.4.1. Variables

Una vez consolidadas las bases de datos mencionadas anteriormente, el repositorio final tiene la siguiente estructura:

Database	Volumen	Cantidad de Variables
Data Work Orders	46814	21
Data WO PM01 Spence	10613	21

Tabla 2.5: Volumen y variables del Data Warehouse final

Así, el volumen de datos final con el que se trabajará es de 10613 registros, al ser consideradas solo ordenes de trabajo del tipo PM01 para minera Spence.

2.4.1.1. Variables Cualitativas

En cuanto a las variables cualitativas disponibles en la base de datos, luego de aplicados todos los filtros relativos a los alcances de este proyecto, es decir, que la operación sea solo Spence y Work Order Type solo PM01, los detalles de las variables son los siguientes:

Variable	Levels	n	%	\sum %
Region	Minerals Americas	10613	100.0	100.0
	all	10613	100.0	
Operation	Spence	10613	100.0	100.0
	all	10613	100.0	
Work.Order.Type	PM01	10613	100.0	100.0
	all	10613	100.0	
Planner.Group	B701/B29	19	0.2	0.2
	B701/B81	119	1.1	1.3
	B701/B82	955	9.0	10.3
	B701/B83	2	0.0	10.3
	B701/B84	6417	60.5	70.8
	B701/B85	2453	23.1	93.9
	B701/B86	215	2.0	95.9
	B701/B87	321	3.0	98.9
	B701/B88	101	0.9	99.9
	B701/B93	8	0.1	100.0
	B701/B94	3	0.0	100.0
WO.Priority	all	10613	100.0	
	Sched Any Rem Cap	4	0.0	0.0
	Sched Day Defined	6381	60.1	60.2
	Sched in Period	4228	39.8	100.0
Business.Impact	all	10613	100.0	
	A-Major Production Stopper	18	0.2	0.2
	B-Minor Production Stopper	60	0.6	0.7
	C-Major Maintenance Impact	141	1.3	2.1
	D-Minor Maintenance Impact	1897	17.9	19.9
	E-No Impact	7530	71.0	90.9
	Not assigned	967	9.1	100.0
MRP.Type	all	10613	100.0	
	ND	7	0.1	0.1
	PD	1447	13.6	13.7
	V1	8113	76.4	90.1
	X0	79	0.7	90.9
	Not assigned	967	9.1	100.0
all	all	10613	100.0	

Tabla 2.6: Variables Cualitativas

2.4.1.2. Variables Temporales

En cuanto a las variables temporales (fechas), los detalles son los siguientes:

Variable	n	Min	q ₁	\tilde{x}	\bar{x}	q ₃	Max	#NA
Work.order.release.date	10613	2017-05-31	17393	17454	2017-10-27	17541	2018-04-19	0
Work.Order.Created.D	10613	2017-06-01	17380	17434	2017-10-06	17508	2018-04-18	0
Basic.Start.Date	10613	2017-06-06	17437	17506	2017-12-10	17593	2018-05-03	0
Work.Order.TECO.date	10613	2017-06-11	17454	17515	2017-12-17	17602	2018-05-07	0

Tabla 2.7: Variables temporales

2.4.1.3. Transformación de variables

En lo que procede, se explica la transformación y creación de 4 nuevas variables. La primera de ellas es la variable *Time to execute* ("tte"), entendida como el horizonte de planificación que usaron los planificadores de inventario al crear la Work Order. Se calcula de la siguiente forma:

$$ttr = BasicStartDate - WorkOrderCreationDay \quad (2.1)$$

Luego, se crea la variable *Time to release* ("ttr"), diseñada para mostrar el tiempo que toma la OT en ser liberada al sistema (lo que da visibilidad a supply para desencadenar el proceso de abastecimiento), es la diferencia entre la fecha de liberación y la fecha de creación y se calcula de la siguiente forma:

$$ttr = WorkOrderReleaseDate - WorkOrderCreationDay \quad (2.2)$$

La variable *Time from today* ("tft") refleja la antigüedad de la OT en el sistema es la diferencia entre la fecha del día actual y la fecha de creación.

$$tft = Today - WorkOrderCreationDay \quad (2.3)$$

Finalmente, se genera una transformación sobre la variable *Material Name*, cuyo formato estandar permite extraer la familia de materiales al que corresponde cada material, tomando la primera palabra de cada nombre, por ejemplo "*SEAL,RING,23.52MM ID*" > "*SEAL*". Esto se hace para las 500 familias de materiales más comunes en el dataset, dejando al resto dentro del grupo ".ºtros"

2.4.1.4. Variables Cuantitativas

Construidas las variables anteriormente mencionadas, se tienen 3 nuevas variables cuantitativas descritas a continuación:

Variable	n	Min	q ₁	\tilde{x}	\bar{x}	q ₃	Max	s	IQR	#NA
ttr	10613	-1	1	3	20.6	23	284	38.2	22	0
tte	10613	0	16	43	65.3	99	324	62.2	83	0
tft	10613	2	22	51	71.8	106	329	63.9	84	0

Tabla 2.8: Variables Cuantitativas

2.4.2. Estadísticas descriptivas

2.4.2.1. WOPIFOT

La data histórica del KPI WOPIFOT se obtiene directamente de una transacción del sistema SAP y es actualizada manualmente cada semana por analistas del área de supply. La base de datos de WOPIFOT consta de 18928 registros contenidos entre Junio 2017 y Abril 2018, y 13 variables descritas en la Tabla 3.2.

La evolución mensual del indicador WOPIFOT en los meses considerados en el dataset se muestra en el gráfico siguiente:

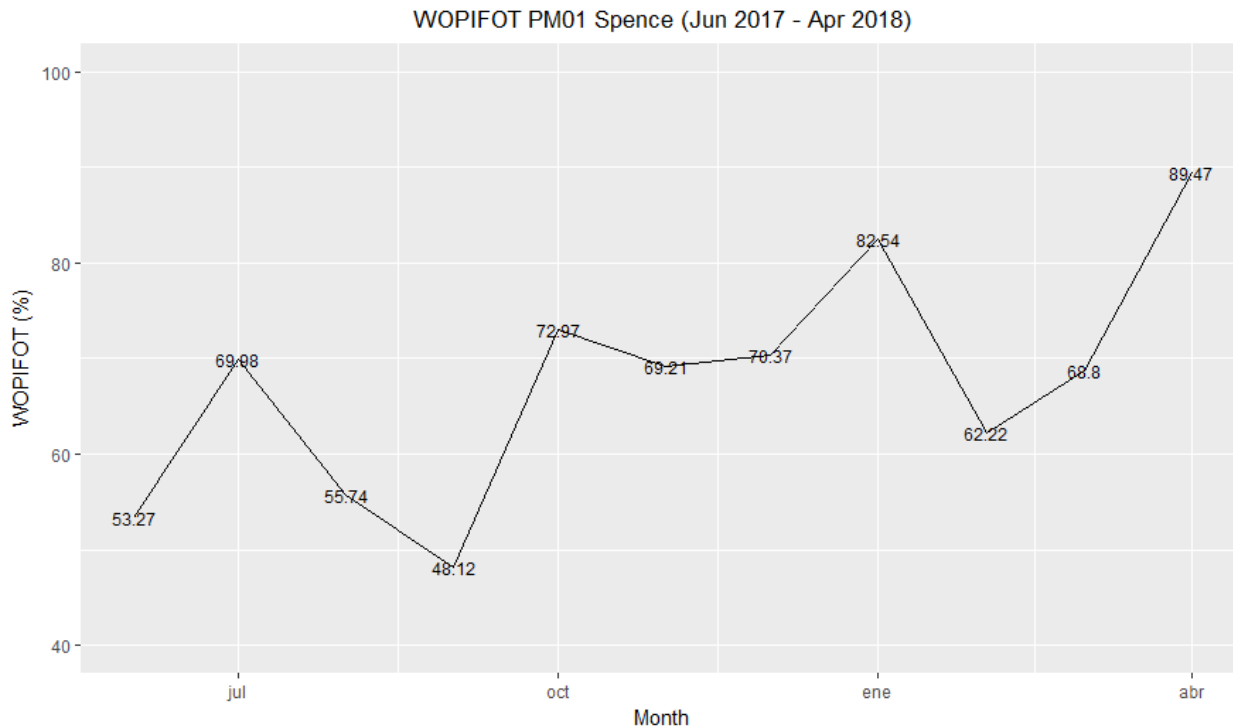


Figura 2.6: Evolución mensual WOPIFOT (6/17-4/18)

De lo anterior se puede notar que existe cierta tendencia al alza en el indicador, pero aún así, en promedio existe un gap de aproximadamente 30% entre el nivel de servicio real y el ideal de 100%, y un gap de aproximadamente 18% en comparación a la meta mensual de 88%.

Al ver el indicador WOPIFOT desagregado por área (Mina, Planta, SSOO) se puede notar que tanto el área de Mina como Planta tienen comportamientos similares, a diferencia de Servicios Operacionales, pero cuyo comportamiento es despreciable para este análisis puesto que la cantidad de OT que generan es mínima.

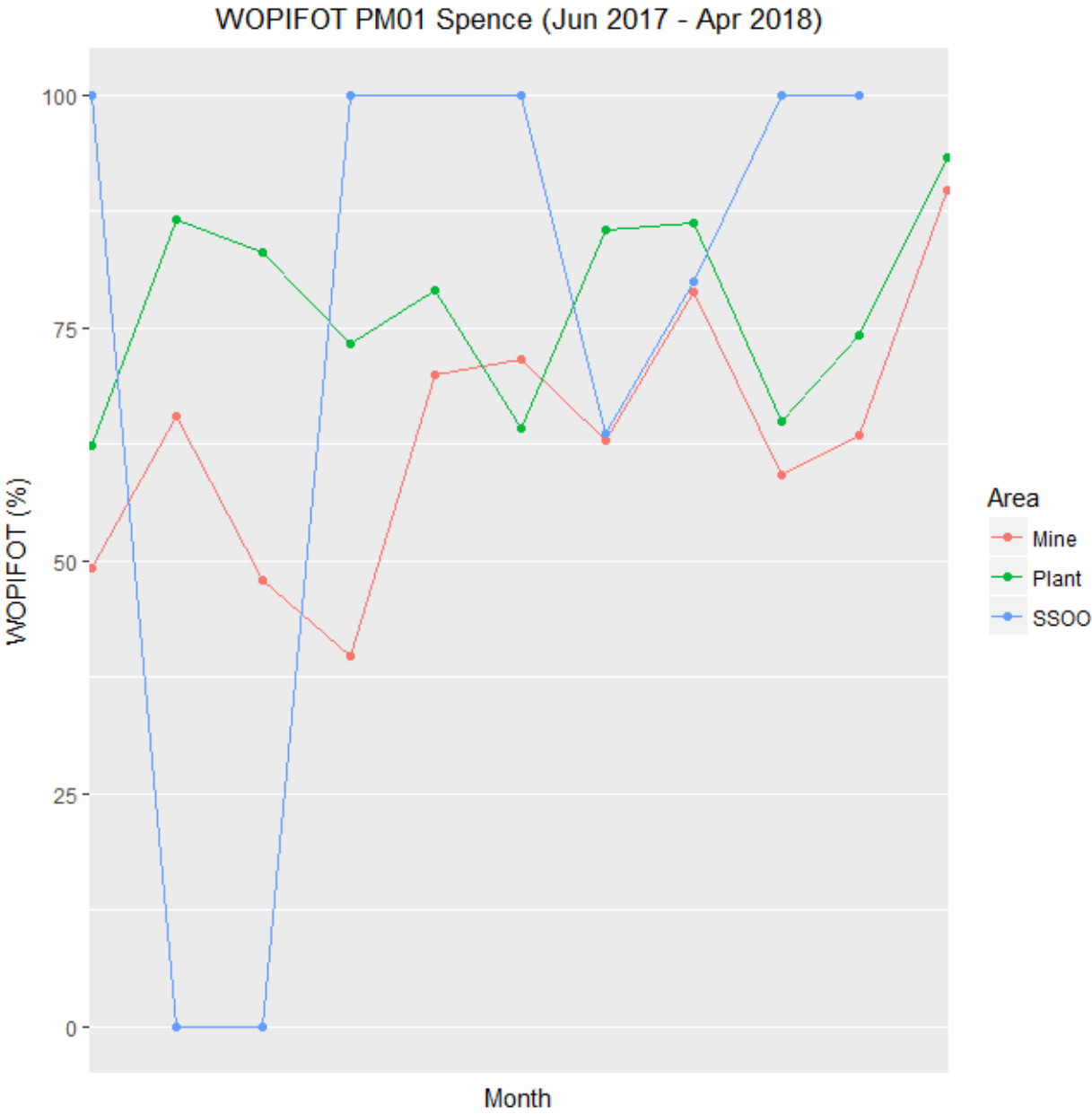


Figura 2.7: Evolución mensual WOPIFOT por área (6/17-4/18)

2.4.2.2. MIFOT

La evolución mensual del indicador MIFOT en los meses considerados en el dataset se muestra en el gráfico siguiente:

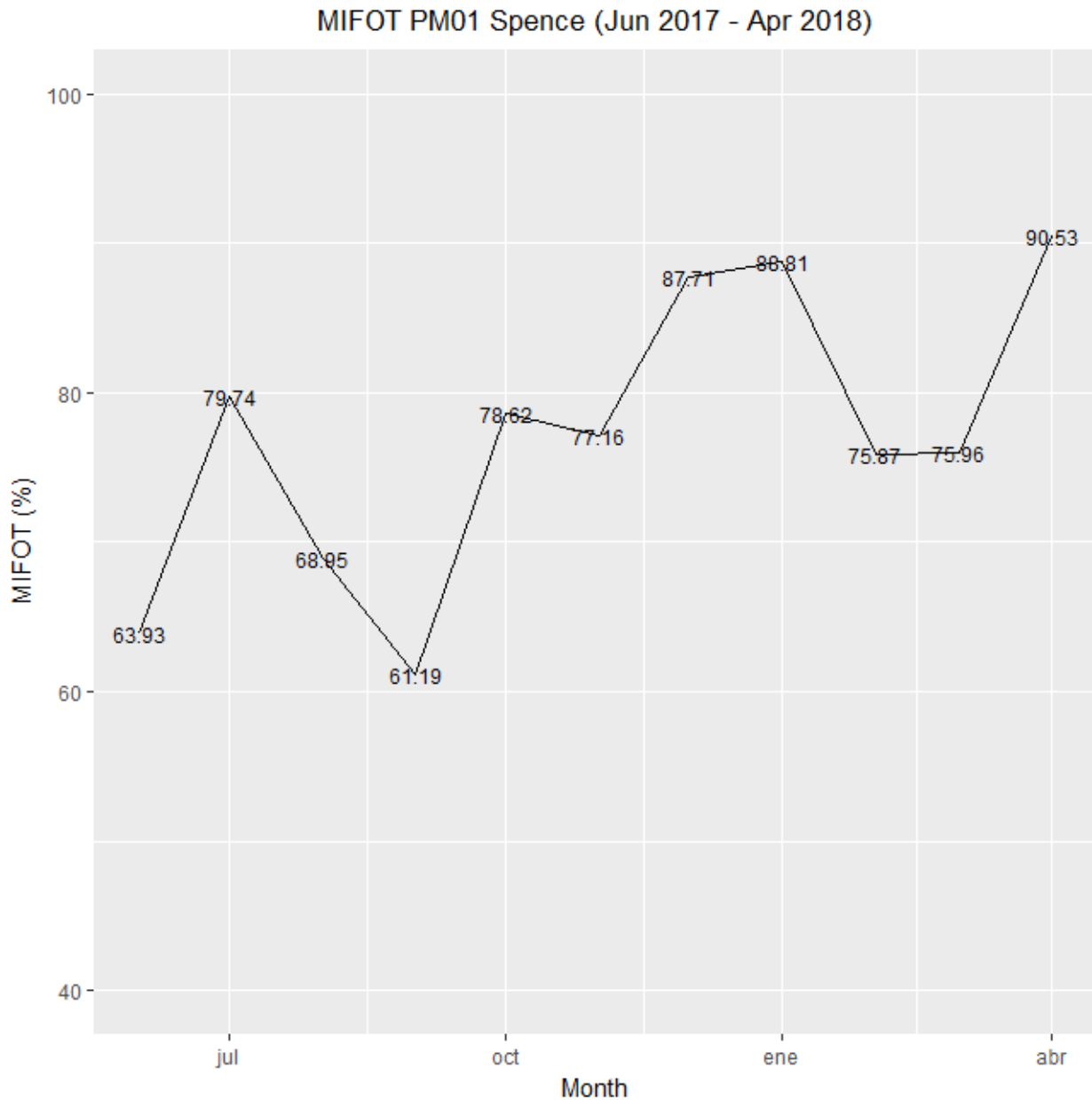


Figura 2.8: Evolución mensual MIFOT (6/17-4/18)

Contrastando los dos gráficos anteriores, se puede notar que MIFOT es consistentemente superior a WOPIFOT todos los meses del set de datos muestral, lo que respalda la idea de que existen causas distintas a la cobertura de materiales que impactan en el indicador de nivel de servicio (aunque de menor impacto en el indicador que la cobertura). Son estas fallas las que diferencian el modelo predictivo de un modelo de stock que solo considera la cobertura de materiales y asume ciertos los leadtime de proveedores, y hacen de un modelo predictivo una

herramienta más completa (aunque también complementaria) para la gestión proactiva de la cadena de suministro.

2.4.3. Desbalance de las clases en los datos

Dado que como se vio anteriormente, el indicador WOPIFOT oscila mayoritariamente al rededor del 70% mensual, existe un desbalance en la variable dependiente del modelo a entrenar. Esto implica que un predictor que clasifique todas las órdenes de trabajo como potencial hit, tendrá en promedio un accuracy de 70%, lo que hace de esta medida una mala aproximación del desempeño del modelo predictivo

Es por esto, que para medir la performance del modelo predictivo, se propone usar las métricas AUC y el Kappa de Cohen para complementar y robustecer el análisis. [12, 13]

2.5. Modelo predictivo como solución al problema planteado

Si bien es evidente que un modelo predictivo basado en machine learning es en teoría capaz de indicar con anticipación, en base a patrones históricos, si una orden de trabajo presenta características o una configuración particular que es propensa a ser una falla a nivel de WOPIFOT, y con esto gestionar proactivamente la cadena de suministro por parte de Supply, hay otra razón por la que resulta una solución útil.

Otra posible aproximación a una solución es un análisis estadístico estático para caracterizar las distintas variables que afectan en el desempeño de la métrica WOPIFOT, sin embargo esto tiene la limitación de que no genera un modelo que analice automáticamente las OTs ni tampoco se adapta a los cambios en los comportamientos humanos o en los patrones históricos como si lo hace un modelo de machine learning.

El hecho de tener un modelo que analice automáticamente cada orden de trabajo que sea creada por Work Management (y que por lo tanto aparezca en el sistema), permite dar mayor visibilidad a la demanda que debe soportar supply.

No existe en Supply BHP MinAm una plataforma que permita una interacción rápida con las ordenes de trabajo por parte de los planificadores de inventario, por lo que la solución propuesta permite tener un punto centralizado para visualización de OTs y junto con esto la posibilidad de focalizar el trabajo solo en las que presenten riesgos indicados por el modelo predictivo.

2.6. Modelo predictivo multi-etapa

Para abordar el problema de las fallas en el nivel de servicio de supply hacia mantenimiento, se propone crear un modelo predictivo basado en machine learning que sea capaz de detectar proactivamente componentes en las órdenes de trabajo cuyo comportamiento histórico revele que existe una probabilidad alta (mayor a 50 %) de que presente algún problema o falla a lo largo de la cadena de suministro.

Para ello se propone un modelo de dos etapas, donde el input es la data histórica del indicador WOPIFOT.

La primera etapa del modelo consiste en un clasificador que permita separar las órdenes de trabajo entre riesgosas y no riesgosas, en base a las características propias de cada OT y de los componentes (materiales) que son requeridos para el trabajo de mantenimiento. Esto permitiría a los planificadores de inventario hacer gestión proactiva en un set menor de órdenes de trabajo que el total, reduciendo también el trabajo manual de análisis y pudiendo enfocar el tiempo en la gestión.

Una segunda etapa del modelo, permite detectar patrones históricos en las combinaciones

de variables en las OT al analizar el reporte de RCA y sugerir una posible causa de falla en las órdenes clasificadas anteriormente como riesgosas. Esta etapa del modelo es de menor relevancia que la primera, puesto que está diseñada solo para guiar el trabajo mediante una sugerencia de la posible falla, pero no reemplaza el trabajo de análisis de las OT anteriormente clasificadas como potenciales fallas.

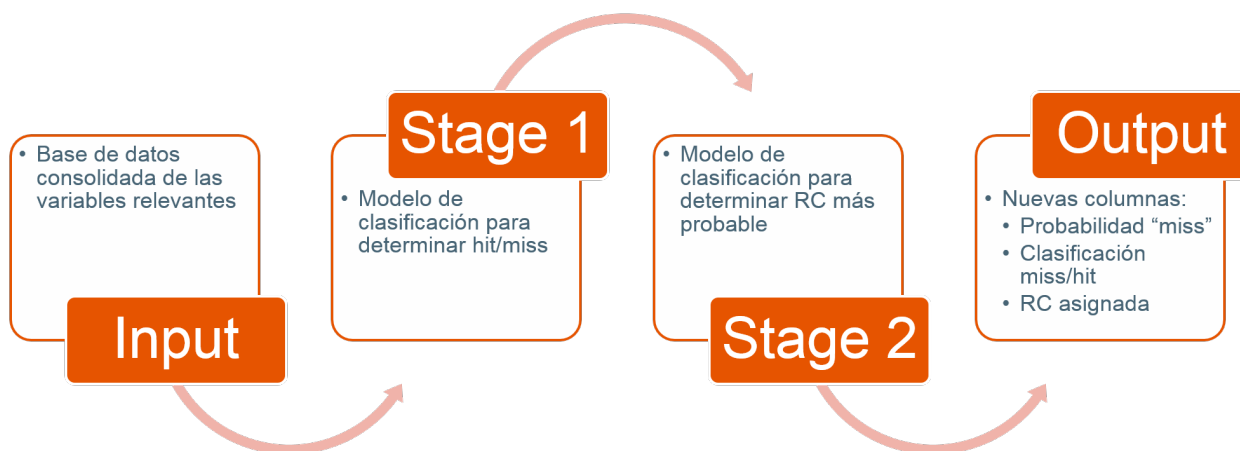


Figura 2.9: Concepto de modelo predictivo multietapa

2.6.1. Selección del algoritmo a utilizar: Challenge de modelos machine learning

En esta sección se comparan 3 modelos para verificar cual de ellos provee mejores resultados en términos de distintas métricas. Se entrenan 3 distintos modelos de aprendizaje de máquinas; Árbol de clasificación, Random Forest, y Gradient Boosting Trees. Los resultados se muestran en la tabla siguiente:

No Information Rate	0.6325				
Algoritmo	Accuracy	Sensitivity	Kappa	AUC	Tiempo
Arbol de Clasificación	59.9 %	29.8 %	7,7 %	0.695	1.686s
Random Forest (n.tree=100)	74.5 %	83.4 %	45.0 %	0,686	1.33s
Random Forest (n.tree=1000)	74.2 %	80.5 %	51.8 %	0,685	177.48s
Random Forest (n.tree=10000)	74.5 %	83.4 %	45.0 %	0,686	627.00s
GBT (n.tree=100)	63.25 %	40.5 %	10.5 %	0.557	1.48s
GBT (n.tree=1000)	69.97 %	50.0 %	23.6 %	0.627	3.61s
GBT (n.tree=10000)	86.4 %	63.21 %	61.48 %	0.783	95.5s

Tabla 2.9: Comparación de resultados de modelos

Cabe destacar que dado que los modelos basados en árboles incorporan en el algoritmo la selección de variables, se consideran todas las variables disponibles en el entrenamiento de los modelos.

Las métricas presentadas en la tabla anterior surgen de un proceso de validación cruzada (K-Folds Cross Validation) con $K = 5$. Es decir, se realiza una partición aleatoria de los datos para entrenamiento y se deja el resto de los datos para testeo, proceso iterativo que se realiza 5 veces con distintas particiones de los datos para finalmente entregar un promedio de las métricas de desempeño de los modelos.

Otra métrica a considerar es el tiempo de entrenamiento de los modelos. En este caso particular, dado que se trata de un volumen pequeño de datos, no es una variable relevante para el proyecto, sin embargo, considerando las oportunidades de escalabilidad de este proyecto, en set de datos más grandes podría jugar un papel fundamental dependiendo de los requerimientos del negocio.

Se puede notar que el modelo de árbol de clasificación es el que presenta peor performance, sin embargo, es el único interpretable visualmente y esto da luces algunos criterios de clasificación.

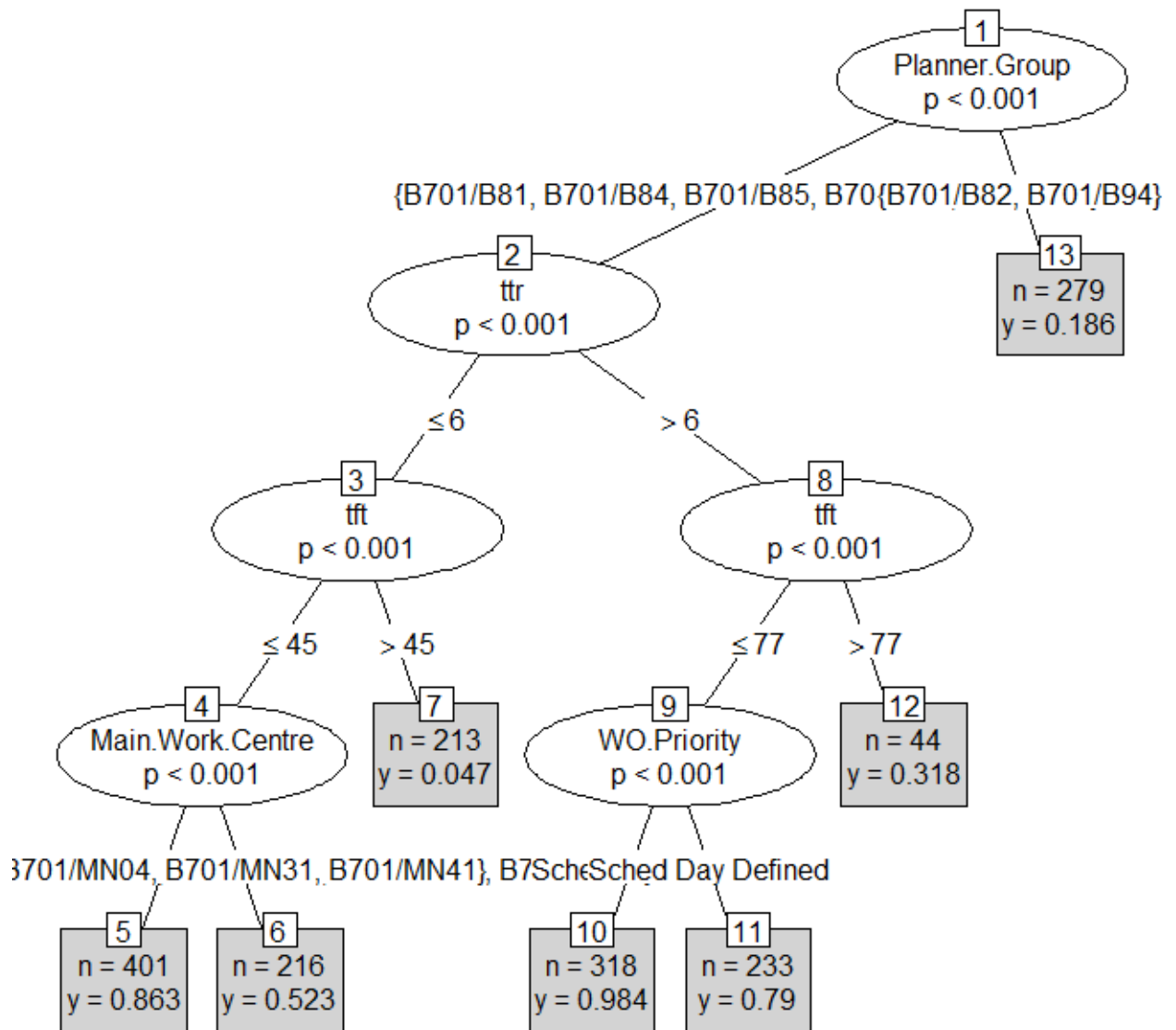


Figura 2.10: Árbol entrenado con todas las variables disponibles

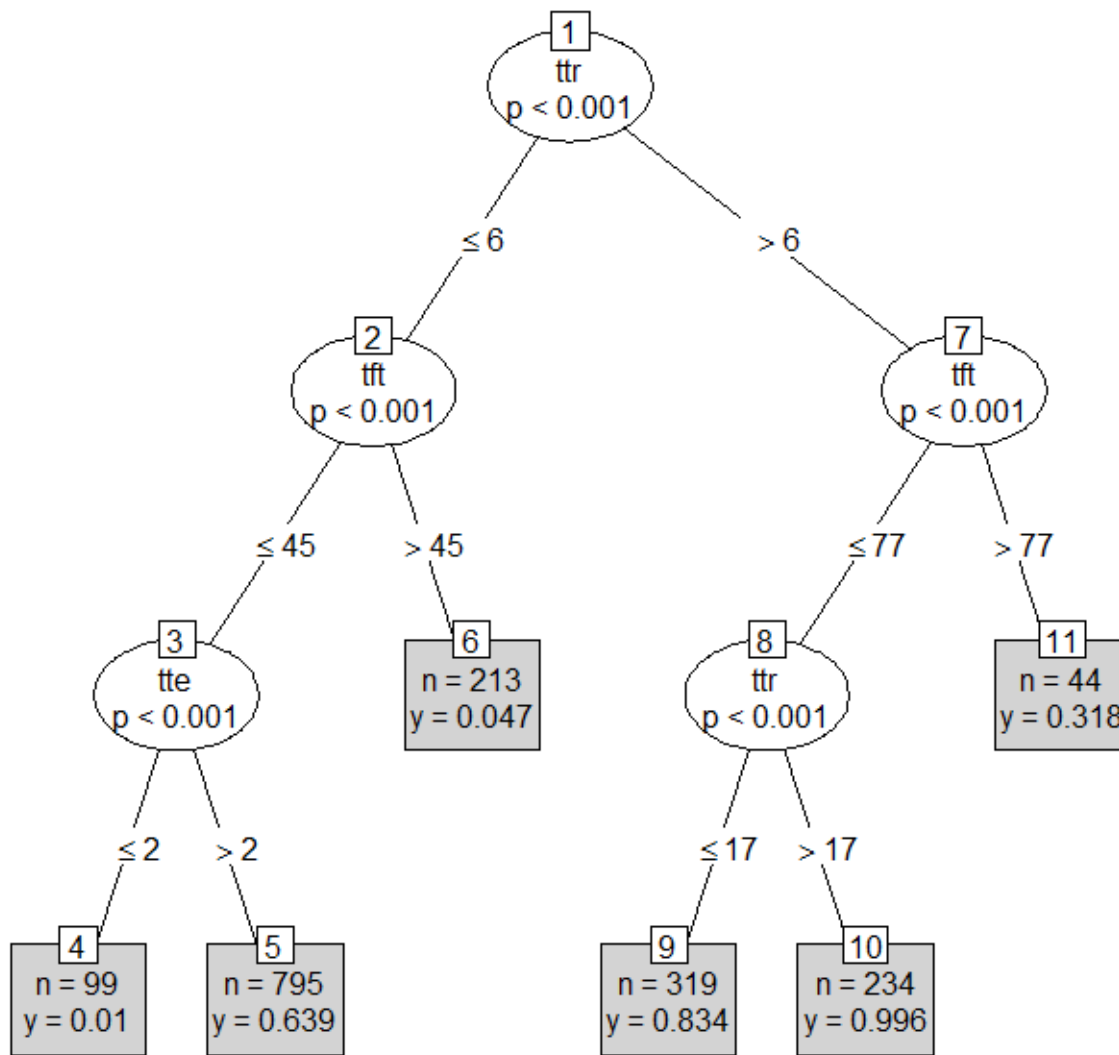


Figura 2.11: Árbol entrenado con variables continuas (numéricas)

En el caso del modelo de random forest, es posible notar que la performance no mejora con un número de árboles mayor. No así con el modelo de boosting que mejora sustancialmente con una mayor cantidad de algoritmos de aprendizaje débiles en el ensamble.

Finalmente, considerando las métricas de desempeño y de tiempo (pensando en la posible escalabilidad del proyecto), el algoritmo a utilizar es el de GBT.

2.7. Construcción del modelo final

2.7.1. Etapa 1: Predicción de Hit o Miss

El modelo a utilizar es el de Gradient Boosting Trees (GBT), implementado vía el software R, RStudio utilizando el paquete de funciones gbm. [11]

Se asume una distribución de tipo Bernoulli dado que la variable dependiente es de carácter binomial.

El paquete Caret permite obtener los hiperparámetros óptimos para un modelo GBT mediante la validación cruzada e iteración de los parámetros. En este caso, los hiperparámetros obtenidos son:

Hiperparámetro	Valor
Shrinkage	0.005
Interaction depth	3
Min. Observaciones	20

Tabla 2.10: Hiperparámetros de GBT obtenidos mediante Caret para etapa 1.

2.7.2. Resultados teóricos

Una de las buenas propiedades de los árboles de decisión es que incorporan en el algoritmo la selección de variables, atributo que se extiende a los modelos de ensamble basados en árboles, en particular el modelo de Gradient Boosting Trees.

Una vez ajustado el modelo como se detalla anteriormente se tienen los siguientes resultados en relación a la importancia relativa de las variables.

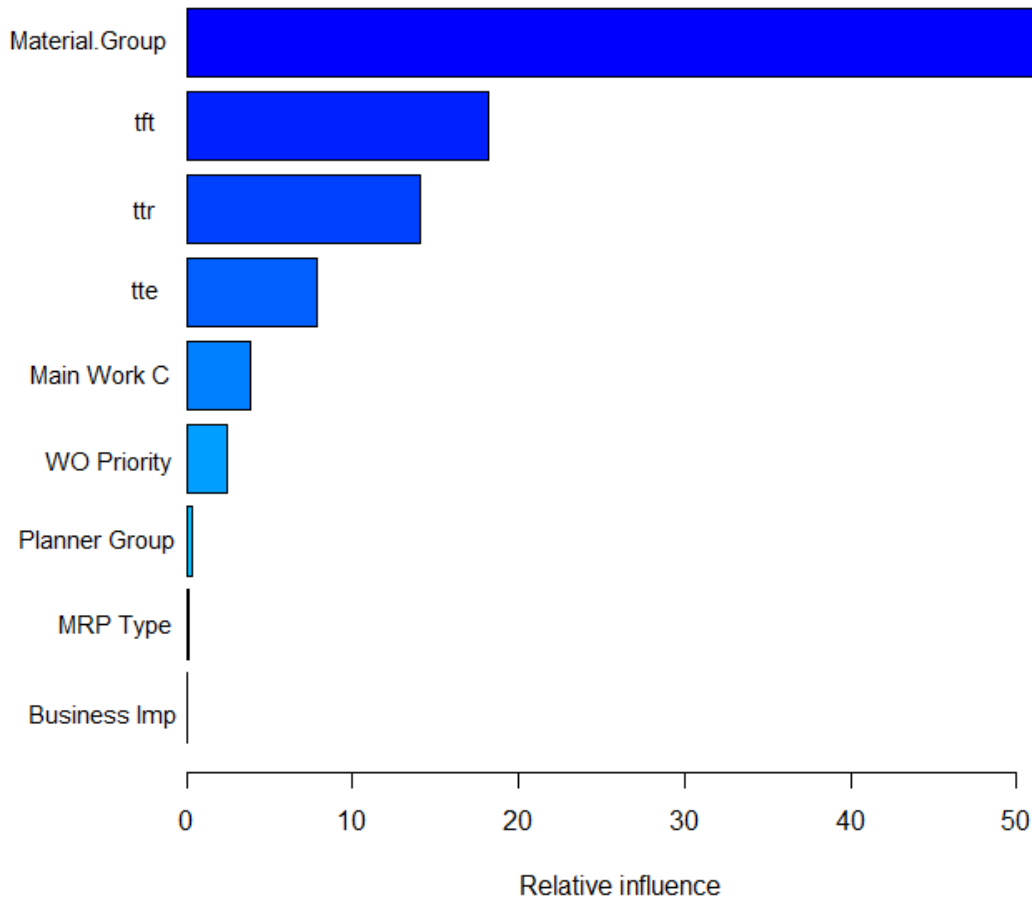


Figura 2.12: Gráfico Importancia relativa de las variables

	var	influencia relativa
factor(Material.Group)	factor(Material.Group)	53.00
	tft	18.20
	ttr	14.07
	tte	7.85
Main.Work.Centre	Main.Work.Centre	3.84
WO.Priority	WO.Priority	2.45
Planner.Group	Planner.Group	0.39
MRP.Type	MRP.Type	0.15
Business.Impact	Business.Impact	0.05

Tabla 2.11: Importancia relativa de las variables

La variable que explica la mayor parte de la variación a nivel de WOPIFOT es la familia de materiales. Esto tiene sentido pues es intuitivo pensar que materiales similares se com-

portan de forma similar para temas de planificación. La segunda variable es tft (Time from today), que refleja la antigüedad de la OT en el sistema. Esta variable tiene también una interpretación intuitiva del por qué de su relevancia. La antigüedad de una OT es un proxy de las replanificaciones de la misma, lo que implica que una OT antigua es muy probable que haya sido replanificada por falta de materiales u otra causa (falta de recursos HH, mala planificación, errores en la evaluación de las operaciones a realizar, etc), lo que implica que de por sí la OT ya ha tenido problemas en el pasado. En el gráfico de dependencia parcial se puede ver como cambia la probabilidad de "Hit." "Miss." en función de la variable tft, manteniendo el de las variables constantes en su valor promedio.

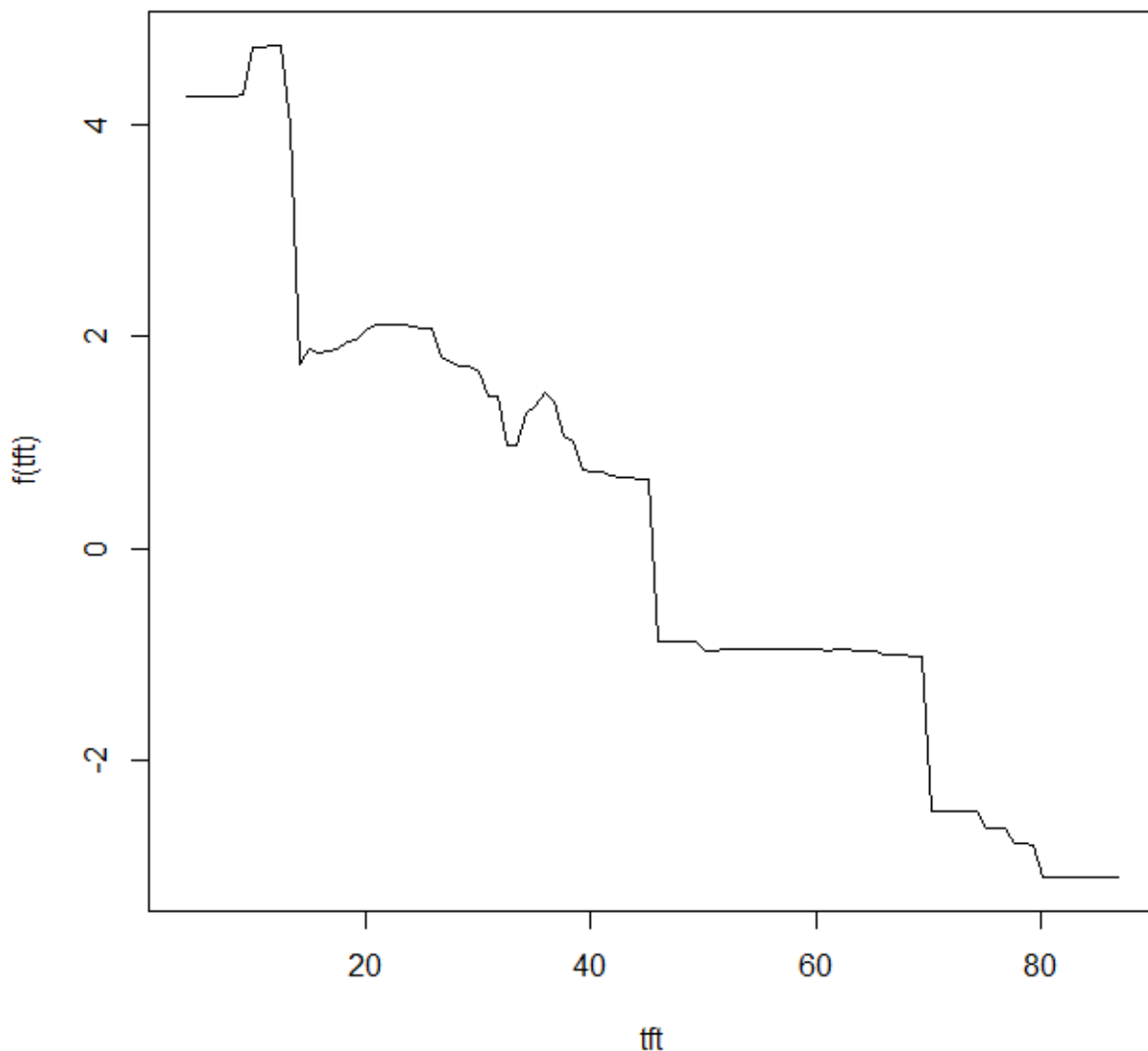


Figura 2.13: Dependencia parcial de Time from Today (tft)

De aquí se puede ver que existe una relación directa entre la antigüedad de la OT y la

probabilidad de "Miss", por las razones explicadas anteriormente.

En cuanto a la variable Time To Release, que explica %14.07 de la varianza, esta variable presenta un comportamiento inicialmente poco intuitivo, cuyo gráfico de dependencia parcial es de la siguiente forma:

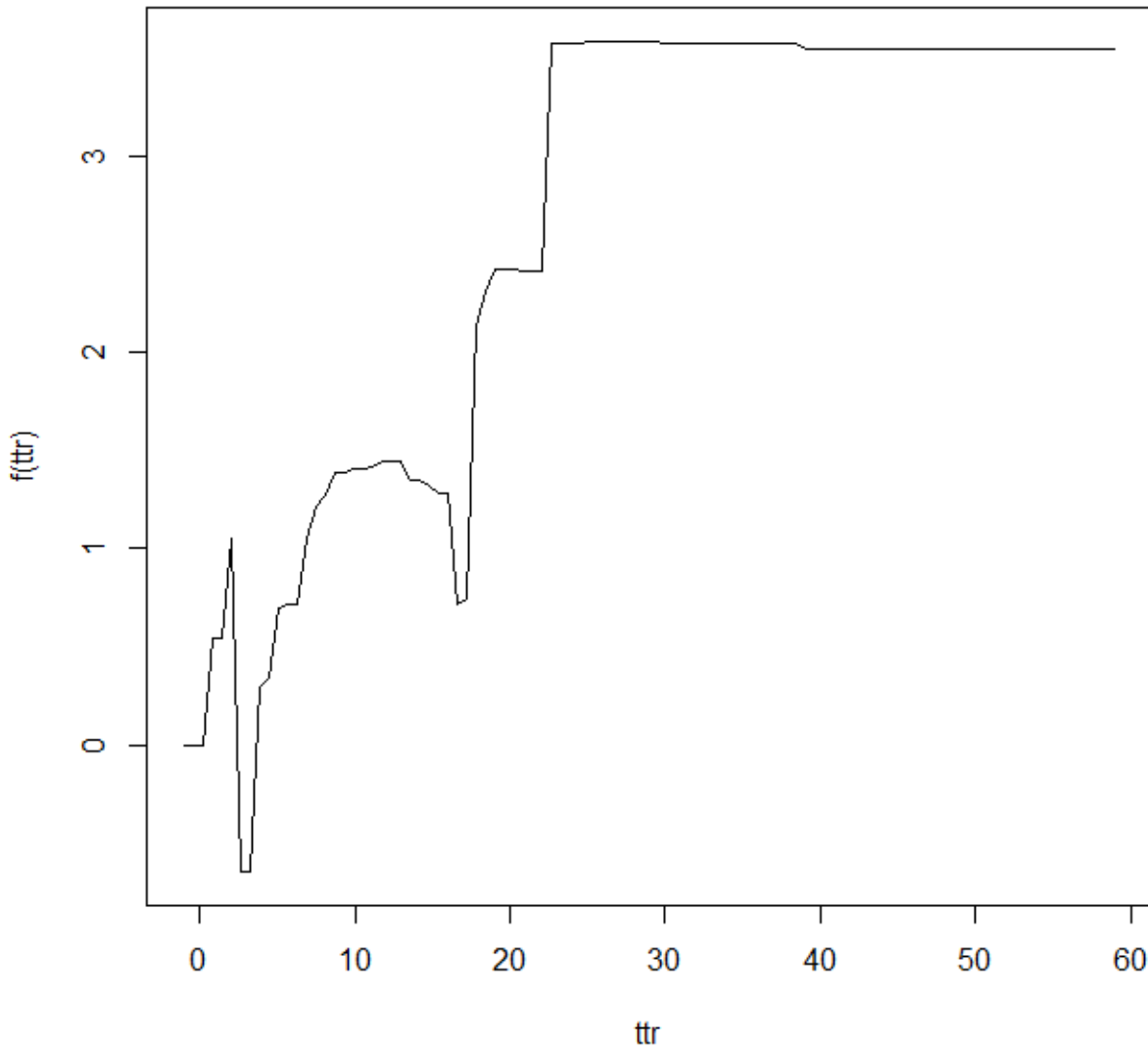


Figura 2.14: Dependencia parcial de Time to Release (ttr)

No obstante, la explicación puede ser que OT cuya liberación no es automática, es decir, requiere de revisión por parte de los planificadores, son menos probables de resultar en un "Miss" puesto que ya han sido revisadas antes de su liberación.

En relación a la variable Time to Execution (tte), que corresponde al horizonte de planifi-

cación por parte de Work Management y cuya importancia relativa en el modelo es de %7.85, se puede ver que el comportamiento tiene un peak alrededor de los 15-20 días de planificación, y las OT con un horizonte corto de planificación (<5 días) son muy propensas a fallas.

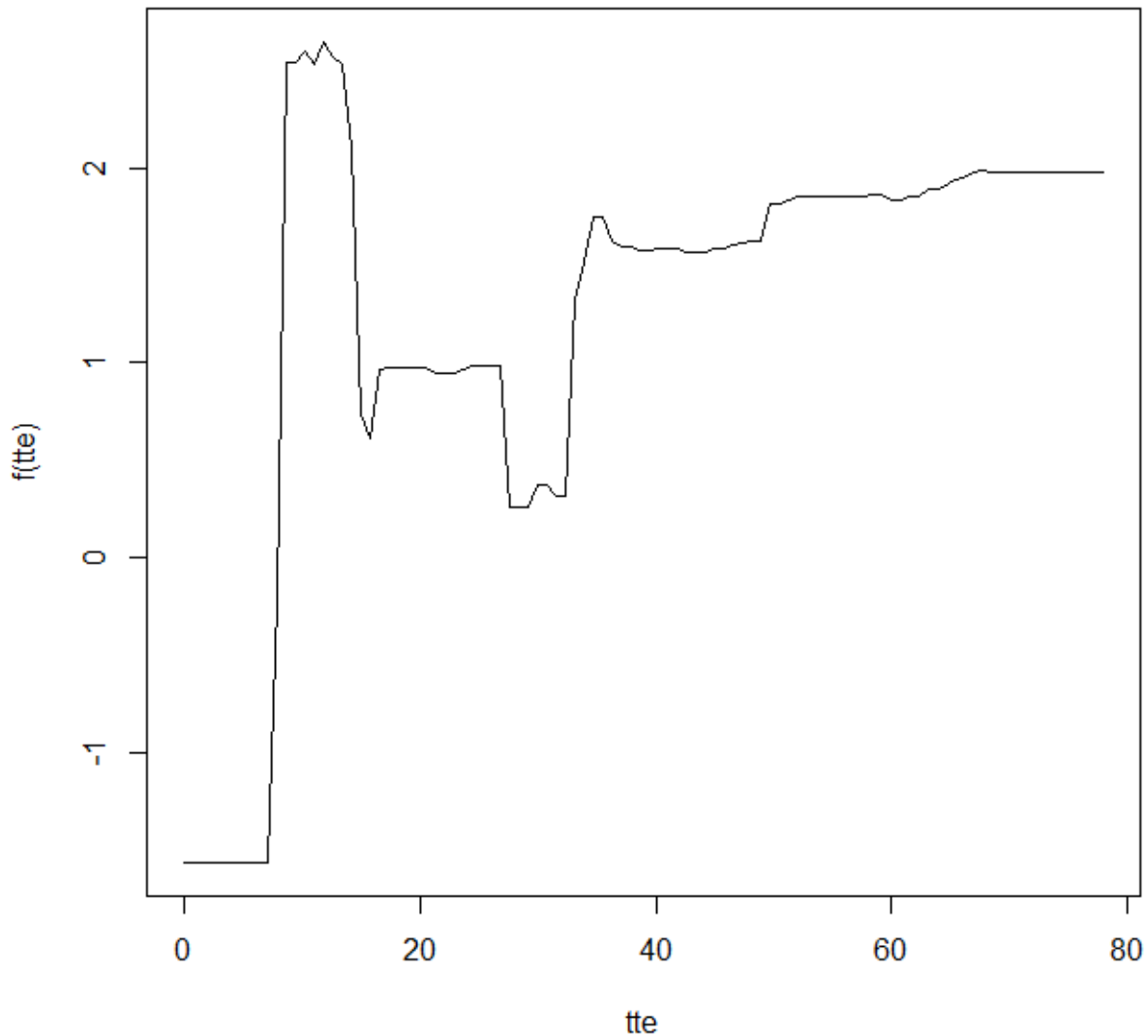


Figura 2.15: Dependencia parcial de Time to Execution (tte)

El testeo teórico del modelo se hizo calibrando un modelo en el mes de Mayo 2018, con toda la data histórica disponible y testeando con la data real del mes de Junio, siendo los resultados ya conocidos, es decir, el testeo se hace con data fuera de muestra para evitar problemas de sobreajuste del modelo.

Los resultados teóricos, incluyendo métricas de Accuracy, AUC, y Cohen's Kappa son los siguientes:

	Reference	
Prediction	0	1
0	86	9
1	44	157

Tabla 2.12: Matriz de Confusión, Testeo fuera de muestra

Donde 0 corresponde a un "Miss" 1 a un "Hit" del indicador WOPIFOT.

Accuracy :	0.8209
95 % CI :	(0.7724, 0.8629)
No Information Rate :	0.5608
P-Value [Acc >NIR] :	<2.2e-16

Tabla 2.13: Accuracy del modelo

Estadísticos	
Kappa :	0.6256
Mcnemar's Test P-Value :	3.008e-06
Sensitivity :	0.6615
Specificity :	0.9458
Pos Pred Value :	0.9053
Neg Pred Value :	0.7811
Prevalence :	0.4392
Detection Rate :	0.2905
Detection Prevalence :	0.3209
Balanced Accuracy :	0.8037

Tabla 2.14: Estadísticos del modelo

Estos resultados significan que, considerando que WOPIFOT PM01 real de Spence en el mes de Mayo fue de 56.06 %, considerando 166 "Hits" sobre un total de 296 Ordenes de trabajo completadas durante el mes, y que el modelo fue capaz de predecir correctamente 86 de las OT fallidas, y levantar falsa alarma sobre 44 OTs, el potencial teórico de utilizar la herramienta predictiva sería pasar de 56.06 % a 85.13 %, en el caso ideal de que fuese posible hacer gestión correctiva sobre todas las OT que el modelo clasifica como potenciales fallas. Esto a costa de la revisión innecesaria de 44 OTs por parte del planificador de inventario. Cabe destacar que durante este mes, fueron aplicadas algunas medidas de cierre de órdenes antiguas en el sistema, lo que explica el excepcional bajo desempeño de WOPIFOT durante el mes.

2.7.3. Etapa 2: Predicción de posible causa de falla

Para la segunda etapa del modelo predictivo, orientada a sugerir una posible causa de falla a las OT previamente clasificadas como riesgosas, se utiliza el mismo algoritmo de GBT utilizado anteriormente.

Para esto, el supuesto fundamental es que existen patrones en la configuración de las variables asociados a las causas raíces de las fallas a nivel de materiales (medidas por MIFOT).

Al entrenar un modelo de GBT, asumiendo una distribución multinomial dado que se trata de un problema de clasificación múltiple, y con los hiperparámetros obtenidos mediante un proceso de validación cruzada gracias al paquete de funciones de R `caret`, se obtienen los siguientes resultados:

Hiperparámetro	Valor
Shrinkage	0.05
Interaction depth	3
Min. Observaciones	10

Tabla 2.15: Hiperparámetros de GBT obtenidos mediante `Caret` para etapa 2.

En este modelo se consideran solo las causas más frecuentes detectadas en el análisis de causas raíces. Estas son:

- BAU-Incorrect Returns Process
- BAU-Possible Stocking Levels/Lead Time problem
- BAU-Possible vendor delivery
- BAU-Required outside lead time
- BAU-Required prior to WO release
- Free Text - Required Outside Lead Time
- PUR-Co-ordinate free text order/delivery
- URGENT-Inventory Review non stocked material miss
- URGENT-Inventory Review stocked material miss
- PUR-Free text not delivered as scheduled

No Information Rate	46.13 %	
Modelo	Accuracy	Kappa
GBT (ntree=100)	66.33 %	0.5214
GBT (ntree=1000)	67.58 %	0.5206
GBT (ntree=10000)	71.07 %	0.5762

Tabla 2.16: Resultados GBT en predicción de RCA.

En cuanto a la importancia relativa de las variables, obtenida gracias a las propiedades del algoritmo GBT, los resultados son los siguientes:

	var	rel.inf
Material.Group	Material.Group	81.30
tte	tte	7.83
MRP.Type	MRP.Type	4.77
ttr	ttr	3.93
WO.Priority	WO.Priority	1.10
Main.Work.Centre	Main.Work.Centre	0.88
Planner.Group	Planner.Group	0.20
Work.Order.Type	Work.Order.Type	0.00

Tabla 2.17: Importancia relativa de variables en predicción de RCA

De donde se obtiene que el 81.3% de la variabilidad es explicada por la familia del material, lo que sustenta el supuesto de que las fallas a nivel de material están fuertemente correlacionadas con el factor humano y el trato que se le da a cada tipo de material en el proceso de planificación.

La variable MRP Type toma más relevancia en este modelo ya que el tipo de material está directamente correlacionado con algunos tipos de falla. Por ejemplo, materiales no catalogados, sin MRP Type están asociados con fallas de free-text, materiales tipo PD están asociados con fallas de leadtime, ya que son materiales comprados a pedido.

2.7.4. Implementación del modelo predictivo

Para llevar el modelo de una construcción teórica a una herramienta aplicable a la gestión de Supply, es necesario utilizar los resultados anteriormente presentados y el modelo desarrollado en una interfaz dinámica.

La utilización de la herramienta predictiva se basa en el entrenamiento del modelo en base a data histórica de las órdenes de trabajo (WOPIFOT, MIFOT, RCA Tool) con una cadencia por definir, de forma que el modelo permita capturar nuevos patrones de comportamiento a medida que las acciones correctivas se apliquen y los procesos se adapten a la gestión.

Una vez entrenado el modelo (ajustado a la data histórica), se puede proceder a analizar las nuevas ordenes de trabajos encontradas en el sistema, y así clasificarlas en un potencial miss o hit a nivel de material, y en una segunda etapa del modelo, se asigna una causa probable de falla de acuerdo a los patrones detectados de RCA Tool.

Finalmente una vez construidas las columnas de predicción de hit/miss y asignada una posible causa a las potenciales fallas, se utiliza la data de inventario para descartar falsas alarmas eliminando las alertas en ordenes de trabajos predichas como potencial falla con causa de quiebre de stock. Una vez hecho esto, la data es llevada a un dashboard de forma de que el usuario pueda interactuar con la plataforma y con esto apoyar la gestión de las órdenes de trabajo.

La figura siguiente ilustra el funcionamiento del modelo y su interacción con el usuario.

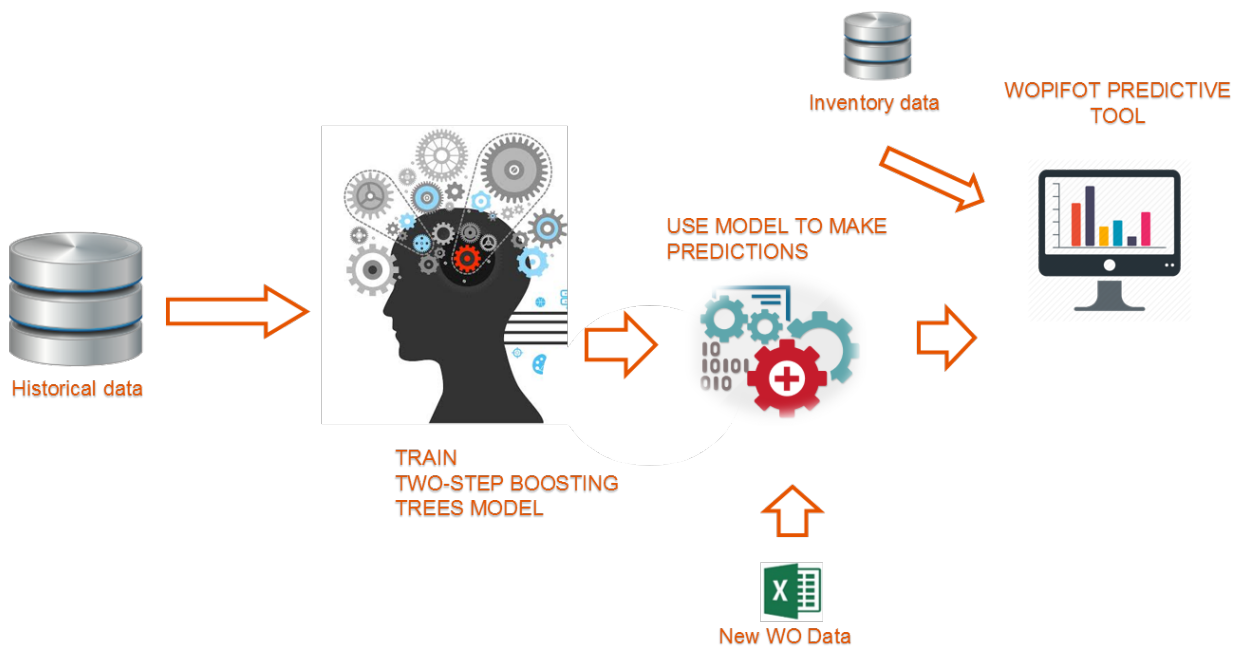


Figura 2.16: Herramienta predictiva e interacción con el usuario

2.8. Diseño del dashboard para la herramienta WPT

Una etapa necesaria para pasar del modelo teórico a una herramienta amigable para el usuario, en este caso planificadores de inventario, es diseñar una interfaz que permita revisar los resultados del modelo predictivo en las rutinas de planificación de forma dinámica.

Así, se hace necesaria la construcción de un dashboard mediante el software TIBCO Spotfire que permite contar con diversas funcionalidades de visualización y filtrado de datos.

Se utiliza TIBCO Spotfire dado que es el software oficial en BHP para visualización de data y reportabilidad y en vista de esto, es también amigable para el cliente final, en este caso, los planificadores de inventario.

El dashboard consta de 5 áreas donde cada una de estas constituye una etapa en la revisión de las órdenes de trabajo futuras durante las rutinas de planificación.

Un primer recuadro permite al usuario realizar los filtros por área de trabajo (Main Work Centre) de forma de poder derivar las acciones necesarias de forma más precisa. Ejemplos de áreas son Palas, Equipos Cat, Área seca, Área húmeda, etc.

Luego del primer filtro por Main Work Centre, el dashboard mostrará las órdenes de trabajo abiertas para esa área en particular, mostrando en un gráfico de barras la dispersión de OTs por semana del año en el eje horizontal, y la cantidad de OTs programadas para esa semana en el eje vertical. Esto permite al planificador tener una primera intuición de la distribución de la carga de trabajo para las semanas futuras.

En este mismo gráfico, se muestra en las barras, divididas por color, las ordenes de trabajo que el modelo predice como riesgosas (en rojo), y las que el modelo no detecta un patrón histórico de riesgo (en color verde). En este recuadro, el planificador puede seleccionar la área roja del gráfico de barras para poder acceder al detalle de las órdenes de trabajo riesgosas en esa semana en particular.

La siguiente etapa corresponde a un gráfico de dispersión de las ordenes de trabajo por día de la semana seleccionada anteriormente. Cada orden de trabajo es representada como un círculo y su tamaño muestra el precio total de la OT (calculada como la suma del costo de todos los materiales requeridos para completar las operaciones descritas en la OT). El color del círculo representa el impacto de cada OT en el negocio, siendo las de color rojo las con impacto mayor (parada de equipos principales), las de color anaranjado con impacto menor (detención de equipos secundarios) y de color verde sin impacto en las operaciones.

Una cuarta etapa corresponde al detalle de cada orden de trabajo, donde el planificador puede revisar los materiales que el modelo predice como riesgosos y la causa asociada a la posible falla.

Finalmente, un último recuadro muestra la distribución de la carga de trabajo asociada a la gestión de las OTs riesgosas, dividiendo las acciones necesarias entre las áreas de Compras, Inventario y Planificación.

2.8.1. Diseño de experimento

Para finalmente probar y medir el valor que agrega WPT al negocio, es necesario un test real del impacto que genera su utilización. Para esto, se propone un experimento que permita comparar en un horizonte de un mes el cambio a nivel de WOPIFOT de un grupo de prueba frente a un grupo de control.

2.8.1.1. Experimento

Dado que dentro de las áreas de Mina y Planta de minera Spence, WOPIFOT presenta un comportamiento similar, se propone utilizar una de estas áreas como grupo de control, y la otra como sujeto de experimento. Así, se utilizará la herramienta WPT por los planificadores de una de estas áreas durante un periodo de un mes, mientras la otra área continuará con sus operaciones regulares.

Para ambas áreas, cualquier directiva asociada a la gestión de materiales será adoptada, lo que no debería alterar los resultados del experimento puesto que la única variable que diferirá entre áreas es la utilización de la herramienta predictiva.

Así, luego del periodo de prueba, se obtienen los resultados y mediante un modelo de diferencias en diferencias es posible determinar si existe un cambio en la tendencia que sea estadísticamente significativo, lo que probaría la validez o no validez de la herramienta.

2.8.1.2. Resultados Esperados

Se espera que luego del periodo de un mes, se evidencie un cambio en la tendencia de WOPIFOT en el área experimental. Sin embargo, por efectos de la reticencia al cambio y el factor humano, y factores aleatorios, este cambio en la tendencia no debería alcanzar el máximo potencial de lo indicado en los resultados teóricos.

La gestión del cambio es una parte fundamental del proceso de incorporación de nuevas herramientas a las rutinas en todo ámbito del negocio, y es este el mecanismo que permitirá aprovechar el máximo potencial de cualquier nueva tecnología.

2.9. Visión estratégica

Uno de los objetivos secundarios del proyecto WOPIFOT Predictive Tool es crear un precedente en BHP MinAm de como las herramientas de analytics pueden agregar valor al negocio mediante la automatización de procesos de análisis y la utilización más eficaz de los datos.

Tanto en el área de Supply de BHP Minerals Americas, como en las demás áreas de la empresa, herramientas analíticas del tipo descriptivas, de diagnóstico y de monitoreo son frecuentemente utilizadas para detectar, entender con base en datos y finalmente abordar los problemas que surgen en las operaciones.

Dicho esto, existe la oportunidad dentro de Supply y las demás áreas de Minerals América de adoptar nuevas tecnologías de analytics y data science más complejas y con esto agregar valor al negocio, ya sea mediante la optimización de procesos o la reducción de costos y horas hombre convirtiendo a los analistas en partners del negocio, y dedicando más tiempo a la gestión que al análisis mismo.

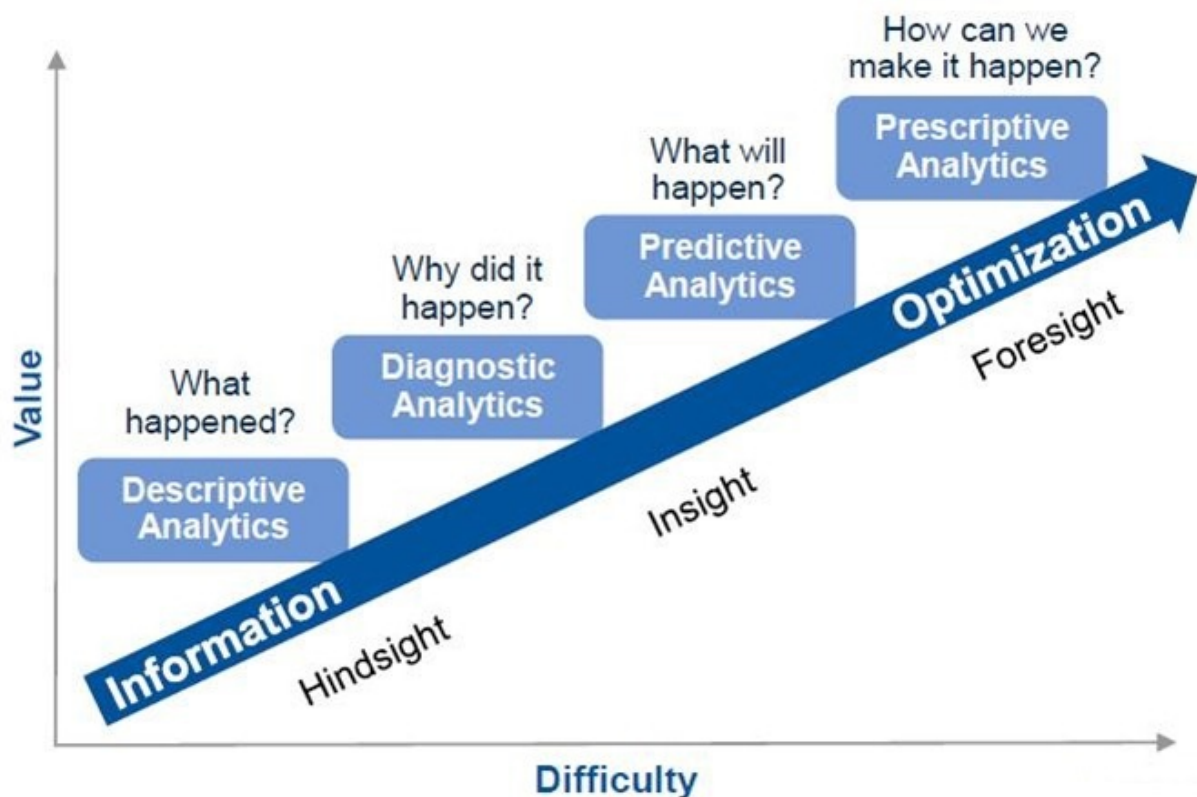


Figura 2.18: Curva de evolución de Analytics. Complejidad vs Valor agregado

La tabla anterior muestra la curva natural de desarrollo del data science, relacionando la complejidad del modelamiento y algoritmos con el potencial valor agregado al negocio. Así, la visión estratégica de la implementación de la herramienta WPT, es mostrar dentro de, en principio, el área comercial de MinAm el potencial de invertir en el desarrollo del data

science.[14]

Junto con esto, parte de la visión es automatizar los análisis de datos y dar la oportunidad a analistas de optimizar el uso de su tiempo e invertirlo en gestión, toma de decisiones e interpretación de resultados, más que en generación de reportes y trabajo manual de bases de datos.

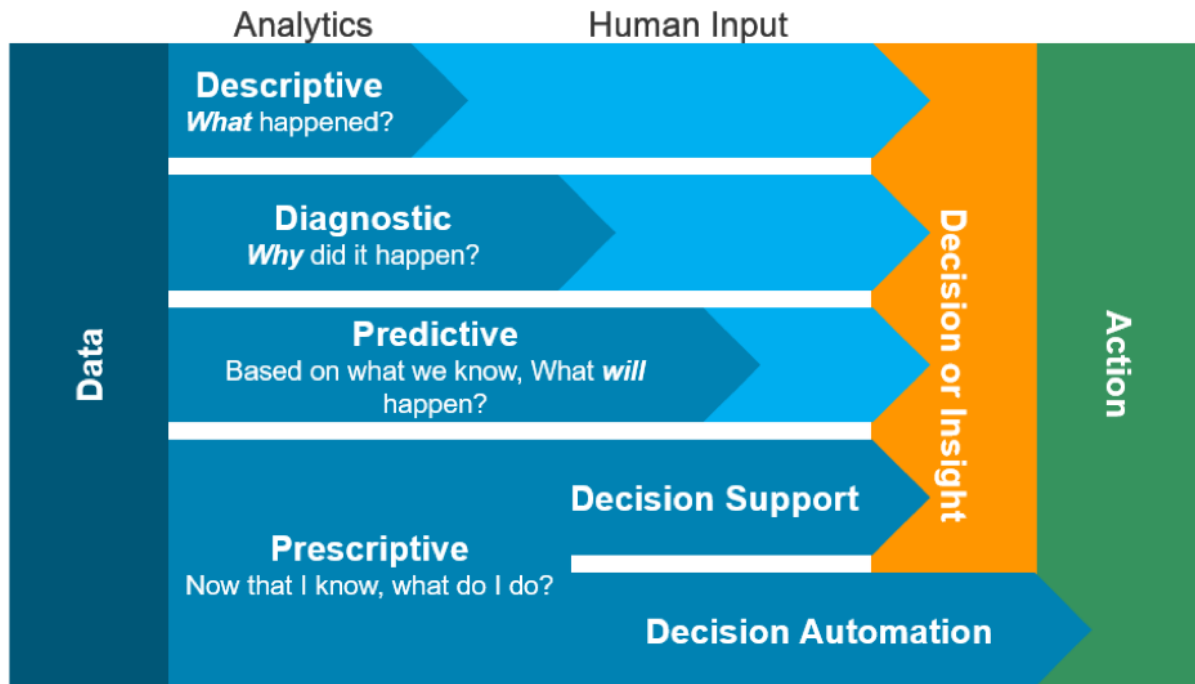


Figura 2.19: Relación entre complejidad de modelos e intervención humana

Como se muestra en la tabla anterior, el propósito de fomentar el desarrollo de nuevas habilidades en el área de ciencia de datos, permite reducir el input humano requerido para el análisis de los datos, y utilizarlo de mejor forma en decisión y finalmente en gestión.

Conclusión

2.10. Cumplimiento de los objetivos propuestos

En relación a los objetivos específicos propuestos al comienzo de este documento, una vez terminado el desarrollo del proyecto el estado de cada uno es el siguiente:

Fue levantado el proceso de supply mediante una metodología BPMN y gracias al conocimiento experto de diferentes profesionales del área de supply. Poder mapear este proceso resulta ser fundamental para el entendimiento del negocio y el posterior entendimiento de los datos.

En cuanto a la gestión de las bases de datos, el resultado del proyecto es un Data Warehouse compuesto de data relativa a las órdenes de trabajos y que alimenta el modelo predictivo.

El algoritmo que mejor resultados entrega es el de Gradient Boosting Trees, tanto en accuracy como performance general en cuanto a tiempo, interpretabilidad de los resultados y recursos computacionales.

El modelo anteriormente mencionado resulta con un precisión de 82,2% lo que permite generar una herramienta predictiva confiable y además automatizar el análisis de las órdenes de trabajo, lo que permite a los planificadores enfocar su trabajo sólo en OTs riesgosas y así optimizar su tiempo.

2.11. Resultados Observados

Considerando que el proceso de supply para mantenimiento comienza con la emisión de una orden de trabajo, es de fundamental importancia la capacidad de ser proactivo en la gestión de estas. El proyecto WPT (WOPIFOT Predictive Tool) permite lograr este objetivo al área de Supply.

Utilizar una metodología de trabajo cuyo núcleo sea el entendimiento del problema a nivel del negocio y el posterior entendimiento de los datos hace que el proyecto tenga sustento y pueda finalmente agregar valor.

Parte fundamental de un proyecto de data mining es el acceso a los datos, y la limpieza y formato de estos. El éxito o fracaso de proyectos de este tipo depende en gran parte de la disponibilidad, calidad y capacidad de trabajar los set de datos de buena forma.

Es importante considerar quien es el cliente final del proyecto de minería de datos y diseñar la solución junto con este, de forma iterativa y considerando de que forma será utilizada la herramienta de forma de crear una interfaz amigable con el usuario.

Finalmente, y muy de la mano con lo anterior, la gestión del cambio es un proceso fundamental en la incorporación de nuevas tecnologías. Esto considera también la creación de nuevas capacidades en la organización de forma de poder soportar el cambio tecnológico.

2.12. Trabajo Futuro

Quedan propuestos como trabajo a futuro los siguientes puntos, que si bien están fuera del alcance de esta memoria, permitirían robustecer el modelo predictivo y agregar valor al negocio mediante una herramienta más confiable.

- Realizar el testeo on-site formulado en el punto (3.7.1) de este documento
- Incorporar en el modelo variables relativas a los vendor (proveedores) y al manejo de materiales en bodegas, esto permitiría contar con un modelo más robusto que de también mejores y más completos insights sobre el proceso de supply.
- Complementar el modelo predictivo con un análisis de stock disponible vs reservas, lo que permitiría disminuir la cantidad de falsos positivos generados por el modelo.
- Realizar el correcto proceso de gestión del cambio, generando las nuevas capacidades necesarias para utilizar y continuar desarrollando modelos de data science más complejo que permitan automatizar análisis, y agregar así valor al negocio
- Hacer el análisis económico de la variación del indicador WOPIFOT en relación a la disponibilidad de equipos y su consecuente impacto en la producción

Bibliografía

- [1] Shmueli, G., Patel, N. R., Bruce, P. C., *Data Mining for Business Intelligence. 2nd ed.* John Wiley and Sons, Hoboken, New Jersey, 2010.
- [2] Rüdiger Wirth., *CRISP-DM: Towards a standard process model for data mining.* Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 2000.
- [3] J. Ross Quinlan. *C4.5: programs for machine learning.* Morgan Kaufmann Publishers Inc. San Francisco, CA, USA 1993.
- [4] Andrew Colin., *Building Decision Trees with the ID3 Algorithm.* Dr. Dobbs Journal, 1996
- [5] Paul E. Utgoff, *Incremental Induction of Decision Trees.* Kluwer Academic Publishers, 1989
- [6] D. Hosmer, S. Lemeshow. 2000. *Applied Logistic Regression.* Second Edition, Wiley Series in Probability and Statistics.
- [7] S. Yanmin, M. Kamel, Y. Wang., *Boosting for Learning Multiple Classes with Imbalanced Class Distribution.* Proceedings of the Sixth International Conference on Data Mining, 2006
- [8] Breiman L. Bagging predictors *Machine Learning forthcoming*, 1996.
- [9] Leo Breiman. 2001. *Random Forests.* Mach. Learn. 45, 1 (October 2001)
- [10] Jerome H. Friedman, *Greedy Function Approximation : A Gradient Boosting Machine.* Stanford, 2010.
- [11] Greg Ridgeway, *Generalized Boosted Models : A guide to the gbm package*, 2005.
- [12] Cheng G. Weng Josiah Poon, *A New Evaluation Measure for Imbalanced Datasets*, School of Information Technologies, J12, University of Sydney, Sydney, NSW, Australia 2006,
- [13] Tom Fawcett, *Learning from Imbalanced Classes, 2016.* <https://www.svds.com/learning-imbalanced-classes/>

- [14] Morgan & Duncan, *Analytics maturity curve*, 2016.
- [15] Kimball et al. John Wiley and Sons *Building the data warehouse (2da edición)*., New York, 1996.
- [16] Kimball, R. y Merx, *The data webhouse toolkit. Wiley Computer Publisher*, New York, USA. 2000.
- [17] Kimball et al. John Wiley and Sons *The data warehouse lifecycle toolkit: expert methods for designing, developing and deploying data warehouses.* , New York, USA. 1998.
- [18] [29] Levene, M. y Loizou, G. *Why is the snowflake schema a good data warehouse design? Information Systems*, 28. pp. 225-240. 2003.
- [19] Vassiliadis, P., Simitsis, A. y Skiadopoulos, S. *On the logical modeling of ETL processes. In Proc. 14th Int. Conf. on Advanced Information Systems Engineering, London, UK.* pp. 782-786.2002.

Anexos

2.13. Causas raices de fallas del indicador MIFOT separadas por rol responsable

Responsable	Causa Raiz
MCoE strategy / Site A&I	Parts added after 'schedule lockdown' - BAU work only Parts (spare) not on a task list Free text with no catalogue match required for urgent PM03 work Parts (spare) added on requirement date for PM02 work Parts (spare) added during execution (except PM03) & work order has task list
Maintenance Execution	Free text (with catalogued V1 match) required for urgent work Part (consumable) added on requirement date Part not collected for urgent work Parts (spare) added on requirement date (except PM03) & work order has no task list
MCoE Planning	Part required on or prior to WO release Free text outside goods handling time for BAU work Catalogued non-stock outside lead time for BAU work
Site Planning	BAU Reservation moved forward Free text added after handover for BAU work Catalogued non-stock part added after lockdown for BAU work
Warehousing	Warehouse Picking issue - urgent work Warehouse Picking issue - BAU work Incorrect Returns Overdue return not processed
Vendor	Possible Vendor DIFOT issue Part not delivered as per lead time

Tabla 2.18: Causas raices de fallas en MIFOT por rol responsable 1

Responsable	Causa Raiz
Vendor	Possible Vendor DIFOT issue Part not delivered as per lead time
Purchasing	Vendor number not mentioned for free text purchase PR to PO conversion issue for DP's PR to PO conversion issue for non-DP's Free text not delivered as scheduled Co-ordinate free text delivery/order
Inventory Analyst	Catalogued non-stocked part required for urgent work Stocked (V1) item not sufficient to meet the demand Stocked item directly purchased Obsolete item on work order Update lead time on material Stocking level at hub not sufficient to meet demand Part (consumable) added after scheduled lockdown Free text (with catalogued non-stock match) required for urgent work
Inventory Planning	Internal STR to STO took more than 3 days Logistics issue from hub to forward store ZD material not delivered on time

Tabla 2.19: Causas raices de fallas en MIFOT por rol responsable 2

2.14. Código WOPIFOT Predictive Tool en R

2.14.1. Extracción, Transformación y Carga de data

```

dataW =
  read.csv('//petstgdc2.americas.ent.bhpbilliton.net/Departmental_Shares/Asset/SpotFire/Supp
  Pablo/12. WOPIFOT Predictive Tool/WOPIFOT DATA.csv', sep=';')
dataM =
  read_xlsx('//petstgdc2.americas.ent.bhpbilliton.net/Departmental_Shares/Asset/SpotFire/Supp
  Pablo/Supply Synch/MIFOT.xlsx' )
# dataM = read_xlsx('MIFOT.xlsx' )

names(dataM)<-make.names(names(dataM))
colnames(dataM)[7]<-"Work.Order.Number"
Oniqua <- read.csv('HISTORICAL MRP TYPE.csv', sep=";")
mwc.area<-read.table(textConnection("Main.Work.Centre Area

```


B201/MN01 Mine
B201/MN02 Mine
B201/MN03 Mine
B201/MN04 Mine
B201/MN05 Mine
B201/MN06 Plant
B201/MN07 Plant
B201/MN08 Plant
B201/MN09 Plant
B201/MN10 Plant
B201/PN01 Plant
B201/PN02 Mine
B201/SN20 Mine
B501/MN01 Perforadoras
B501/MN02 Perforadoras
B501/MN03 Palas
B501/MN04 Palas
B501/MN05 Camiones
B501/MN06 Camiones
B501/MN07 Camiones
B501/MN09 Camiones
B501/MN11 PTT.EEAA
B501/MN12 EEAA
B501/MN13 EEAA
B501/MN15 Chancadores
B501/MN16 Chancadores
B501/MN17 Chancadores
B501/MN18 Camiones
B501/MN20 Concentradoras
B501/MN21 Concentradoras
B501/MN22 Concentradoras
B501/MN23 Concentradoras
B501/MN24 Concentradoras
B501/MN30 Concentradoras
B501/MN31 Concentradoras
B501/MN32 Concentradoras
B501/MN33 Concentradoras
B501/MN34 Concentradoras
B501/MN35 Ctodos
B501/MN36 Ctodos
B501/MN37 Ctodos
B501/MN38 Ctodos
B501/MN39 Ctodos
B501/MN40 Ctodos
B501/MN41 CHO
B501/MN45 Ctodos
B501/MN46 CHO
B501/MN50 Ctodos
B501/MN51 NPI
B501/MN52 NPI

B501/MN60 NPI
B501/MN65 NPI
B501/MN72 NPI
B501/MN73 NPI
B501/MN80 Concentradoras
B501/MN81 Palas
B501/MN82 Concentradoras
B501/MN83 Concentradoras
B501/MN84 Others
B501/MN85 Concentradoras
B501/MN87 Concentradoras
B501/MN88 Palas
B501/PN01 NPI
B501/PN02 OperMina
B501/PN03 Chancadores
B501/PN05 Concentradoras
B501/PN06 Ctodos
B501/PN06 CHO
B501/PN34 Others
B501/PN35 Ctodos
B501/PN37 Ctodos
B501/PN38 Others
B501/SN20 Concentradoras
B501/SN26 Others
B501/SN34 Concentradoras
B501/SN35 Concentradoras
B521/MN42 CHO
B521/MN43 CHO
B521/MN44 CHO
B521/MN45 CHO
B521/MN55 NPI
B521/MN66 NPI
B521/PN07 NPI
B701/MN01 Mine
B701/MN01 Mine
B701/MN02 Mine
B701/MN03 Mine
B701/MN04 Mine
B701/MN21 Plant
B701/MN31 Plant
B701/MN35 Plant
B701/MN36 Mine
B701/MN41 Plant
B701/MN51 Plant
B701/MN53 Plant
B701/MN61 SS00
B701/PN01 Mine
B701/PN32 Plant
B701/PN32 Plant
B701/PN33 Plant

```

B701/SN17 SS00
CY04/MN90 Others
CY08/MN13 Others
FX5C/MN61 Others
B501/MN86 OGP1
B501/MN90 Mechanics.Shutdown.LC
B501/MN91 Mechanics.Shutdown.LS.L1
B501/MN92 Mechanics.Shutdown.LS.L2
B501/MN93 Electrical.Shutdown.LS

"), header=TRUE)

```

```

#Filter SPENCE only (Spence=B701, CMCC=B201)
Oniqua <- Oniqua[Oniqua$Plant=="B701",]

#Rename some columns
colnames(dataW)[27]<-"WOPIFOT"
colnames(dataM)[15]<-"MIFOT"

#Variables to keep
#From Wopifot
WopVar<-c("Region",
          "Operation",
          "Work.Order.Number",
          "Work.order.created.b",
          "Work.Order.Type",
          "Planner.Group",
          "Main.Work.Centre",
          "Work.order.release.date",
          "X.2",
          "Work.Order.Created.D",
          "Basic.Start.Date",
          "Work.Order.TECO.date",
          "WOPIFOT"
)
#From Mifot
MifVar<-c("Work.Order.Number",
          "Material",
          "Material.Name",
          "Calculated.Reservation.Requirement.Date",
          "MIFOT")

dataW<-dataW[WopVar]#Keep only variables listed above
dataM<-dataM[MifVar]

#Merge WOPIFOT and MIFOT reports

dataWO=merge(dataM,dataW,"Work.Order.Number")

```

```

colnames(dataWO)[2]<-"Material.Number"
colnames(dataWO)[13]<-"WO.Priority"

#Remove data from older than 2017 WOs (As they behave different from 2017 ones. No
  predictive value)
dataWO$Year<-format(as.Date(dataWO$Work.Order.Created.D, format="%d.%m.%Y"), "%Y")
  #Create YEAR column
dataWO$Month<-format(as.Date(dataWO$Work.Order.Created.D,
  format="%d.%m.%Y"), "%m") #Create MONTH column
dataWO<-dataWO[as.numeric(dataWO$Year)>=2017,]

#Correct Busines Impact nomenclature
Bimp<-function(s){
  myLetters <- LETTERS[1:26]
  if (nchar(s)>1){return(s)}
  else{ n<-match(s, myLetters)
  BI<- switch(n,
    "A-Major Production Stopper",
    "B-Minor Production Stopper",
    "C-Major Maintenance Impact",
    "D-Minor Maintenance Impact",
    "E-No Impact"
  )
  return(BI)
}
}
Oniqua$Business.Impact<-
  as.factor(sapply(as.character(Oniqua$Business.Impact),Bimp))

dataWO$Month<-as.numeric(dataWO$Month)
dataWO$Month<-sapply(month.abb[dataWO$Month],toupper)

#Merge dataWO and Oniqua to assignt MRP and BImp to each material according to date
dataWO<-dataWO[!(dataWO$Year=="2017" & dataWO$Month %in% c("JAN", "FEB", "MAR",
  "APR", "MAY")),] # Remove data prior to JUN
dataWO$KEY<-paste(dataWO$Material.Number,dataWO$Month,dataWO$Year,sep="")
Oniqua$KEY<-paste(Oniqua$Material.Number,Oniqua$Month,Oniqua$Year,sep="")
data<-merge(dataWO,Oniqua,"KEY",all=TRUE)

data <- data[!is.na(data$Operation),] #Delete matched data w/o WONumber
data$Plant<-NULL
data$Year.y<-NULL
data$Month.y<-NULL
data$Material.Number.y<-NULL
data$KEY<-NULL

```

```

data<-join(data,mwc.area)

#Replace MRPTYPE and Business.Impact NAs as "Not Assigned"
levels(data$Business.Impact)<-c(levels(data$Business.Impact),"Not assigned")
data$Business.Impact[is.na(data$Business.Impact)]<-"Not assigned"
levels(data$MRP.Type)<-c(levels(data$MRP.Type),"Not assigned")
data$MRP.Type[is.na(data$MRP.Type)]<-"Not assigned"

#Group all system Work Orders as User=SISTEMA
group<-function(x){
  if (substr(as.character(x),1,3)=="IP1"){fact= "SISTEMA"}
  else {fact=as.character(x)}
  return(fact)
}

data$User.Type=data$Work.order.created.b
data$User.Type=as.factor(unlist(lapply(data$User.Type,group)))

#Replace wopifot factors as hit/miss
replacehm<-function(x){
  if (x==0){fact= "Miss"}
  else {fact="Hit"}
  return(fact)
}
data$WOPIFOT_N<-as.numeric(gsub("^([0-9]+),([0-9]+)$","\\1.\\2", data$WOPIFOT))
data$WOPIFOT<-unlist(lapply(data$WOPIFOT_N,replacehm))
data$WOPIFOT=as.factor(data$WOPIFOT)

#Material description to string
data$Material.Name <- unlist(data.frame(lapply(data$Material.Name, as.character),
  stringsAsFactors=FALSE))

#Material description aggregate groups function
matgroup<-function(m){
  if (is.na(m)) return("B701/Not assigned")
  if (m=="Not assigned") return("Free Text")
  if (m=="B701/Not assigned" | m=="B201/Not assigned"){return("Free Text")}
  else {return(gsub(",.*$", "", m))}
}
data$Material.Group<-unlist(lapply(data$Material.Name,matgroup))

#Identify Duplicate materials
dupli<-function(x){
  if (substr(as.character(x),1,3)=="{DU}") {fact= "DUPLICATED"}
  else {fact=as.character(x)}
  return(fact)
}

```

```

}

data$Material.Group<-sapply(data$Material.Group,dupli)

data$UNO<-1
#Define list of common materials, label the rest as OTHERS
rm<-500 #Cantidad de materiales relevantes
<-----Adjust
Common.Materials<-names(sort(table(data$Material.Group),decreasing=T))[1:rm]

commonmat<-function(m){
  ifelse(m %in% Common.Materials,return(m),return("OTHERS"))
}

data$Material.Group<-unlist(lapply(data$Material.Group,commonmat))
data$Material.Group<-factor(data$Material.Group)

#Create Time variables
data$today<-Sys.Date()
data$Basic.Start.Date=as.Date(data$Basic.Start.Date, format = "%d.%m.%Y")
data$Work.Order.Created.D=as.Date(data$Work.Order.Created.D, format = "%d.%m.%Y")
data$Work.order.release.date=as.Date(data$Work.order.release.date, format =
"%d.%m.%Y")
data$Work.Order.TECO.date=as.Date(data$Work.Order.TECO.date, format="%d.%m.%Y")

data$ttr=as.integer(data$Work.order.release.date -data$Work.Order.Created.D) #Time
to release
data$tte=as.integer(data$Basic.Start.Date-data$Work.Order.Created.D) #Time to
execute
data$tft=as.integer(data$Work.Order.TECO.date-data$Work.Order.Created.D) #Time
from today /TECO

data$MIFOT_MH<-data$MIFOT
data$MIFOT_MH[data$MIFOT==100]<-"Hit"
data$MIFOT_MH[data$MIFOT!=100]<-"Miss"
data$MIFOT_MH<-factor(data$MIFOT_MH)

data<-data[data$Area %in% c("Mine","Plant"),]
str(data)

#Compute last month wopifot and mifot
LM.Wopifot.Mat<-ddply(data[data$Month.x=="NOV"|data$Month.x=="DEC",],
.(Material.Group),summarize,LM.Wopifot=mean(WOPIFOT_N))
LM.Mifot.Mat<-ddply(data[data$Month.x=="NOV"|data$Month.x=="DEC",],
.(Material.Group),summarize,LM.Mifot=mean(MIFOT))
LM.Dif<-ddply(data[data$Month.x=="NOV"|data$Month.x=="DEC",],
.(Material.Group),summarize,LM.Mifot=mean(MIFOT),LM.Wopifot=mean(WOPIFOT_N))

```

```

LM.Dif$M.Dif<-LM.Dif$LM.Mifot-LM.Dif$LM.Wopifot
LM.Dif[2:3]<-list(NULL)
data2<-data
data2<-merge(data2,LM.Mifot.Mat,by="Material.Group", all=TRUE)
data2<-merge(data2,LM.Wopifot.Mat,by="Material.Group", all=TRUE)
data2<-merge(data2,LM.Dif,by="Material.Group", all=TRUE)

table(ceiling(data2$MIFOT/100))
table(data2$WOPIFOT_N/100)
table(ceiling(data2$MIFOT/100),data2$WOPIFOT_N/100)
data2$WM<-ceiling(data2$MIFOT/100)*data2$WOPIFOT_N/100

data<-data2

data$TECO.Month<-format(as.Date(data$Work.Order.TECO.date,
  format="%d.%m.%Y"),"%m") #Create MONTH column
data$WOPIFOT_N<-data$WOPIFOT_N/100

```

2.14.2. Estadística descriptiva

```

#-----ESTADSTICA
  DESCRIPTIVA-----
xtable.confusionMatrix <- function(x, caption = NULL, label = NULL, align = NULL,
  digits = NULL, display = NULL, ...){
  tech <- attr(x, "tech")
  nFeaturesOptim <- attr(x, "nFeaturesOptim")
  mat <- x[2] # assumes two-class problems
  dimnames(mat) <- list(observed = paste("observed", attr(x, "row.vars")[[1]]),
    predicted = paste("predicted", attr(x, "col.vars")[[1]])

  if(is.null(caption)) caption <- paste("Confusion matrix for classifier: ", tech,
    "; Optimal number of features: ",
    nFeaturesOptim, sep = "")
  xtable:::xtable.matrix(mat, caption = caption, label = label, align = align,
    digits = digits, display = display, ...)
}

install.packages("reporttools")
require(reporttools)
require(ggplot2)
library(lubridate)

```

```

library(ggpmisc)
library(reshape2)

data<-droplevels(data)
vars1<-data[, c("Region", "Operation","Work.Order.Type", "Planner.Group",
  "WO.Priority","Business.Impact","MRP.Type")]
cap1<-"Variables Cualitativas"
tableNominal(vars = vars1, cap = cap1, vertical = FALSE, longtable = TRUE)

vars2<-data[,c("ttr","tte","tft")]
cap2<-"Variables Cuantitativas"
tableContinuous(vars = vars2, cap = cap2, lab = "varcuant" ,longtable = FALSE)

vars3<-data[,c("Work.order.release.date","Work.Order.Created.D","Basic.Start.Date","Work.Order
cap3<-"Variables temporales"
tableDate(vars = vars3, print.pval = TRUE, cap = cap3, lab = "vardate", longtable
  = FALSE)

dataWO$WOPIFOT_N<-as.numeric(gsub("^([0-9]+),([0-9]+)$","\\1.\\2", dataWO$WOPIFOT))

wop_month<-as.data.frame(ddply(dataWO[dataWO$Work.Order.Type=="PM01",],
  .(Month,Year,Area),summarize, WOPIFOT=mean((as.numeric(WOPIFOT_N))))))
wop_month<-na.omit(wop_month)
wop_month$Month<-factor(wop_month$Month, levels = toupper(month.abb))
wop_month<-wop_month[order(wop_month$Year,wop_month$Month),]
wop_month$YM<-as.factor(paste(wop_month$Year,wop_month$Month,sep=" "))
wop_month$YM<-as.Date((ymd(wop_month$YM, truncated = 1L)))

ggplot(wop_month, aes(x=YM, y=WOPIFOT, label=sprintf("%0.2f", round(WOPIFOT,
  digits = 2)))) + geom_line() + ylim(40,100) +
  geom_text(size = 3, position = position_stack(vjust = 1),parse = T) +
  xlab("Month") + ylab("WOPIFOT (%)") +ggtitle("WOPIFOT PM01 Spence (Jun 2017 -
  Apr 2018)") + theme(plot.title = element_text(hjust = 0.5))

ggplot(wop_month, aes(x=YM, y=WOPIFOT, color=Area, label=sprintf("%0.2f",
  round(WOPIFOT, digits = 2)))) +
  geom_line() + ylim(40,100) +
  # geom_text(size = 3,parse = T) +
  xlab("Month") + ylab("WOPIFOT (%)") +ggtitle("WOPIFOT PM01 Spence (Jun 2017 -
  Apr 2018)") +
  theme(plot.title = element_text(hjust = 0.5)) + scale_x_discrete() +
  geom_point() + scale_y_continuous()

```



```

mif_month<-as.data.frame(ddply(dataWO[dataWO$Work.Order.Type=="PM01",],
  .(Month,Year),summarize, MIFOT=mean((MIFOT))))
mif_month<-na.omit(mif_month)
mif_month$Month<-factor(mif_month$Month, levels = toupper(month.abb))
mif_month<-mif_month[order(mif_month$Year,mif_month$Month),]
mif_month$YM<-as.factor(paste(mif_month$Year,mif_month$Month,sep=" "))
mif_month$YM<-as.Date((ymd(mif_month$YM, truncated = 1L)))

ggplot(mif_month, aes(x=YM, y=MIFOT, label=sprintf("%0.2f", round(MIFOT, digits =
  2)))) + geom_line() + ylim(40,100) +
  geom_text(size = 3, position = position_stack(vjust = 1),parse = T) +
  xlab("Month") + ylab("MIFOT (%)") +ggtitle("MIFOT PM01 Spence (Jun 2017 - Apr
  2018)") + theme(plot.title = element_text(hjust = 0.5))

```

2.14.3. Entrenar Gradient Boosting Trees

```

set.seed(135)
' %!in%' <- function(x,y)!(' %in%'(x,y)) #define "not in" operator
data<-data[data$Month.x %in% c("SEP","OCT","NOV","DEC","JAN","FEB","MAR","APR"),]
data<-data[data$Work.Order.Type=="PM01",]
data<-data[data$Area %in% c("Plant","Mine"),]

OTN<-unique(data$Work.Order.Number)
ind = sample(2, length(OTN), replace=T, prob=c(0.8,0.2)) #Index for data splitting
OTNtrain<-OTN[ind==1]
OTNtest<-OTN[ind==2]

wopifot.train = data[data$Work.Order.Number %in% OTNtrain ,] #Training set
wopifot.test = data[data$Work.Order.Number %in% OTNtest ,] #Testing set

# wopifot.train = data[data$Month.x %!in% c("MAR") ,] #Training set
# wopifot.test = data[data$Month.x %in% c("MAR") ,] #Testing set

start.time.gbm<-Sys.time()

wopifot.boost=gbm((WOPIFOT_N)~
  tft+ttr+WO.Priority+tte+factor(Material.Group)+Main.Work.Centre+MRP.Type+Business.Impact+Planne
  data=wopifot.train,
  distribution="bernoulli",
  n.trees = 50000,
  shrinkage = 0.005,
  interaction.depth =3,
  n.minobsinnode=20)

```

```

summary(wopifot.boost) #Most relevant variables
gbm.perf(wopifot.boost)
wopifot.boost$var.names

plot.gbm(wopifot.boost, (4) , )

pr.boost<-predict(wopifot.boost,wopifot.test,n.trees = 50000,
  type="response",single.tree = FALSE)
#pr.boost_<-apply(pr.boost,1,which.max)
pr.boost_<-round(pr.boost)

# pred.wop<-ifelse(pred.wop==0,"Miss","Hit")
# table(pred.wop,wopifot.test$WOPIFOT) ->tab
# tab
roc(wopifot.test$WOPIFOT_N,pr.boost_)
plot(roc(wopifot.test$WOPIFOT_N,pr.boost_))
# (tab[1,1]+tab[2,2])/sum(tab)
#confusionMatrix(ifelse(pred.wop==0,"Miss","Hit"),wopifot.test$WOPIFOT)
wop.cm<-confusionMatrix(pr.boost_,wopifot.test$WOPIFOT_N)
wop.cm
end.time.gbm<-Sys.time()

time.taken.gbm<-end.time.gbm-start.time.gbm
time.taken.gbm
#OTtest is the actual WOPIFOT for each WORK ORDER
OTtest<-ddply(wopifot.test, .(Work.Order.Number),summarize,
  WOPIFOT_R1=prod(WOPIFOT_N))

OTs<-as.data.frame(cbind("Work.Order.Number"=dataW$Work.Order.Number,"WOPIFOT_R"=as.numeric(gsub(
  dataW$WOPIFOT))))
Cross<-join(OTtest,OTs)
FullLOT<-as.data.frame(cbind(wopifot.test,"PredWOP"=pred.wop))
FullLOT_<-ddply(FullLOT, .(Work.Order.Number),summarize, WOPIFOT_P=prod(PredWOP))
OTconf.mat<-join(FullLOT_,OTtest)

confusionMatrix(ifelse(OTconf.mat$WOPIFOT_P==0,"Miss","Hit"),ifelse(OTconf.mat$WOPIFOT_R==0,"M
roc(OTconf.mat$WOPIFOT_P,OTconf.mat$WOPIFOT_R)

time.taken.gbm

start.time.gbm<-Sys.time()
library(party)
wop.tree<-ctree(WOPIFOT_N~
  tft+ttr+tte,#+Main.Work.Centre+MRP.Type+Business.Impact+Planner.Group+W0.Priority,
  data=wopifot.train,

```

```

        controls = ctree_control(mincriterion = 0.99,
                                minsplit=500, minbucket = 10))
plot(wop.tree, type="simple")

wop.tree.r<-predict(wop.tree,wopifot.test)
confusionMatrix(wopifot.test$WOPIFOT_N,round(wop.tree.r))
roc(wopifot.test$WOPIFOT_N,wop.tree.r)
end.time.gbm<-Sys.time()
time.taken.gbm<-end.time.gbm-start.time.gbm
time.taken.gbm

```

```

start.time.gbm<-Sys.time()
# install.packages("randomForest")
library(randomForest)
wop.rf<-randomForest(WOPIFOT_N~
    tft+ttr+tte+MRP.Type+Business.Impact+Planner.Group+WO.Priority,
    data=wopifot.train,
    ntree=100)

```

```

wop.rf.r<-predict(wop.rf,wopifot.test)
confusionMatrix(wopifot.test$WOPIFOT_N,round(wop.rf.r))
roc(wopifot.test$WOPIFOT_N,wop.rf.r)
end.time.gbm<-Sys.time()
time.taken.gbm<-end.time.gbm-start.time.gbm
time.taken.gbm

```

```

#----- RCA PREDICTION-----
dataRCA = read.csv('fulldata.csv',sep=";")

```

```

#Function to assign cause as text to factor levels
num.to.cause<- function(n){
  cause<-switch(n,
    # "BAU-Additional Part added on Requirement Date",
    "BAU-Incorrect Returns Process",
    "BAU-Parts Not Collected",
    "BAU-Possible Stocking Levels/Lead Time problem",
    "BAU-Possible vendor delivery",
    "BAU-Required outside lead time",
    "BAU-Required prior to WO release",
    "DP-Possible PR to PO Conversion",
    "Free Text - Required Outside Lead Time",
    "PUR-Co-ordinate free text order/delivery",

```

```

        "URGENT-Inventory Review non stocked material miss",
        "URGENT-Inventory Review stocked material miss",
        "URGENT-Parts Not Collected",
        "URGENT-Work Management Process Issues",
        "PUR-Free text not delivered as scheduled")
    return(cause)
}

#Function to assign Create.to.Requirement interval
create.to.requirement<-function(n){
  if (n<=7){m<-1}
  else if (n<=14){m<-2}
  else if (n<=21){m<-3}
  else {m<-4}

  ctr<-switch(m,
    "Inside 7 Days",
    "8 to 14 Days",
    "15 to 21 Days",
    "Over 21 Days")
  return(ctr)
}

dataRCA <- subset(dataRCA, dataRCA$Operation=="Spence")
dataRCA$Calculated.Reservation.Requirement.Date=as.Date(dataRCA$Calculated.Reservation.Requirement.Date,
  format = "%d-%m-%Y")
dataRCA$Work.Order.Created.D=as.Date(dataRCA$Work.Order.Created.D, format =
  "%d-%m-%Y")
dataRCA$Work.order.release.date=as.Date(dataRCA$Work.order.release.date, format =
  "%d-%m-%Y")
dataRCA$tte=as.integer(dataRCA$Calculated.Reservation.Requirement.Date-dataRCA$Work.order.release.date)
dataRCA$ttr=as.integer(dataRCA$Work.order.release.date
  -dataRCA$Work.Order.Created.D)

#Material description to string
dataRCA$Material.Description <-
  unlist(data.frame(lapply(dataRCA$Material.Description, as.character),
    stringsAsFactors=FALSE))

dataRCA$Material.Group<-unlist(lapply(dataRCA$Material.Description,matgroup))

```

```

dataRCA$Material.Group<-unlist(lapply(dataRCA$Material.Group,commonmat))
dataRCA$Material.Group<-unlist(lapply(dataRCA$Material.Group,as.factor))

#Set level structures for neural net training
levels(dataRCA$Material.Group)<-c(levels(dataRCA$Material.Group),levels(data$Material.Group))
levels(dataRCA$MRP.Type)<-c(levels(dataRCA$MRP.Type),levels(data$MRP.Type))
levels(dataRCA$WO.Priority)<-c(levels(dataRCA$WO.Priority),levels(data$WO.Priority))
levels(dataRCA$Material.Group)<-c(levels(dataRCA$Material.Group),levels(data$Material.Group))

str(dataRCA)

names(dataRCA)[72]<-"RC"
levels(dataRCA$RC)<-c(levels(dataRCA$RC),"PUR-Free text not delivered as
  scheduled")
dataRCA<-dataRCA[dataRCA$Work.Order.Type == "PM01",]

#
# -----
# -----BOOSTING TREES IMPORTANT VARIABLE
# -----DETECTION-----
#
# -----
ind = sample(2, nrow(dataRCA), replace=T, prob=c(0.5,0.5)) #Index for data
splitting
rca.train = dataRCA[ind==1,] #Training set
rca.train = rca.train[rca.train$RC != "BAU-Additional Part added on Requirement
  Date",]
rca.test = dataRCA[ind==2,] #Testing set
rca.test = rca.test[rca.test$RC != "BAU-Additional Part added on Requirement
  Date",]

RCA.boost=gbm(RC~
  ttr+tte+Work.Order.Type+Planner.Group+WO.Priority+Main.Work.Centre+MRP.Type+Material.Group,
  data=rca.train,
  distribution="multinomial",
  n.trees = 10000,
  shrinkage = 0.05,
  interaction.depth = 3,
  n.minobsinnode=10)

summary(RCA.boost) #Most relevant variables

```

```

pr.boost.rca<-predict(RCA.boost,rca.test,n.trees = 10000,
  type="response",single.tree = FALSE)
pr.boost.rca_<-apply(pr.boost.rca,1,which.max)
pr.boost.rca_<-sapply(pr.boost.rca_,num.to.cause)
table(pr.boost.rca_,rca.test$RC) ->tab.rca
tab.rca
mean(pr.boost.rca_==rca.test$RC)
confusionMatrix(pr.boost.rca_,rca.test$RC)

save.image("WPT.RData")
----- NEW DATA PREDICTION-----

```

2.14.4. Análisis y predicción de resultado de nuevas OT

```

#-----New data to predict-----

setwd('C:/Users/barrjg9/Desktop/WorkSpace')

OT <- read_xlsx("OTs.xlsx")
OT <- subset(OT, grepl("REL",OT$'System status'))
OTm<- read_xlsx("OT_material.xlsx")
OniquaFeb<-read.csv("OniquaFEB.csv", sep=";")
OniquaFeb$Business.Impact<-
  as.factor(sapply(as.character(OniquaFeb$Business.Impact),Bimp))

# KeepOT<- c("Order",
#           "Created on",
#           "Actual release date",
#           "Basic start date",
#           "Basic finish date",
#           "Main work center",
#           "Order Type",
#           "Priority text",
#           "Description",
#           "User Status",
#           "System status",
#           "Planning plant",
#           "Planner group",
#           "Entered by")
KeepOT<- c("Order",
          "Created on",
          "Release",
          "Bsc start",

```

```
"Basic fin.",
"Mn.wk.ctr",
"Type",
"Priority text",
"Description",
"UserStatus",
"System status",
"Plnt",
"PG",
"Entered by")
```

```
KeepOTm<- c("Order",
            "Reservation",
            "Material",
            "Material Description",
            "Requirements date",
            "Plant",
            "Requirement Quantity")
```

```
OT<-OT[,KeepOT]
OTm<-OTm[,KeepOTm]
```

```
names(OT)<- c("Work.Order.Number",
             "Work.Order.Created.D",
             "Work.order.release.date",
             "Basic.Start.Date",
             "Basic.Finish.Date",
             "Main.Work.Centre",
             "Work.Order.Type",
             "WO.Priority",
             "Description",
             "User.Status",
             "System.Status",
             "Planning.Plant",
             "Planner.Group",
             "User.Type")
```

```
names(OTm)<- c("Work.Order.Number",
              "Reservation",
              "Material.Number",
              "Material.Name",
              "Calculated.Reservation.Requirement.Date",
              "Plant",
              "Total.Reservation.Qty")
```

```

OTm$Material.Number[is.na(OTm$Material.Number)]<-"#"
OT$Main.Work.Centre<-paste(OT$Planning.Plant,OT$Main.Work.Centre,sep="/")
OT<-join(OT,mwc.area,match="first")
OTdata<-join(OTm,OT,match="first")

OTdata$KEY<-paste(OTdata$Material.Number,toupper(month.abb[as.integer(format(as.Date(OTdata$W
  "%m"))])),format(as.Date(OTdata$Work.Order.Created.D), "%Y"),sep="")

OTdata<-join(OTdata,OniquaFeb,match="first")
#Create Time variables
OTdata$today<-Sys.Date()
OTdata$Basic.Start.Date=as.Date(OTdata$Basic.Start.Date, format = "%Y-%m-%d")
OTdata$Work.Order.Created.D=as.Date(OTdata$Work.Order.Created.D, format =
  "%Y-%m-%d")
OTdata$Work.order.release.date=as.Date(OTdata$Work.order.release.date, format =
  "%Y-%m-%d")
OTdata$Calculated.Reservation.Requirement.Date=as.Date(OTdata$Calculated.Reservation.Requireme
  format = "%d.%m.%Y")

OTdata$ttr=as.integer(OTdata$Work.order.release.date -OTdata$Work.Order.Created.D)
  #Time to release
OTdata$tte=as.integer(OTdata$Basic.Start.Date-OTdata$Work.Order.Created.D) #Time
  to execute
OTdata$tft=as.integer(OTdata$today-OTdata$Work.Order.Created.D) #Time from today
OTdata$ctr=as.integer(OTdata$Calculated.Reservation.Requirement.Date-OTdata$Work.Order.Created
  #Time from today

#Create Material Groups (Material family)
#Material description to string
OTdata$Material.Group<-unlist(lapply(OTdata$Material.Name,matgroup))

OTdata$Material.Group<-sapply(OTdata$Material.Group,dupli)

#Define list of common materials, label the rest as OTHERS

OTdata$Material.Group<-unlist(lapply(OTdata$Material.Group,commonmat))
OTdata$Material.Group<-factor(OTdata$Material.Group)

#Replace MRPTYPE and Business.Impact NAs as "Not Assigned"
levels(OTdata$Business.Impact)<-c(levels(OTdata$Business.Impact),"Not assigned")
OTdata$Business.Impact[is.na(OTdata$Business.Impact)]<-"Not assigned"
levels(OTdata$MRP.Type)<-c(levels(OTdata$MRP.Type),"Not assigned")
OTdata$MRP.Type[is.na(OTdata$MRP.Type)]<-"Not assigned"
OTdata$MRP.Type[is.na(OTdata$MRP.Type)]<-"Not assigned"
levels(OTdata$Material.Group)<-c(levels(OTdata$Material.Group),"Free Text")

```



```

OTdata$Material.Group[OTdata$Material.Number=="#"]<-"Free Text"

#Create create.to.requirement interval variable
OTdata$Create.to.Requirement <- sapply(OTdata$ctr,create.to.requirement)

#-----COMPUTE PREDICTED RESULTS-----
prediction<-predict(wopifot.boost,OTdata,n.trees = 30000,
  type="response",single.tree = FALSE)
prediction_<-apply(prediction,1,which.max)
prediction_<-abs(prediction_-2)

#-----COMPUTE PREDICTED RCA-----
Predicted.OTm<-OTdata
#Predicted.OTm<-unique(Predicted.OTm)

Predicted.OTm$Prediction<-prediction_

predictionRCA<-predict(RCA.boost,Predicted.OTm,n.trees = 20000,
  type="response",single.tree = FALSE)
predictionRCA_<-apply(predictionRCA,1,which.max)
predictionRCA_<-sapply(predictionRCA_,num.to.cause)

Predicted.OTm$RCA<-predictionRCA_
Predicted.OTm[Predicted.OTm$Prediction==1,]$RCA<-"-"

#-----ADD CONSIDERATIONS-----
Predicted.OTm$Considerations <- "None"
Predicted.OTm[Predicted.OTm$MRP.Type=="PD",]$Considerations<- "PD Material";
Predicted.OTm[Predicted.OTm$MRP.Type=="ZD",]$Considerations<- "ZD Material";

Predicted.OTm[Predicted.OTm$Material.Group=="Free Text",]$Considerations<- "Free
Text";
Predicted.OTm[Predicted.OTm$Material.Group=="DUPLICATED",]$Considerations<-
"Duplicated Material Number";
Predicted.OTm[Predicted.OTm$Considerations=="Duplicated Material
Number",]$Prediction<-0;
Predicted.OTm[Predicted.OTm$Considerations=="Duplicated Material
Number",]$RCA<-"Duplicated Material Number";

###OLD OTS Marked as MISS?
Predicted.OTm[Predicted.OTm$tte>365,]$Prediction<-0
Predicted.OTm[Predicted.OTm$tte>365,]$Considerations<-"WO is too old"
if (Predicted.OTm[Predicted.OTm$tte>365,]$RCA %in% c("BAU-Parts Not
Collected","")) {
  Predicted.OTm[Predicted.OTm$tte>365,]$RCA <- "WO is too old"
}

```

```
}
```

```
###Parts Not Collected as HIT  
Predicted.OTm[Predicted.OTm$RCA=="BAU-Parts Not Collected",]$Prediction<-1
```

2.14.5. Incorporación y cruce de data de inventario

```
#-----MAR  
INTERACTION-----  
  
MAR<-read_xlsx("MAR.xlsx", skip=13, sheet="MAR Tracker")  
names(MAR)[3]="Work.Order.Number"  
MAR$'Overall Status'<-as.numeric(as.character(MAR$'Overall Status')) #Overall  
status to numeric  
sum(is.na(MAR$'Overall Status')) #Total of NA Overall Status  
# MAR[is.na(MAR$'Overall Status'),]$'Overall Status'<-0.1 #NA in overall status to  
0.1  
  
#Join MAR and WPT  
Predicted.OTm<-join(Predicted.OTm,  
as.data.frame(cbind("Work.Order.Number"=MAR$Work.Order.Number,  
"Material.Number"=MAR$Material,  
"Overall Status"= MAR$'Overall  
Status')), match="first")  
  
unique(Predicted.OTm$RCA)  
prop.table(table(Predicted.OTm$RCA))*100 # Percentage of each cause  
prop.table(table(Predicted.OTm[Predicted.OTm$RCA!="-"],$RCA))*100 # Percentage of  
each cause excluding hits  
table(Predicted.OTm[Predicted.OTm$RCA!="-"],$RCA)  
mean(Predicted.OTm$Prediction==Predicted.OTm$'Overall Status', na.rm = T)  
#Predictvie model vs MAR concordance  
  
# #Remove bugged data from overall status (dates)  
# unique(Predicted.OTm$'Overall Status')  
# Predicted.OTm[Predicted.OTm$'Overall Status' %!in% c(0,0.1,0.5,1),]<- NA  
#  
  
table(Predicted.OTm$'Overall Status')  
prop.table(table(Predicted.OTm$'Overall Status')) #MIFOT estimation
```

```

# #Calculate WOPIFOT predicted by MAR
# Predicted.OT$WOPIFOT.MAR<- ddply(Predicted.OTm, .(Work.Order.Number), summarize,
  WOPIFOT=prod('Overall Status')[,2]
# Predicted.OT$WOPIFOT.MAR<-ceiling(Predicted.OT$WOPIFOT.MAR)
# mean(Predicted.OT$WOPIFOT.MAR)

#Predicted.OTm<-na.omit(Predicted.OTm)
Predicted.OTm[is.na(Predicted.OTm$'Overall Status'),]$'Overall Status'<-0.25

#Remove false positives (Incorrect miss prediction)
Predicted.OTm$Accountability<-""
Predicted.OTm[Predicted.OTm$RCA=="BAU-Required outside lead time" &
  Predicted.OTm$'Overall Status'==1,]$Prediction<-1
Predicted.OTm[Predicted.OTm$RCA=="BAU-Possible Stocking Levels/Lead Time problem"
  & Predicted.OTm$'Overall Status'==1,]$Prediction<-1
Predicted.OTm[Predicted.OTm$RCA=="BAU-Possible vendor delivery" &
  Predicted.OTm$'Overall Status'==1 ,]$Prediction<-1
Predicted.OTm[Predicted.OTm$RCA=="BAU-Required outside lead time" &
  Predicted.OTm$'Overall Status'==1,]$RCA<--"
Predicted.OTm[Predicted.OTm$RCA=="BAU-Possible vendor delivery" &
  Predicted.OTm$'Overall Status'==1 ,]$RCA<--"

row(subset(Predicted.OTm, Predicted.OTm$RCA=="BAU-Possible vendor delivery" &
  Predicted.OTm$'Overall Status'==1))

#Add accountability
Predicted.OTm$Accountability<-""
Predicted.OTm[na.omit(Predicted.OTm$RCA=="BAU-Required outside lead time" &
  Predicted.OTm$Prediction!=1),]$Accountability<-"Inventory Optimization"
Predicted.OTm[na.omit(Predicted.OTm$RCA=="BAU-Possible Stocking Levels/Lead Time
  problem" & Predicted.OTm$Prediction!=1),]$Accountability<-"Inventory
  Optimization"
Predicted.OTm[na.omit(Predicted.OTm$RCA=="BAU-Parts Not Collected" &
  Predicted.OTm$Prediction!=1) ,]$Accountability<-"Inventory Planner"
Predicted.OTm[na.omit(Predicted.OTm$RCA=="BAU-Possible vendor delivery" &
  Predicted.OTm$Prediction!=1) ,]$Accountability<-"Focal Point"
Predicted.OTm[na.omit(Predicted.OTm$Considerations=="Duplicated Material
  Number"),]$Accountability<-"Inventory Optimization"
Predicted.OTm[na.omit(Predicted.OTm$Considerations=="WO is too
  old"),]$Accountability<-"Inventory Planner"
Predicted.OTm[na.omit(Predicted.OTm$Prediction==1),]$Accountability<--""
Predicted.OTm[na.omit(Predicted.OTm$Prediction==1),]$RCA<--""
Predicted.OTm[na.omit(Predicted.OTm$Prediction==1),]$Recomendation<--""

```

```

#Add recomendation
Predicted.OTm$Recomendation<-" "
Predicted.OTm[na.omit(Predicted.OTm$RCA=="BAU-Required outside lead time" &
  Predicted.OTm$'Overall Status'!=1) ,]$Recomendation<-"Update material
  parameters"
Predicted.OTm[na.omit(Predicted.OTm$RCA=="BAU-Parts Not Collected" &
  Predicted.OTm$'Overall Status'!=1) ,]$Recomendation<-"Contact WO User"
Predicted.OTm[na.omit(Predicted.OTm$RCA=="BAU-Possible vendor delivery" &
  Predicted.OTm$'Overall Status'!=1) ,]$Recomendation<-"Expediting"
Predicted.OTm[na.omit(Predicted.OTm$Considerations=="Duplicated Material
  Number"),]$Recomendation<-"Remove duplicates"
Predicted.OTm[na.omit(Predicted.OTm$Considerations=="WO is too
  old"),]$Recomendation<-"Check for replanification causes"

#Business Impact number code
BimpN<-function(s){
  ifelse(is.na(s), return(0),
  BIN<- ifelse(s=="Not Assigned",0,
    ifelse(s=="A-Major Production Stopper", 3,
      ifelse(s=="B-Minor Production Stopper",2,
        ifelse(s=="C-Major Maintenance Impact",3,
          ifelse(s=="D-Minor Maintenance Impact",2,
            ifelse(s=="E-No Impact",1,0))))))
  return(BIN)
}

Predicted.OTm$Business.Impact.N<-sapply(Predicted.OTm$Business.Impact,BimpN)
Predicted.OTm$Business.Impact.N<-Predicted.OTm$Business.Impact.N

#Add material prices
prices<-read_xlsx("prices.xlsx")
names(prices)[1]<- "Material.Number"
Predicted.OTm<-join(Predicted.OTm,prices, by="Material.Number", match="first")
Predicted.OTm[is.na(Predicted.OTm$'Average Unit Price'),]$'Average Unit Price'<-0
Predicted.OTm$WPTMAR<- Predicted.OTm$Prediction##Predicted.OTm$'Overall Status'

Predicted.OT<- ddply(Predicted.OTm, .(Work.Order.Number),summarize,
  WOPIFOT=prod(WPTMAR), Price=sum('Average Unit Price'), "Business
  Impact"=max(Business.Impact.N))
Predicted.OT$WOPIFOT<-trunc(Predicted.OT$WOPIFOT)
Predicted.OT<- join(Predicted.OT,OT[,c("Work.Order.Number",
  "Basic.Start.Date","Area", "Description", "Main.Work.Centre")],match="first")
Predicted.OT<-na.omit(Predicted.OT)

write.csv(Predicted.OT,"OT Prediction.csv")
write.csv(Predicted.OTm, "OT Material Prediction.csv")

```

```
Dup<-Predicted.OTm[Predicted.OTm$RCA=="Duplicated Material Number",]  
LTime<-Predicted.OTm[Predicted.OTm$RCA=="BAU-Required outside lead time",]  
Old<-Predicted.OTm[Predicted.OTm$Considerations=="WO is too old",]  
Vendor<-Predicted.OTm[Predicted.OTm$RCA=="BAU-Possible vendor delivery",]  
Stock<-Predicted.OTm[Predicted.OTm$RCA=="BAU-Possible Stocking Levels/Lead Time  
problem",]  
Freetext<-Predicted.OTm[Predicted.OTm$Considerations=="Free Text",]
```

```
write_xlsx(Dup,"WPT/Duplicados.xlsx")  
write_xlsx(LTime,"WPT/Possible LeadTime Problem.xlsx")  
write_xlsx(Old,"WPT/Old Work Orders.xlsx")  
write_xlsx(Vendor,"WPT/Possible Vendor Miss.xlsx")  
write_xlsx(Stock,"WPT/Possible Stockout.xlsx")  
write_xlsx(Freetext,"WPT/Free Text.xlsx")
```

```
IO<-unique(as.data.frame(Predicted.OTm[Predicted.OTm$RCA=="BAU-Required outside  
lead time" & Predicted.OTm$'Overall Status'!=1 ,]))  
IP<-unique(as.data.frame(Predicted.OTm[Predicted.OTm$RCA=="BAU-Parts Not  
Collected" & Predicted.OTm$'Overall Status'!=1 ,]))  
FP<-unique(as.data.frame(Predicted.OTm[Predicted.OTm$RCA=="BAU-Possible vendor  
delivery" & Predicted.OTm$'Overall Status'!=1 ,]))
```
