

Desarrollo de un modelo de análisis de improductividad en una línea de ensamble de refrigeración para empresas de manufactura en el Ecuador

Orellana Cordero, Fabián Roberto

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios

Trabajo de titulación, previo a la obtención del título de Magíster en Gestión de Sistemas de Información e Inteligencia de Negocios

Msc. Zaldumbide Proaño, Juan Pablo

16 de septiembre del 2020

URKUND

Document Information

Analyzed document	TesisFinal_FabianOrelana.docx (D79194781)
Submitted	9/15/2020 7:17:00 PM
Submitted by	
Submitter email	frelana3@espe.edu.ec
Similarity	2%
Analysis address	vcparraga_espe@analysis.orkund.com

Sources included in the report

SA	Universidad de las Fuerzas Armadas ESPE / Tesis_ArevaloMateo_v010_Final.pdf Document Tesis_ArevaloMateo_v010_Final.pdf (D45995365) Submitted by: mpdlaz@espe.edu.ec Receiver: mpdlaz.espe@analysis.orkund.com	 5
W	URL: https://repositorio.espe.edu.ec/bitstream/21000/18792/1/T-ESPE-038816.pdf Fetched: 7/1/2020 6:02:57 AM	 1
SA	Universidad de las Fuerzas Armadas ESPE / Tesis_VivianaParraga_orkund.docx Document Tesis_VivianaParraga_orkund.docx (D40820534) Submitted by: ertonseca@espe.edu.ec Receiver: ertonseca.espe@analysis.orkund.com	 1
W	URL: https://www.paradigmadigital.com/tecbiz/comparativa-plataformas-business-intelig... Fetched: 9/15/2020 7:17:00 PM	 2

Firma:



Ing. Zaldumbide Proaño, Juan Pablo MSc.

Director

C.C.: 1715467948



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA**

CENTRO DE POSGRADOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, **“Desarrollo de un modelo de análisis de improductividad en una línea de ensamble de refrigeración para empresas de manufactura en el Ecuador”** fue realizado por el ingeniero **Orellana Cordero, Fabián Roberto** el mismo que ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 15 de septiembre de 2020

Firma:

Ing. Zaldumbide Proaño, Juan Pablo MSc.

Director

C.C.: 1715467948



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

RESPONSABILIDAD DE AUTORÍA

Yo, **Orellana Cordero, Fabian Roberto** con cédula de ciudadanía n° 0103663233, declaro que el contenido, ideas y criterios del trabajo de titulación: **“Desarrollo de un modelo de análisis de improductividad en una línea de ensamble de refrigeración para empresas de manufactura en el Ecuador”** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 16 de septiembre de 2020

Firma

Ing. Orellana Cordero Fabián Roberto
C.C. 0103663233



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA
CENTRO DE POSGRADOS**

AUTORIZACIÓN DE PUBLICACIÓN

Yo **Orellana Cordero, Fabian Roberto** autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **“Desarrollo de un modelo de análisis de improductividad en una línea de ensamble de refrigeración para empresas de manufactura en el Ecuador”** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 16 de septiembre de 2020

Firma

Ing. Orellana Cordero Fabián Roberto
C.C. 0103663233

Dedicatoria

A mis padres por haberme dado las herramientas para ser una mejor persona; muchas de mis logros los agradezco a Uds. entre de los que incluyo este. Me formaron con las mejores reglas y me motivaron constantemente para alcanzar mis ideales.

Agradecimiento

Agradezco a mis profesores, personas de grandes conocimientos que me han ayudado a llegar a este punto en el que me encuentro. No ha sido fácil el llegar a este punto, muchas fronteras son las que he tenido que pasar para tener mi tesis con éxito y obtener mi titulación profesional.

A mis padres, gracias por su apoyo y cariño brindado al estar siempre en los momentos más importantes de mi vida. Este logro es de Uds.

A mis hijos, que con su diario amor me han hecho una mejor persona y sobre todo impulsar a que pueda ser cada día mejor.

A mi novia, por todo el apoyo, comprensión y confianza que me ha brindado en los momentos difíciles y hacer de los buenos más placenteros. Por motivarme a ser mejor cada día de mi vida.

Índice de contenidos

<i>Carátula</i>	1
<i>Certifico del director</i>	2
<i>Responsabilidad de Autoría</i>	4
<i>Autorización</i>	5
<i>Dedicatoria</i>	6
<i>Agradecimiento</i>	7
<i>Índice de contenidos</i>	8
<i>Índice de tablas</i>	10
<i>Índice de figuras</i>	11
<i>Resumen</i>	12
<i>Abstract</i>	13
<i>Capítulo I. Introducción</i>	14
Antecedentes	14
Justificación e importancia.....	15
Planteamiento del problema.....	15
Objetivos General	17
Objetivos Específicos	17
Hipótesis	18
Alcance	19
Preguntas de investigación.....	20
Fundamentación teórica	21
<i>Capítulo II. Revisión de Literatura</i>	28
Antecedentes del estado del arte	28
Herramientas.....	33
Técnicas de minería de datos	38
Algoritmos de minería de datos.....	39
<i>Capítulo III. Metodología</i>	41
Metodología de investigación	41

Metodología de minería de datos	47
Esquema general del proyecto	50
<i>Capítulo IV. Desarrollo del estudio</i>	<i>52</i>
Recolección y análisis de la información.....	52
Limpieza de los datos	56
Elección herramientas y algoritmos	68
Dashboard descriptivos	74
Creación del Modelo.....	80
Evaluación del modelo.....	85
<i>Capítulo V. Conclusiones y Recomendaciones</i>	<i>87</i>
Conclusiones.....	87
Recomendaciones	88
<i>Bibliografía</i>	<i>89</i>
<i>Anexo A. Artículos revisados por repositorio académico</i>	<i>94</i>

Índice de tablas

Tabla 1. Construcción de la cadena de búsqueda	30
Tabla 2. Tabla comparativa de metodologías del basadas en el diseño	42
Tabla 3. Pros & Contras de Tableau.....	72

Índice de figuras

Figura 1. Relación de variables y fundamentación teórica	21
Figura 2. Ciclo metodología CRISP-DM.....	24
Figura 3. Artículos encontrados en los repositorios académicos.	31
Figura 4. Productos disponibles Tableau.....	36
Figura 5. Investigación mediante ciclos.....	45
Figura 6. Flujo de la metodología CRISP-DM.....	48
Figura 7. Esquema general del proyecto.....	51
Figura 8. Base de datos inicial.....	54
Figura 9. Modelo entidad relación	55
Figura 10. Modelo multidimensional.....	56
Figura 11. Proceso limpieza de datos	57
Figura 12. Paso uno del proceso.....	58
Figura 13. Programación en Python.....	58
Figura 14. Datos traídos desde MySQL	59
Figura 15. Código de depuración de la data, revisión instancias	60
Figura 16. Código depuración de la data, agrupaciones	61
Figura 17. Código de inserción data.....	62
Figura 18. Ejemplo código carga de datos a dimensión	63
Figura 19. Comandos de población masiva	64
Figura 20. Código carga de datos a FACT.....	65
Figura 21. Datos cargados de la FACT.....	66
Figura 22. Datos cargados de la dimensión Gestión	66
Figura 23. Datos cargados de la dimensión Empleados	67
Figura 24. Datos cargados de la dimensión Empresa	67
Figura 25. Datos cargados de la dimensión Tiempo.....	68
Figura 26. Cuadrante Mágico de Gartner 2020 herramientas BI.....	70
Figura 27. Interfaz Tableau.....	71
Figura 28. Cuadrante Mágico de Gartner para herramientas de Minería de Datos.....	73
Figura 29. Dashboard en herramienta Tableau	75
Figura 30. Informe descriptivo general.....	76
Figura 31. Informe descriptivo Paradas.....	77
Figura 32. Informe descriptivo Llamadas Bodega.....	78
Figura 33. Informe descriptivo Llamadas Mantenimiento	79
Figura 34. Informe descriptivo Llamadas Producción	80
Figura 35. Ejemplo para modelo.....	82
Figura 36. Modelo con algoritmo Lineal	83
Figura 37. Modelo con algoritmo Deep Learning	84
Figura 38. Modelo con algoritmo Decision Tree.....	85
Figura 39. Resultados evaluación modelos	86

Resumen

Actualmente, las líneas de ensamble de empresas de manufactura son factores en la cadena de valor y se refieren a las maquinarias, personal calificado y un sin número de materiales para su fabricación. En la empresa caso de estudio se había detectado que en algunos casos no se tienen los resultados esperados, por la deficiencia de los insumos antes mencionados, ocasionando retrasos, deficiencia y por ende costos adicionales.

El propósito del proyecto fue mejorar la efectividad de las líneas de producción de la empresa. Según los lineamientos de la metodología orientada al diseño y CRISP DM, se comenzó con el entendimiento del negocio y datos de la empresa de manufactura, que permitió adquirir información almacenada en bases de datos referente al comportamiento de las líneas de producción. Estos datos fueron depurados en la plataforma Oracle, para luego presentar un conjunto de dashboards desarrollados en Tableau, para visualizar el comportamiento de los “tiempos de para” en la línea de ensamble. Para identificar posibles patrones predictivos que provocan las paras de planta, se elaboró un modelo de minería de datos utilizando Rapidminer. Como resultado de esta implementación la empresa cuenta con un sistema de apoyo, que les permite prevenir las paras de planta, de acuerdo con variables como, falta de materiales por gestión logística, prevención de mantenimiento de equipos y problemas de la gestión de la calidad.

Palabras claves:

- **LÍNEA DE ENSAMBLAJE**
- **TIEMPO DE PARA**
- **TABLEAU**
- **ORACLE**
- **RAPIDMINER**

Abstract

At present, the lines of ensemble of companies of manufacture are factors in the chain of value and they refer to the machineries, qualified personnel and one without number of materials for his manufacture. In the company had detected that in any cases do not have the awaited results, for the shortcoming of the inputs earlier mentioned, causing delays, it is like that, that this company took as a case of study.

The intention of the project was to improve the effectiveness of the lines of production of the company. According to the lineaments of the methodology faced to the design and CRISP DM, it was begun by the understanding of the business and information of the company of manufacture, which allowed to acquire information stored in bases of information regarding the behavior of the lines of production. This information was purified in the platform Oracle, then to present a set of dashboards developed in Tableau, to visualize the behavior of the “times of for” in the line of ensemble. To identify possible patterns predictive that they provoke you stop them of plant, there was prepared a model of mining of information using Rapidminer. As result of this implementation the company relies on with a system of support, which he allows them to prepare that you stop them of plant, in accordance with variables as, it is absent of materials for logistic management, prevention of maintenance of teams and problems of the management of the quality.

Key words:

- **ASSEMBLY LINE**
- **TIME TO**
- **TABLE**
- **ORACLE**
- **RAPIDMINER**

Capítulo I. Introducción

Antecedentes

Con la masificación de las metodologías ágiles y sistemas de control muchas empresas manufactureras como Toyota y General Motors se han proliferado ofreciendo miles de productos a los usuarios. El gran catálogo de productos que actualmente se ofrecen, nacen debido al uso de la metodología *Lean Manufacturing*, que se ha convertido en una ayuda para eliminar o reducir actividades en los procesos que no agreguen valor dentro de la producción (Liker, 2010).

Históricamente la revolución industrial permitió el crecimiento de compañías de producción masiva con métodos complejos para la producción de sus productos, generando varios procesos innecesarios como la sobreproducción, tiempo de esperas entre procesos, excesos de inventarios, transportes innecesarios, defectos, tareas y movimientos innecesarios, etc. Además, muchas de las empresas que fueron líderes en la industria no lograron innovar al nivel que las metodologías y las tecnologías avanzaban y se quedaron rezagadas en relación con su competencia, en algunos casos fueron absorbidas por otras compañías y en otros casos quebraron (Ravé, 2019)

En el Ecuador las empresas dedicadas al desarrollo de productos están buscando la manera de innovar para evitar quebrar o ser absorbidas por empresas más grandes. Una de las estrategias ha sido utilizar las tendencias existentes en el mundo como la implementación de metodologías ágiles, adquisiciones de nuevas tecnologías y personal capacitado para desarrollarse. Luego de realizar revisiones sobre empresas internacionales se ha detectado que la tendencia se va por incorporar metodologías que

permitan mejorar los procesos, ya que los procesos actuales así sean funcionales no garantizan toda la rentabilidad de la empresa (Liker, 2010).

Para contribuir con una solución efectiva al inconveniente de no tener información estructurada. El presente proyecto utiliza técnicas de inteligencia de negocios donde encontró las relaciones entre la pérdida de dinero y paras en la producción. Que permiten determinar mediante el uso de indicadores analíticos, acciones en la línea y evitando el no cumplimiento del proceso de producción.

Justificación e importancia

Actualmente el Ecuador se encuentra en un proceso de cambio y mejora impulsado desde el gobierno, con el cambio de la matriz productiva. Esto ha obligado a las empresas a optimizar sus procesos y mejorar sus prácticas para ser más competitivas (Senplades, 2012). Entre estas prácticas está la optimización de la productividad de las empresas, teniendo como uno de sus parámetros la optimización y desarrollo de la industrialización, ya que en la actualidad se realizan grandes esfuerzos económicos para generar un producto o marca.

Las empresas se están ajustando a los cambios tecnológicos y aprovechan la información histórica, antes que las mismas desaparezcan debido a los costos operativos que incurren en tenerlas almacenadas. Además, el desperdicio de recursos humanos y económicos en paras de procesos productivos, pasan al costo de sus productos, dejando de tener productos competitivos.

Planteamiento del problema

La empresa caso de estudio tiene como giro de negocio la manufactura de productos de línea blanca para el hogar. La empresa mantiene una línea de producción

de electrodomésticos de refrigeración. Dentro de la línea de producción se tienen varias etapas para el ensamble de una refrigeradora, cada etapa tiene una serie de pulsantes que se presionan cuando se tiene algún inconveniente. Cuando se realiza la pulsación se genera información a un sistema transaccional. Esta información carecía de estandarización y no se contaba con más reportería que la verbalización de cualquier evento por parte de un supervisor.

Las empresas actualmente no se manejan de esta manera, si no que usan la información histórica, actual y preventiva para indagar en las falencias de sus líneas de producción para tomar decisiones de las posibles causas de paras en la línea y con esto aumentan el número unidades en la línea de producción y por ende sus ganancias.

Es así como empresas como las dedicadas a la producción de electrodomésticos gastan grandes sumas de dinero anualmente por falta del cumplimiento en su presupuesto de producción. Estas paras en la línea generalmente ocurren por problemas logísticos, calidad o daños en la maquinaria. Por este motivo las empresas de manufactura cuentan con personal especializado para realizar revisiones de la línea de producción por medio de normativas de las metodologías *Lean Manufacturing*.

Los problemas con este tipo de revisiones es que son apaga incendios, son métodos correctivos que no solventaban las paras de las líneas de producción; por ejemplo, la línea de producción tenía una meta de producir por cada minuto un producto de refrigeración, esta tarea se la cumplía en cada una de las estaciones de trabajo que son cinco en la línea de refrigeración, cada uno de los operarios contaba con un documento, llamado hojas de trabajo estandarizado, en donde se tenían los pasos a seguir para el correcto cumplimiento de la actividad de la que es responsable. Cuando estos pasos no se realizaban en su cabalidad se debía a que existían problemas con el proceso de producción o se terminó el material y no había una generación de información

por el área logística o no existía revisión temprana por parte de mantenimiento en el daño de un equipo.

Por estos motivos, la línea de producción no contaba con los resultados esperados en su productividad planificada para su día de trabajo, al no realizar los resultados esperados se buscó alternativas, como el uso de herramientas de inteligencia de negocios, para implementar sistemas que les permitió generar información descriptiva así como predictiva de parámetros de línea, de acuerdo, a horas de jornadas de trabajo, equipos de trabajo menos productivos, falta de capacitación del personal, etc. para ofrecer la productividad que se tenía esperada.

Objetivos General

Implementar un sistema descriptivo de información en una planta de ensamble de productos de línea blanca que, mediante el análisis de información histórica permita identificar cuáles son las posibles causas para que se detenga su línea de producción, con la finalidad de mejorar la efectividad en su meta de producción.

Objetivos Específicos

OE1: Realizar una revisión inicial de la literatura, acerca del proceso de producción en empresas de manufactura y la calidad de los productos en sus líneas de fabricación para determinar las técnicas de sistemas descriptivos de información y seleccionar la que mejor se ajuste a las necesidades de la organización.

OE2: Recolectar y analizar la información existente en la línea de producción de refrigeración de la empresa en cuestión, para identificar valores concurrentes que puedan ratificar los resultados.

OE3: Identificar los algoritmos de minería de datos que serán utilizados en el modelo a diseñar utilizando una herramienta de minería de datos, para buscar la que mejor se adapte a los objetivos a los datos a analizar.

OE4: Crear el modelo de minería de datos mediante el algoritmo para determinar los patrones que afectan la baja producción y calidad de los productos.

OE5: Evaluar el modelo analítico – predictivo a través del uso de técnicas de validación implementadas en minería de datos, para determinar el nivel de confianza en un resultado predictivo.

Hipótesis

La creación de un sistema descriptivo y de análisis de productividad basado en información histórica permitió identificar las principales variables para asegurar el cumplimiento de la producción.

Categorización de las variables de investigación:

Variable dependiente: Factores que influyen en el aprovechamiento de la producción y calidad de los productos a producir.

Variable independiente: Modelo de minería de datos.

La hipótesis planteada consideró utilizar información de maquinaria, personal, logística y materiales que son de naturaleza descriptiva y cuantificables lo que permitió establecer predicciones en los resultados.

Para la demostración de la hipótesis planteada se consideró evaluar los resultados del modelo analítico predictivo aplicando técnicas de validación de minerías de datos que muestran el tipo de predicciones correctas e incorrectas cuando se aplica un modelo sobre un conjunto de datos de prueba.

Además, la demostración de la hipótesis planteada se dio mediante la comprobación estadística, evaluando los resultados del modelo usando técnicas de validación de minería de datos.

Alcance

El presente proyecto disminuyó los costos de producción al mejorar los resultados de productividad, mediante el uso de la Inteligencia de Negocios. Para lograr estos objetivos se utilizó metodologías de inteligencia de negocios, para generar predictivos de parámetros de planta a un segmento específico, de las posibles causas que tenga mayor probabilidad de afectar el mismo. Este análisis generó un valor agregado para la empresa mediante la mejora de su productividad y el aumento de calidad de sus productos.

Esta investigación recolectó y analizó la información para definir las fuentes de datos a ser utilizadas, así como también, los datos con valores atípicos para que sean excluidos, a fin de que no alteren los resultados. Luego se procedió a la selección e implementación del modelo a ser utilizado, acompañado del formato de los datos, para que los mismos puedan ser utilizados de acuerdo con los requerimientos del modelo. A continuación, se procedió a evaluar los resultados juntamente con los expertos del negocio, quienes determinaron si los resultados obtenidos son óptimos, para esto se tomó en cuenta el 70% de la data de la base de datos del sistema de la organización para las pruebas y para el aprendizaje del sistema, el 30% restante se utilizó para verificar la eficacia del sistema mediante la métrica seleccionada.

El presente proyecto generó un listado de métricas o indicadores para los equipos de trabajo con los incidentes que puedan ser de interés para ellos, esto permitió que la sección busque la mejor manera de contactar al supervisor para mostrar los posibles fallos del proceso

Preguntas de investigación

Para la consecución de los objetivos específicos del proyecto para el desarrollo de un modelo de análisis de improductividad en una línea de ensamble de refrigeración para empresas de manufactura en el Ecuador, requirió se respondan las siguientes preguntas para cada objetivo específico:

OE1 -RQ1.1: ¿Cuáles son los estudios existentes en la actualidad sobre sistemas de información descriptiva para empresas de Manufactura en Refrigeración?

OE1- RQ1.2: ¿Cuáles son las técnicas para la generación de sistemas de analíticos que mejor se ajustan a la realidad de la organización y por qué?

OE2 – RQ2.1: ¿Generación de búsqueda de información para poder revisar cuáles son las fuentes de datos con las que cuenta la organización?

OE2 – RQ2.2: ¿Existen valores atípicos que puedan afectar el resultado de la ejecución de los algoritmos de inteligencia de negocios?

OE3 – RQ3.1: ¿Qué tipo de datos necesita el algoritmo de inteligencia de negocios para sistemas analíticos para su correcto funcionamiento?

OE3 – RQ3.2: ¿Es posible acoplar los datos existentes para que se ajusten a las necesidades del algoritmo de inteligencia de negocios para sistemas de análisis predictivos?

OE4 – RQ4.1: ¿Cuál es el nivel de confianza con el que se va a trabajar para el análisis de los resultados?

OE4 – RQ4.2: ¿Cuál es el tipo de métrica a utilizar para el modelo seleccionado?

OE5 – RQ5.1: ¿Si es posible validar el modelo, se puede determinar un nivel de confianza en los resultados que permita demostrar una mejora en los mismos?

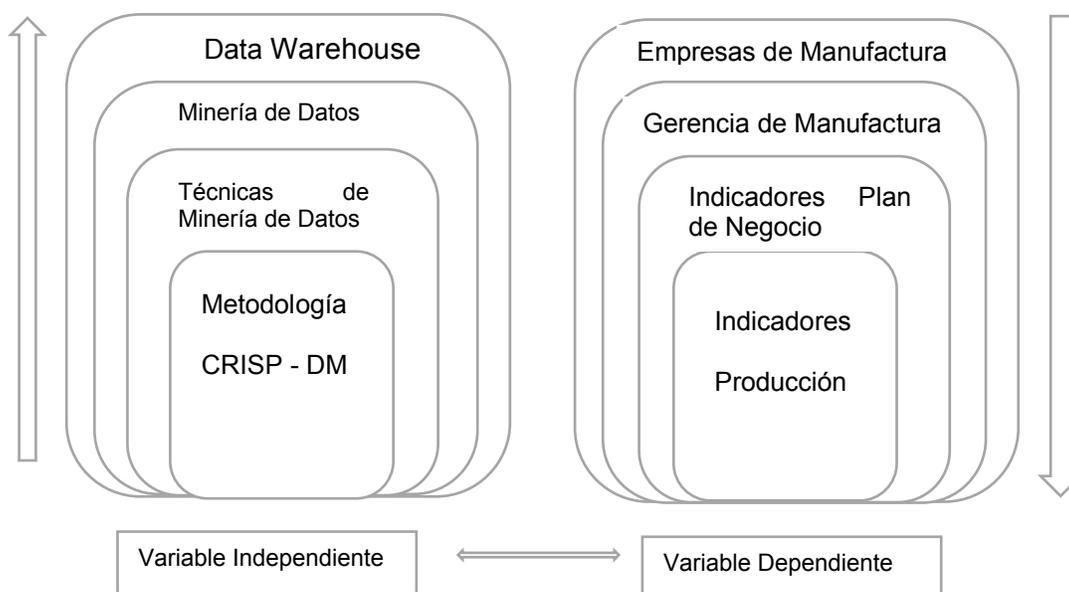
OE5 – RQ5.2: ¿Cuál es el margen de error que se debe considerar al implementar modelos de analítica predictiva?

Fundamentación teórica

La fundamentación teórica buscó la congruencia de los conceptos teóricos con la hipótesis planteada en este proyecto, por esta razón se hizo un análisis de la teoría usando las variables del problema. El propósito de este análisis fue investigar jerárquicamente cada categoría hasta llegar a la categoría que comprendió y explicó las variables dependientes e independientes del tema de estudio, para esto se propuso la jerarquía de estudio mostrada en la **Error! Reference source not found.**

Figura 1.

Relación de variables y fundamentación teórica



Nota. El gráfico presenta la relación entre variables dependientes e independientes para la definición del marco teórico.

Fundamentación de la variable independiente

Data Warehouse

Conocido también como almacén de datos y utilizado como un sistema de almacenamiento de diversas fuentes de datos para administrarlos y proveer información significativa a la empresa. De acuerdo con las necesidades de consulta de la organización, un *Data Warehouse* almacena los datos durante un período de tiempo determinado, con una recopilación centralizada para dar apoyo a la toma de decisiones.

En una bodega de datos se debe verificar que los datos estén limpios y entonces el sistema de dar una rápida gestión de gran capacidad de almacenamiento. Esta herramienta puede responder a cualquier consulta compleja relacionada con los datos (Acosta, 2019).

Minería de Datos

La minería de datos hace referencia al proceso de extraer conocimiento nuevo y útil de grandes volúmenes de información. La minería de datos se usa para resolver varios problemas de negocios como perfilamiento de mantenimiento de máquinas, modelamiento del comportamiento de equipos de trabajo, cálculo del historial de paras de línea, recomendación de puestos críticos entre otros (Chen et al., 2007).

Técnicas de Minería de Datos

Las técnicas de minería de datos se dividen en: técnicas de aprendizaje supervisado y técnicas de aprendizaje no supervisado.

Técnicas de aprendizaje supervisado: En las técnicas de aprendizaje supervisado se tiene variables de entrada y salida. Lo que se hace es ajustar las variables de entrada para predecir las variables de salida.

Técnicas de aprendizaje no supervisado: En las técnicas de aprendizaje no supervisado no tienen una variable de salida para predecir, solo se tiene variables de entrada. En vez de ajustar el modelo a las variables de entrada para predecir la variable de salida, estas técnicas buscan descubrir patrones dentro de los volúmenes de información (Gorakala & Usuelli, 2015).

Machine Learning

Esta es una disciplina de la inteligencia de negocios cuyo término en español significa aprendizaje de máquina, se usa para referir a los algoritmos y técnicas que aprenden automáticamente a identificar patrones en los datos para predecir comportamientos futuros. Las categorías principales de *machine learning* son: clasificación, descubrimiento de relaciones, regresión, *clusterización*, aprendizaje reforzado (Gallagher et al., 2015).

Deep Learning

Es un subconjunto de *machine learning* que es usado para describir algoritmos y técnicas capaces de modelar conceptos abstractos de alto nivel utilizado principalmente en áreas de reconocimiento de imágenes y procesamiento de texto. Un ejemplo de estas técnicas son las redes neuronales las cuales son capaces de reconocer dígitos escritos a mano (Dean, 2014).

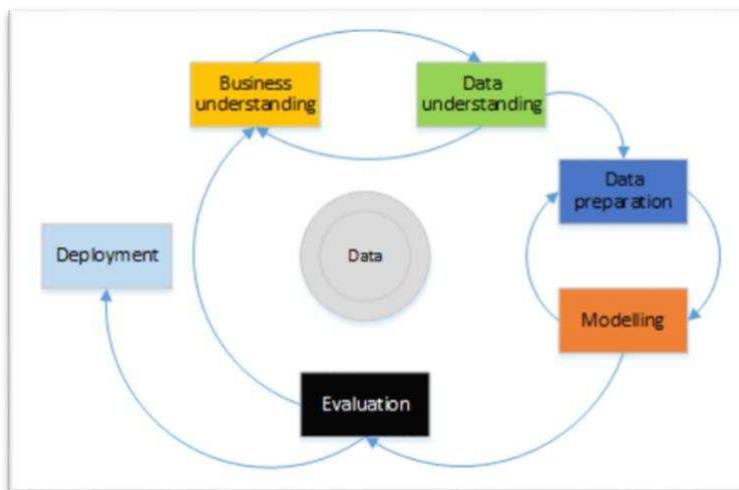
Metodología CRISP-DM

Esta metodología se estructura el ciclo de vida de un proyecto minería de datos mediante seis fases que interactúa entre sí de forma iterativa durante el desarrollo del proyecto ver la *Figura 2*. La metodología se organiza en fases, donde cada fase se estructura en varias tareas genéricas de segundo nivel, y estos hacen una proyección

hacia tareas específicas que describía las acciones que deberían desarrollarse para completar el proyecto (Palacios et al., 2017).

Figura 2.

Ciclo metodología CRISP-DM



Nota. Proceso del ciclo de las fases de la metodología CRISP-DM. Tomado de (Palacios et al., 2017).

Fundamentación de la variable dependiente

Industria 4.0

El término industria 4.0 se refiere a un nuevo modelo de organización y de control de la cadena de valor, a través del ciclo de vida del producto y a lo largo de los sistemas de fabricación apoyado por las tecnologías de la información.

El término industria 4.0 se utiliza de manera generalizada en Europa, si bien se acuñó en Alemania. También es habitual referirse a este concepto con términos como "Fábrica Inteligente" o "Internet industrial". En definitiva, se trata de la aplicación a la industria del modelo "Internet de las cosas" (IOT). Todos estos términos tienen en común el reconocimiento de que los procesos de fabricación se encuentran en un proceso de

transformación digital, una "revolución industrial" producida por el avance de las tecnologías de la información y, particularmente, de la informática y el software (Jose Luis Val Rolvan, 2015).

Sistema ANDOM Lean Manufacturing

Andon es una expresión de origen japonés que significa "lámpara" y que se relaciona con el control visual. A su vez es considerado como un elemento de la filosofía Lean Manufacturing, el cual agrupa un conjunto de medidas prácticas de comunicación utilizadas con el propósito de plasmar, de forma evidente y sencilla, el estado de algún sistema productivo.

El control visual como técnica de comunicación tiene múltiples aplicaciones, quizá las más importantes se relacionan con la identificación de anomalías y despilfarros; y sus principales propósitos consisten en facilitar tanto la toma de decisiones, como la participación del personal, proporcionando al mismo, información acerca de cómo su desempeño influye en los resultados, logrando así que pueda tener un mayor control sobre sus metas. Puede afirmarse entonces que el control visual empodera y motiva al personal a través de la información (Brayan Salazar Lopez, 2016).

Empresas de Manufactura

Según datos del INEC publicados en 2016 por su revista Evolución del Sector Manufacturero Ecuatoriano 2010-2013, la industria manufacturera ha presentado un crecimiento promedio inferior al crecimiento total de la economía en el periodo 2007-2014. Esto ha hecho que pierda participación en el PIB del país. No obstante, esta reducción se explica en gran medida por la caída de la producción de refinados de petróleo como parte de un plan de repotenciación de la principal refinería del país. Sin contar con refinación,

la participación de la manufactura en el PIB del país se ha mantenido estable alrededor del 12%.

El perfil manufacturero del país está mayormente basado en actividades intensivas en mano de obra y en recursos naturales, mientras que aquellas actividades intensivas en ingeniería presentan aún una participación reducida. Asimismo, la manufactura ecuatoriana está lejos de la frontera productiva internacional al contrastar su productividad con la de EE. UU. La información del año 2013 confirma la existencia de 76.803 empresas en el país, de las cuales la industria manufacturera representa el 9,7%. La industria manufacturera ha sido la de menor expansión, al situarse por debajo del total de la economía, con un crecimiento neto promedio anual de 9,9%. No obstante, dicha tasa de crecimiento sigue siendo significativa en términos de una comparación histórica e internacional (Natalia Garzón, 2016).

Gerencia de Transformación

El área de transformación de la empresa de manufactura está a cargo de asegurar que los departamentos de una organización van a alcanzar las metas de producción dentro del presupuesto y con un alto nivel de calidad. Los directivos utilizan los recursos dentro de la planta de fabricación para cumplir los objetivos para la producción de la empresa. El gerente cuenta con amplia experiencia empresarial en un entorno de fabricación (Kelchner Luanne, 2018).

Indicadores planes de negocios

La empresa de manufactura ha realizado un gran esfuerzo económico para consolidar sus indicadores productivos de tal manera que les permita tener información en tiempo real para la toma de decisiones, sean estas para realizar inversiones,

financiación en base a la evaluación de resultados que les permita realizar un análisis de su situación actual y proyectarse a futuro.

Los indicadores manufactureros utilizados son vistos desde una perspectiva financiero, procesos, calidad, Gente, etc. De estos indicadores para la investigación se usaron solo los indicadores pertenecientes a la parte manufacturera, de los cuales algunos sirven para determinar la efectividad de su producción.

Indicadores manufactura

Los indicadores de manufactura permiten medir la producción de la empresa ya que proveen de datos de personas por producto tanto a nivel de refrigeración como cocción, los indicadores utilizados son Costo del Producto, Gasto del Área, Calidad de los Productos, Mejorar el LEAD TIME y flexibilidad de los procesos, cumplir con cronogramas del área. Estos indicadores permiten realizar comparaciones entre los presupuestos reales y los presupuestos proyectados para determinar la situación actual de la empresa y tomar acciones correctivas a tiempo a nivel diario. Otros indicadores son la comparación entre los gastos reales y el gasto presupuestado a fin de revisar si las áreas de la organización están llevando sus gastos según lo planificado y poder tomar medidas correctivas en caso de que el gasto supere lo presupuestado.

El nivel de granularidad de los indicadores de manufactura de la empresa se lo puede reducir al punto que permite determinar la efectividad de cuantas unidades debe generar por persona, lo cual es útil ya que esto permite determinar si una determinada estrategia de producción tuvo o no la efectividad esperada.

Capítulo II. Revisión de Literatura

Antecedentes del estado del arte

Para el análisis del estado del arte se usaron las fases de criterios de inclusión y estrategia de búsqueda que son parte de un SMS¹, como fuentes de búsqueda de la información para la investigación se usaron los siguientes repositorios académicos Scopus, Springer, IEEExplore.

Definición del objetivo

El objetivo del estado del arte está enfocado en resolver las preguntas de investigación basadas en los objetivos específicos planteados anteriormente.

Definición de los criterios de inclusión y exclusión

Las búsquedas en las bases digitales dependiendo del tema retornan una gran cantidad de artículos relacionados, por lo cual es importante definir las características idóneas de los artículos a ser tomados en cuenta, para el presente análisis, se tomaron en cuenta los siguientes criterios.

Criterios de Inclusión

Con el fin de analizar metodologías utilizadas en la actualidad, fueron tomados en cuenta artículos a partir del 2008.

¹ Systematic Mapping Study (SMS): El SMS permite realizar un análisis de la literatura existente sobre un determinado tema a fin de identificar el estado del arte de un tema determinado.

Se tomaron en cuenta solamente artículos científicos y documentos de conferencias publicados en el idioma inglés.

Que el artículo contenga información referente al uso de metodologías de inteligencia de negocios para el desarrollo de modelos de análisis de improductividad en una línea de ensamble para empresas de manufactura.

En su mayoría se tomaron en cuenta artículos científicos y documentos de conferencias.

Criterios de Exclusión

Artículos que tengan temas de inteligencia de negocios no relacionados con analíticos o análisis predictivo.

Artículos que no estén en el idioma inglés.

Definición de la estrategia de búsqueda

Revisión Inicial

Se realiza una búsqueda inicial en los distintos repositorios académicos para buscar estudios relacionados con las preguntas de investigación.

Validación cruzada de estudios

En esta fase se procede a verificar que los estudios cumplan con los criterios de inclusión y exclusión, con lo cual finalmente se obtienen el listado inicial de documentos académicos con los cuales se va a trabajar en las siguientes fases del estudio.

Integración del Grupo de Control

El grupo de control está conformado por los estudios que cumplen con los criterios de inclusión y exclusión para lo cual se procede a realizar un análisis inicial del título de los estudios, introducción, conclusiones y palabras claves. Los estudios seleccionados para el grupo de control se muestran en la *Tabla 1*.

Tabla 1*Construcción de la cadena de búsqueda*

Grupo	Título	Palabras Clave
Control		
EC1	Applying Data Mining Techniques to Address Critical Process Optimization Needs in Advanced Manufacturing	Advanced Manufacturing, Big Data, Datamining Platform, Process Optimization
EC2	Data Mining and Analytics in the Process Industry: The Role of Machine Learning	Data mining, data analytics, machine learning, process industry,
EC3	Adaptive KNN based Recommender System through Mining of User Preferences	Recommender system _ Personalization, Adaptive K-nearest neighbor, Ontology, Web mining, Clustering
EC4	Data mining in manufacturing: a review based on the kind of knowledge	Knowledge discovery, Data mining, Manufacturing, Text mining, Literature review
EC5	Service innovation and smart analytics for Industry 4.0 and big data Environment	Manufacturing, servitization, predictive maintenance, industrial big data
EC6	Towards Industry 4.0 Utilizing Data-Mining Techniques: A Case Study on Quality Improvement	Industry 4.0, Datamining, Intelligent, Manufacturing

Nota. Tabla de resume los grupos de control establecido, así como las palabras claves de cada uno de ellos.

La cadena de búsqueda está formada por la unión de las palabras claves que más se repiten en cada contexto, los conectores usados son OR para las palabras que están

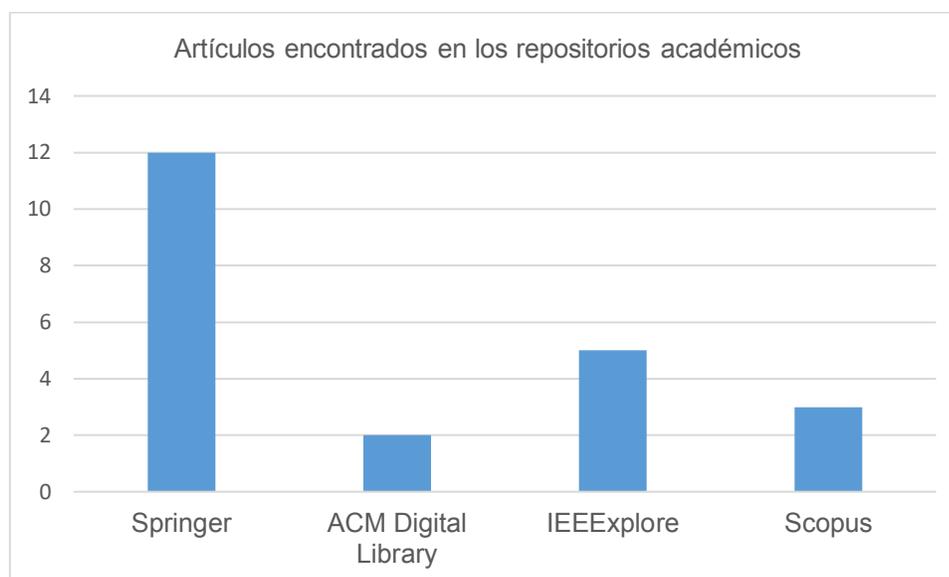
dentro del mismo contexto y el conector AND para las palabras que están en contextos distintos, de esta manera se establece la siguiente cadena de búsqueda.

((manufacturing) OR (Process Optimization)) AND (data mining) AND (Big Data) AND (Industria 4.0)

Al aplicar la cadena de búsqueda en los repositorios académicos seleccionados para el estudio se procede a filtrar solo los documentos que estén en el idioma inglés, cuya fecha de publicación sea mayor al 2008 y que sean artículos científicos o documentos de conferencias obteniendo como se observa en la Figura 3.

Figura 3.

Artículos encontrados en los repositorios académicos.



Nota. El gráfico indica el número de artículos revisados en los repositorios académicos de donde se seleccionó luego los artículos primarios.

Luego de filtrar los artículos en los repositorios académicos se obtuvo un total de 22 artículos repartidos como lo indica la Figura 3 y detallados en el Anexo A. Finalmente, se realizó una revisión a nivel de resumen para determinar los artículos primarios,

determinando que serían tres. Una vez obtenidos los resultados primarios se realizó la revisión de los documentos encontrados, los cuales se detallan a continuación:

(Wei Ji, Lihue Wang,2017) Big data analytics-based fault prediction for shop floor scheduling

En este documento los autores identifican como método común de recomendación el enfoque de predicción de fallas, basado en el análisis de datos para la programación de una planta de manufactura, de acuerdo con los datos obtenidos en la línea de producción se encuentran posibles patrones de falla ya sea por errores de máquina, mantenimiento o problemas de programación. Con los patrones extraídos se calculan sus similitudes o diferencias, sobre esta información se puede extraer probabilidades de falla en las tareas programadas o las tareas de mecanizadas actuales y proporciona una referencia para la toma de decisiones para programar o reprogramar tareas evitando posibles errores por esto propone análisis por medio de técnicas análisis de datos.

(Usmani et al., 2017) The production of prediction: What does machine learning want?

Los autores proponen realizar un mapeo de las principales formas de acción material, narrativa y problematización que se ejecutan a través de técnicas de modelado de datos como algoritmos de Regresión, árboles de decisión y clasificadores Naive Bayes con esto se pretende que el enfoque de los datos de minería de datos sea usado para solventar problemas económicos y de conocimiento de varios ámbitos, mediante esto podrían ayudarnos a preguntar el cómo de varios entornos.

(A. K. Choudhary · J. A. Harding · M. K. Tiwari, 2008) Data mining in manufacturing: a review based on the kind of knowledge

Los autores proponen revisar la literatura que trata sobre el descubrimiento del conocimiento y aplicaciones de minería de datos en el amplio dominio del ámbito de la

fabricación. Revisando las principales funciones de la minería de datos como técnicas de caracterización y descripción, asociación, clasificación, predicción, agrupación y análisis en su evolución. Esta revisión revela las aplicaciones progresivas y las brechas existentes identificadas en el contexto de la minería de datos en fabricación. Además de los vacíos de investigación y encontrar los vínculos entre el conocimiento y las herramientas de minería de datos aplicadas y sus técnicas en la actualidad.

Conclusión

Al realizar la revisión de literatura se identifica que la manera como las empresas de manufactura mejoran su producción es por medio de sistemas de análisis predictivos. Es común encontrar dentro de estos sistemas de análisis predictivos algoritmos de clusterización, adicionando varias técnicas de minería de datos, manejos estadísticos que permiten segmentar la información de las líneas de producción, tomando en cuenta las características explícitas e implícitas de las fábricas, así como el contexto de los recursos humanos. Es por este motivo que para solventar el problema de las empresas de producción en el Ecuador se propone la implementación de un algoritmo de inteligencia de negocios que analice la información histórica de las líneas de producción para realizar recomendaciones de posibles parámetros de planta tomando en cuenta las características de los productos y usuarios con la finalidad de mejorar la efectividad de la fuerza de producción de la empresa.

Herramientas

Para el desarrollo del presente proyecto de titulación se requirió determinar las herramientas de software con las que se trabajarán, en primera instancia se definió realizar los modelos relacionales y multidimensional utilizando la herramienta Power Designer.

Power designer

Herramienta que permite analizar, manipular y sobre todo visualizar la arquitectura de metadatos de un negocio, empresa o simplemente bases de datos convencionales. Esta herramienta permite el diseño y construcción de modelo de datos a nivel físico y conceptual para su posterior análisis buscando un alto rendimiento. Es considerada una herramienta líder en modelamiento empresarial por características como: interfaz gráfica altamente personalizable, alinea el negocio con el área de tecnología, sistema de documentación, reduce impacto ante cambios, entre otras que la convierten en una herramienta destacada y de fácil uso (Ecured, 2019).

Oracle

Se trata de una herramienta cliente/servidor para la gestión de bases de datos. Es un producto vendido a nivel mundial, aunque debido a la gran potencia que tiene y su elevado precio, hace que mayormente sea utilizado por empresas muy grandes y multinacionales.

Oracle con su tecnología cliente/servidor, requiere primero la instalación de la herramienta servidor y posteriormente instalar la base de datos en otros equipos con herramientas de desarrollo como Oracle Designer y Oracle Developer, que son las herramientas básicas de programación sobre Oracle.

Para desarrollar en Oracle se utiliza PL/SQL, bastante potente para tratar y gestionar la base de datos, también por norma general se suele utilizar SQL al crear un formulario.

El Developer es una herramienta que permite crear formularios en local, es decir, mediante esta herramienta se puede crear formularios, compilarlos y ejecutarlos. La principal ventaja de esta herramienta es que es bastante intuitiva y dispone de un modo

que permite componer el formulario, tal y como lo haríamos en Visual Basic o en Visual C.

Designer que es una herramienta que se conecta a la base de datos y por tanto crea formularios en ella, de esta manera todo el mundo se conecta mediante Designer a la aplicación que contiene todos los formularios y no hay problemas de diferentes versiones, esto es muy útil y perfecto facilitar el trabajo con otros. Pero el principal y más notable problema es la falta de un entorno visual para diseñar el formulario, es decir, posee una estructura como de árbol en la cual insertamos un formulario, a la vez dentro de éste insertamos bloques o módulos que son las estructuras que contendrán los elementos del formularios, que pueden estar basados en tablas o no (Masip, 2002).

Tableau

Se trata de una herramienta de software con una solución completa de *Business Intelligent*. Durante los últimos años esta herramienta ha sido líder en el mercado, por la capacidad que da al usuario de interactuar con los datos de una manera intuitiva, ya que con pocos clics se puede generar potentes visualizaciones o crear Dashboards para realizar control de KPI y así extraer información importante de una organización (Gil, 2020).

Tableau con su logo visto en la **Error! Reference source not found.**, es una herramienta excelente para la analítica de datos, puesto sus capas de visualización y presentación tienen una muy alta clasificación, por su facilidad de uso, simplificando el trabajo para presentar los datos de una manera comprensible e intuitiva.

Entre sus características más principales se puede destacar la facilidad para integrar datos desde diferentes orígenes, como software líder en plataformas de Business Intelligence, Tableau permite integrar toda la información de la empresa en un único

modelo, de forma que se puede analizar al mismo tiempo datos de sistemas ERP como JDE o SAP, base de datos Oracle, SQL Server, hojas de cálculo, etc. (Neteris, 2019).

Además, la facilidad con la que funciona Tableau permite al usuario manipular fácilmente lo que está pasando en la interfaz, permitiendo realizar gráficas al instante. Otras de las características que denota esta herramienta es el control de versiones, la ejecución de consultas sin necesitar código, diferentes precios, independencia del departamento de TI, adaptable a diferentes situaciones y entornos, rapidez, tiene una comunidad enorme y su gran popularidad ante otras herramientas (Bordera, 2018). Presenta a sus usuarios los productos disponibles mostrados en la Figura 4.

Figura 4.

Productos disponibles Tableau.



Nota. Soluciones de software que presenta Tableau para los usuarios. Tomado de (Neteris, 2019).

Tableau Desktop: Permite realizar potentes visualizaciones desde nuestro equipo en pocos segundos.

Tableau Server: Permite compartir con otros usuarios la información generada en la versión de escritorio. Es perfecto para trabajar en equipos y a distancia.

Tableau Online: Se trata de una plataforma de análisis alojada en la nube. No se necesita instalar el programa para poder trabajar con el software, ya que toda la plataforma de análisis se encuentra hospedada en la nube (Neteris, 2019).

Rapidminer

Considerada como una de las principales herramientas de desarrollo de modelos minería de datos, disponible en edición gratuita, así como en edición comercial. Además, posee la extensión Rapidminer Server diseñada para implementar y mantener modelos facilitando la colaboración empresarial. El portafolio de soporte de Rapidminer también incluye Rapidminer Real-Time Scoring y Rapidminer Radoop. También presenta características aumentadas como Rapidminer Turbo Prep, Rapidminer Auto Model y Rapidminer Automated Model Ops, mientras que Rapidminer AI Cloud ofrece opciones de implementación basadas en la nube (Gartner, 2020).

Rapidminer con la comunidad de usuarios masivos, está activo y reforzada por los esfuerzos de habilitación del proveedor, incluidas las certificaciones de habilidades. Rapidminer ofrece soporte empresarial las 24 horas, los 7 días de la semana para clientes que pagan con SLA según la gravedad del problema.

Actualmente, esta herramienta se posiciona como un visionario en el cuadrante mágico de Gartner, pasando del cuadrante de líderes del año pasado debido principalmente al lento crecimiento con respecto a otros proveedores en este cuadrante mágico. Mantiene un fuerte enfoque en la innovación y la excelencia de los productos, lo que servirá en la lucha por un espacio más concurrido y competitivo en el mercado. La nueva misión de Rapidminer y la nueva estrategia de lanzamiento al mercado está dirigida a clientes empresariales que pueden ayudar a acelerar el crecimiento (Gartner, 2020).

Técnicas de minería de datos

Las técnicas de minería de datos deben ser escogido basándose en el tipo de datos a analizar, el tipo de conocimiento o patrones a extraer de los datos y la forma en la que se utilizará el conocimiento. Además, para pasar los datos por un proceso de minería de datos es necesario que éstos hayan sido preparados. Considerando el problema que se desea solucionar se selecciona técnica más adecuada.

Los problemas planteados, generalmente tienen dos tipos de enfoque: problemas descriptivos y problemas predictivos.

Técnicas de data mining para problemas descriptivos

El objetivo de estas técnicas es resolver un problema mediante la descripción de los datos. Por ejemplo se tienen los siguientes:

Análisis de Segmentación: Mediante esta técnica se busca agrupar datos similares y homogéneos, para dar solución de estos problemas se construye modelos de aprendizaje no supervisado o clustering.

Análisis de Asociación: El objetivo de esta técnica es analizar datos desconocidos para encontrar relaciones que tenga sentido entre ellos.

Técnicas minería de datos para problemas predictivos

Son aquellas técnicas usadas para aplicar a datos futuros, es decir, predicen comportamientos de patrones a partir de los datos, se las conocen también como aprendizaje supervisado y las variables utilizadas pueden ser categóricas y numéricas. Entre algunas de éstas se tiene:

Clasificación: Esta técnica crea modelos en los que la variable a predecir tiene valores definidos y son contables, mayormente se usan variables categóricas.

Predicción de Valores: Intenta predecir valores, por lo que más ocupa son variables numéricas. Con frecuencia lo que se obtiene es la probabilidad de un evento.

Clasificación según el enfoque

De acuerdo con su función las técnicas se pueden clasificar en:

Clasificadores: Clasifican datos en clases predefinidas.

Algoritmos de regresión: A partir de los datos generan una función predictiva.

Descubrimiento de reglas de asociación: Búsqueda de relaciones entre variables.

Modelado de dependencias: Generación de modelos que expliquen las dependencias entre atributos.

Agrupamiento: Creación de grupos cuando las clases son desconocidas.

Aprendizajes basados en casos: Se establecen en indexar y recordar los casos más significativos, de forma que los nuevos casos son clasificados según el descriptor más próximo.

Compactación: Búsqueda de descripciones más compactas de los datos.

Detección de desviaciones: Basado en la búsqueda de desviaciones más importantes de los datos respecto a valores anteriores (Romero, 2019).

Algoritmos de minería de datos

Para resolver cualquier problema de minería de datos se requiere utilizar la técnica adecuada de acuerdo a los datos analizados, y en base a esa técnica se escoge el respectivo algoritmo entre los cuales se tienen los siguiente.

Árboles de decisión

Algoritmo que da como respuesta una secuencia de decisiones organizadas de manera jerárquica, a manera de ramas de un árbol. Para utilizar estos algoritmos los

datos de entrada de pueden ser numéricos o categóricos. Pueden ser aplicados para técnicas de clasificación de categorías (árboles de clasificación), clasificación de datos numéricos (árboles de regresión), agrupación y pronóstico.

Redes neuronales artificiales

Son algoritmos muy potentes, y ayudan a modelar prácticamente con cualquier tipo de problemas. Son utilizados para técnicas de clasificación, predicción y agrupación. Los datos que utilizan este algoritmo tienen que ser numéricos caso contrario es necesario discretizar los mismos.

Estadísticos descriptivos

Genera reglas que demuestran cómo se asocian los diferentes atributos ingresados mediante la data (Romero, 2019).

El algoritmo K-means

Basado en el análisis de grupos, puesto que trata de dividir los datos recogidos en "clusters" agrupados por características similares.

Máquinas de vectores de soporte

Los algoritmos de máquinas de vectores de soporte toman datos de entrada y predicen cuál de las dos posibles categorías incluyen los datos de entrada.

Apriori

Este algoritmo es comúnmente utilizado para controlar los datos de transacciones, por ejemplo, para una tienda de ropa mediante este algoritmo se podría controlar qué camisas suelen comprarse juntas los clientes.

Naive Bayes

El algoritmo Naive Bayes predice la salida de una identidad basándose en los datos de observaciones conocidas (Greg, 2018).

Capítulo III. Metodología

Metodología de investigación

La investigación científica es aquel proceso investigativo que mediante la aplicación del método científico busca encontrar información confiable, relevante, veraz y fidedigna con la utilidad de entender, verificar, corregir o aplicar el conocimiento. Para alcanzar el objetivo de esta investigación, se utiliza una metodología, comprendida por un conjunto de aspectos operativos necesarios para la realización de un estudio. Es así como esta se ocupa de problemas reales detectados desde la práctica (Cabezas et al., 2018).

Esta metodología busca obtener contribuciones científicas o técnicas genéricas, pero de utilidad borrosa o no declarada y aquellas orientadas a la relevancia que son útiles, pero no suficientemente formales, transparentes o validadas con criterio científico (González & Pomares Quimbaya, 2012).

Pero cuando se tiene problemas que requieren someterse a recurrentes pruebas, validaciones, mejoras y ajustes para poder ser publicados como potenciales soluciones, como es el caso de proyectos de minería de datos, una metodología de la investigación científica común, no aporta significativamente a la consecución del objetivo de la misma, puesto que no orienta la misma a una investigación cuantitativa, orientada a la generalización de resultados. Es así como se requiere de una investigación cuyo énfasis es la solución de problemas y la construcción de conocimiento para futuras situaciones similares (Sánchez et al., s. f.).

Es así como autores como (De Benito Crosetti & Salinas Ibáñez, 2016) han citado a la investigación basada en diseño como un tipo de investigación orientado hacia las nuevas tendencias educativas, basadas en la innovación, que buscan introducir un

elemento nuevo para transformar una situación. Por otro lado (Sánchez, Nahón, y Ramírez s. f.) indican que este tipo de investigación diseña procesos, estrategias o cualquier otro producto necesario para analizar, diseñar, desarrollar e implementar soluciones dando aportes significativos en un área del conocimiento.

Se plantea utilizar la investigación científica basada en el diseño que busca hacer aportes significativos en un área del conocimiento y a la vez dar solución a problemas relevantes, utilizando el análisis de problemas aún no resueltos en un ambiente del mundo real y su resolución de una manera novedosa y rigurosa a través del diseño de artefactos (Winter, Zhao, Aierσ, & DESRIST, 2012).

Dentro de algunos estudios realizados en busca de la mejor metodología orientada al diseño se puede visualizar las fases que lo componen en la tabla 2 .

Tabla 2

Tabla comparativa de metodologías del basadas en el diseño

Fases	(Easterda y et al., 2018)	(De Benito & Crosetti Salinas Ibáñez, 2016)	(Pagano & Buitron, 2009)	(Sánchez et al., s. f.)	(Gonzalez & Pomares Quimbaya, 2012)
1	Enfocar	Análisis de la situación. Definición del problema.	Preparación del diseño.	Análisis de la situación. Definición del problema.	Entorno (Personas, tecnología y organización)
2	Entender	Desarrollo de soluciones de acuerdo con una fundamentación teórica.	Implementación del diseño.	Desarrollo de soluciones de acuerdo con una fundamentación teórica.	Relevancia (Requerimientos o aplicación)

Fases	(Easterday et al., 2018)	(De Benito & Crosetti Salinas Ibáñez, 2016)	(Pagano & Buitron, 2009)	(Sánchez et al., s. f.)	(Gonzalez & Pomares Quimbaya, 2012)
3	Definir	Implementación	Análisis retrospectivo	Implementación	Diseño / Investigación (construir-diseño-evaluar) proceso interno recurrente
4	Concebir	Validación		Validación	Rigor (Contribución o conocimiento aplicable)
5	Construir	Producción de documentación y principios de diseño		Publicación de resultados que propongan principios para futuras investigaciones.	Base de conocimiento (Teorías, modelos, métodos, artefactos, experiencia)
6	Testear				
7	Presentar				

Nota. Tabla que compara las fases citadas por algunos autores para el uso de la metodología basada en el diseño.

Las metodologías propuestas por (Easterday, Rees Lewis, y Gerber 2018) y (De Benito Crosetti y Salinas Ibáñez 2016) proponen que cada fase pueda volver a cualquiera de sus fases anteriores, mientras que (Pagano y Buitron 2009) muestran un proceso recursivo de sus fases. Las fases según (Sánchez et al. s. f.) se someten a un proceso recursivo donde siempre se validan sus fases. Por otro lado (González y Pomares Quimbaya 2012) proponen un modelo en donde si un ciclo se mueve los otros dos se mueven a la vez, lo que permite que el proceso esté en continuo descubrimiento del conocimiento y a la vez se esté validando y comprobando el entorno.

La investigación de este trabajo se elaboró y ejecutó utilizando la metodología de investigación científica basada en el diseño de González y Pomares Quimbaya, que está enfocada al desarrollo del proyecto planteado, puesto recurrentemente se necesita estar evaluando el entorno del negocio, por otro lado construyendo y evaluando el modelo creado y por último publicando el conocimiento creado. Además, esta metodología también permite volver a las fases anteriores acoplándola más al proyecto.

Esta investigación se distingue de la investigación aplicada por el énfasis en la construcción de artefactos innovadores y la retroalimentación que proveen para revisar, extender o recontextualizar los fundamentos teóricos.

(González y Pomares Quimbaya 2012) han propuesto una arquitectura general para proyectos de investigación centrados en diseño, mediante la articulación de tres ciclos: rigor, relevancia y diseño, cómo se puede observar en la *Figura 5*.

Figura 5.
Investigación mediante ciclos



Nota. La figura representa el ciclo de la metodología orientada al diseño propuesta por los autores Gonzalez & Pomares Quimbaya, 2012. Tomado de (Gonzalez & Pomares Quimbaya, 2012).

Los tres ciclos consisten en primero integrar el entorno, lugar donde se dará solución al problema utilizando los requerimientos de personas, tecnología y organización para que mediante la construcción y evaluación de un diseño se contribuya a la base del conocimiento con teorías, modelos, métodos, experiencia y artefactos existentes.

Se debe tomar en cuenta que este es un proceso cíclico en cada una de sus fases y si un ciclo se está moviendo los otros dos lo harán a la vez, no como sucede en las fases de investigación (Gonzalez & Pomares Quimbaya, 2012).

Algunos autores han tomado como referencia la metodología propuesta por Gonzalez y Pomares para el desarrollo de proyectos de investigación. Por ejemplo Ramirez J. se basó en esta metodología para desarrollar un sistema experto basado en ontología para la detección de fraude en tarjetas de crédito (Ramirez, 2019). Otro ejemplo del uso de la metodología lo expuso (Mendoza Mendoza et al., 2019) en su desarrollo de un portal web para la consulta y análisis de datos abiertos relacionados con la

contratación pública : ADACOP, para la explotación sobre los datos abiertos del gobierno colombiano. Bello y Leonardo por su parte también utilizaron esta metodología en el trabajo sobre la clasificación de historias provenientes de familiares de víctimas y sobrevivientes del conflicto sociopolítico de Colombia (Bello & Leonardo, 2016). Además, (Fontaines et al., 2020) dentro de su libro titulado “Tendencias en investigación” también resalta esta investigación como metodología para proyectos de ingeniería.

Los ciclos de esta metodología están enmarcados al desarrollo del proyecto planteado en tres fases.

Entorno

En esta fase se realizó un análisis de los requerimientos del proyecto, es decir, se recopiló datos existente en la base de datos y necesidades que los actores principales de una empresa de manufactura, requieren obtener como resultado del proyecto. A continuación, se determinó las posibles herramientas ETL utilizadas, para realizar la limpieza y transformación de los datos recopilados, seleccionando la más factible para el diseño y creación de una bodega de datos. Además, se identificó los algoritmos de minería de datos más adecuados para el modelo diseñado, utilizando una herramienta de minería de datos, de acuerdo con lo descrito en los objetivos específicos OE1 y OE2.

Además, se utilizaron los tipos de investigación experimental y descriptiva. La experimental relacionó las variables de estudio con la que se determinó las herramientas y modelos necesarios para hacer la minería de datos. Y la descriptiva recopiló datos relevantes de estudios posteriores sobre el diseño de minería de datos.

Obtenidos los requerimientos para el desarrollo del proyecto es posible pasar a la segunda fase.

Diseño

Fase en la que se crea el modelo de minería de datos mediante los requerimientos definidos en la fase de entorno, para determinar los patrones que afectan el desempeño de las líneas de producción. Finalmente, se evaluó el modelo a través del uso de técnicas de validación implementadas en minería de datos y así determinar el nivel de confianza del resultado predictivo. Como lo describe los objetivos OE3,OE4 y OE5.

Con el diseño realizado y evaluado se realizó la contribución de acuerdo con la metodología y así pasar a la última fase.

Bases de conocimiento

En esta fase se observó los resultados obtenidos en la investigación, como es un modelo de minería de datos de las parámetros de línea de la empresa manufacturera en el Área de transformación, donde se identificó los patrones de comportamiento de los productos que influyen en la productividad y calidad, de acuerdo con lo planteado en el objetivo general.

Para la visualización de los resultados, bases del conocimiento se realizaron cuadros de mando para poder presentar el modelo creado en este proyecto.

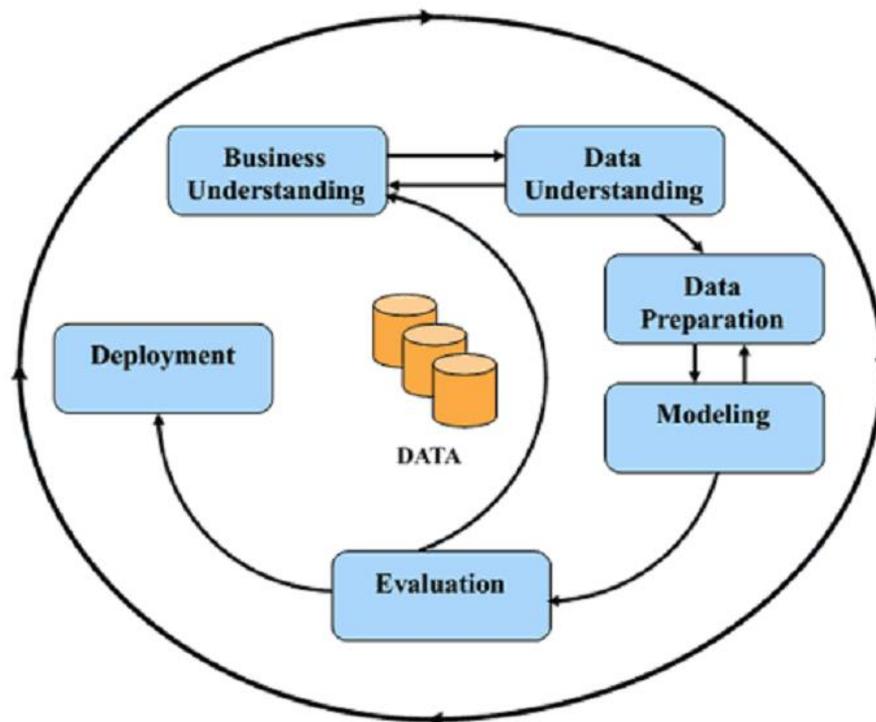
Metodología de minería de datos

CRISP-DM es una metodología robusta utilizada en el proceso de minería de datos. El enfoque de esta metodología es estructurado para la planificación de un proyecto de minería de datos.

El esquema de este modelo se basa en la secuencia de eventos con algunas tareas que pueden realizarse en un orden diferente y, a menudo, será necesario retroceder a tareas anteriores y repetir ciertas acciones como se lo observa en la Figura 6.

Figura 6.

Flujo de la metodología CRISP-DM



Nota: La figura detalla el flujo de las fases de la metodología de minería de datos CRISP-DM. Tomado de (Manasson, 2020).

Entendimiento del negocio

En la primera fase de esta metodología se encuentra el entendimiento comercial. Es común observar que para iniciar un proyectos no se definen primero el problema y los objetivos adecuados, considerando que no se habla específicamente del análisis de datos, qué es uno de los errores que siempre se comete. Es así como la mayoría de metodologías incluida CRISP-DM, así énfasis en colocar primero la definición de problemas o la comprensión empresarial.

Basados en este proyecto la primera fase es comprender, cuál es el proceso que sigue la empresa para obtener la data a analizar, y así comprender el propósito de análisis y proporcionar una definición clara del problema (Manasson, 2020).

Es aquí donde se requiere acercarse a los encargados de la empresa y a los empleados que proceden a cargar la data en el sistema, para comprender cuál es el propósito de esta y determinar cómo afecta está en la actividad del negocio.

Comprensión de datos

La siguiente etapa del proceso CRISP-DM es la comprensión de los datos. Normalmente implica los siguientes pasos: determinar qué datos se necesitan, recopilar datos, si los datos requeridos no están disponibles, explorar datos, verificar la calidad de los datos, entre otros.

En esta fase se recopilará toda la información que esté relacionada al caso de estudio, con el propósito de analizar las bases de datos involucradas, el tipo de columnas que estas tienen y observar las características de las mismas.

Es importante tener una vista previa de los datos para seleccionar sólo las características necesaria y de esta manera formatear los datos para cargar en un archivo. Una vez que se cargan los datos, se exploran con el objetivo de comprender sus dimensiones, tipos de datos y distribución de valores (Manasson, 2020).

Finalmente, es importante determinar cuál es la estructura que van a tener los datos, mediante la creación de un modelo relacional, para luego pasarlo a un modelo multidimensional.

Preparación de datos

Ahora que se conoce los datos con los que se está tratando, es necesario pasar a la fase de análisis y modelado. Debido a la mezcla de datos, es necesario pasar a la

fase más larga pero importante del proceso, como lo es el proceso de limpieza de los datos. Esta etapa del proceso CRISP-DM se puede dividir en los siguientes pasos:

Datos limpios: Generar nuevos atributos necesarios para el análisis (también llamado Ingeniería de características).

Integrar datos (fusionar y agregar): Aquí se decide qué hacer con los valores faltantes, los valores atípicos y otros datos inutilizables identificados en la etapa anterior (Manasson, 2020).

Modelado

Esta es la parte más interesante del proceso CRISP-DM donde realmente se encuentran las respuestas a las preguntas planteadas mediante la realización de modelos y análisis estadísticos (Manasson, 2020).

En esta fase se buscó determinar la predicción de los datos mediante un modelo de minería de datos que permitió visualizar los patrones de comportamiento de los datos proporcionados.

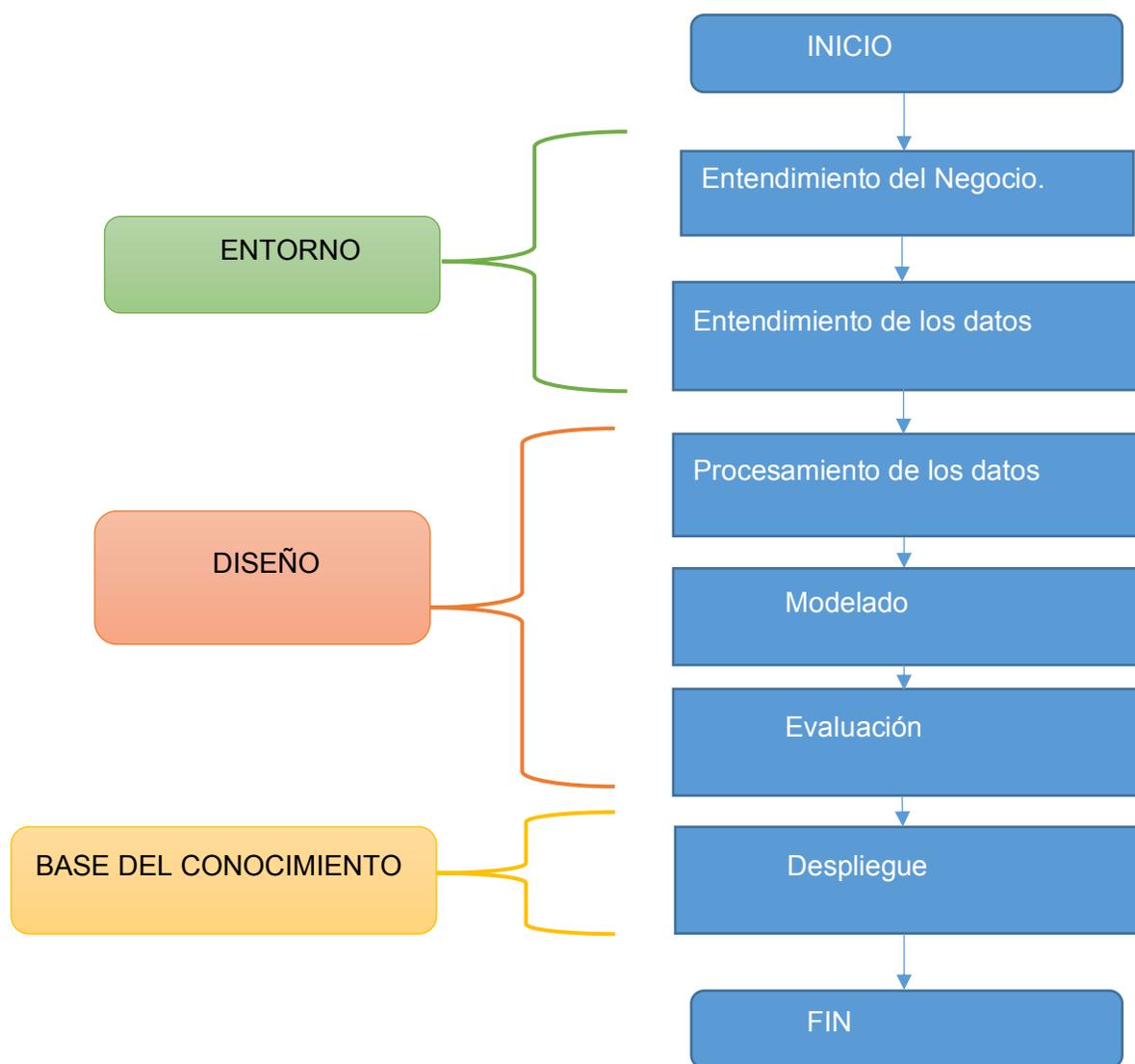
Despliegue

Con el análisis completo, se pasa a la etapa final del proceso CRISP-DM: implementación. En este caso de análisis exploratorio, la implementación es significativamente ya que se obtuvo un modelo predictivo que determina los patrones de para del proceso de producción de una empresa (Manasson, 2020).

Esquema general del proyecto

En base a las metodologías antes revisadas, y considerando la importancia de seguir un proceso organizado y estandarizado del manejo de los datos para así obtener un modelo eficaz y confiable se siguieron los siguientes pasos:

Figura 7.
Esquema general del proyecto



Nota. La figura representa el flujo seguido en el proyecto desarrollado en base a las metodologías propuestas.

Capítulo IV. Desarrollo del estudio

Recolección y análisis de la información

Descripción del negocio

La empresa cuyo giro de negocio es la manufactura de productos de línea blanca para el hogar, es una empresa con cuarenta y siete años en el mercado ecuatoriano y cuenta con nueve países para la comercialización a nivel nacional e internacional. La empresa se dedica a la producción y venta de productos electrodomésticos de refrigeración y de cocción. Por motivos de confidencialidad el nombre de la empresa se mantendrá reservado.

En la actualidad la empresa mantiene una línea de producción de refrigeración. Dentro de la línea de producción se tienen varias etapas para el ensamble de una refrigeradora, cada etapa tiene una serie de pulsantes en donde según el inconveniente que se tenga se presiona uno de estos. Cuando se realiza la pulsación se genera información a un sistema transaccional.

Estas paras en la línea generalmente ocurren por problemas logísticos, calidad o daños en la maquinaria. Por este motivo las empresas de manufactura cuentan con personal especializado para realizar revisiones de la línea de producción por medio de normativas de las metodologías *Lean Manufacturing*.

Los problemas con este tipo de revisiones es que son apaga incendios, son métodos correctivos que no solventan las paras de las líneas de producción; por ejemplo, la línea de producción tiene una meta de producir por cada minuto un producto de refrigeración, esta tarea se la cumple en cada una de las estaciones de trabajo que son cinco en la línea de refrigeración, cada uno de los operarios cuenta con un documento llamado hojas de trabajo estandarizado en donde se tienen los pasos a seguir para el

correcto cumplimiento de la actividad de la que es responsable. Cuando estos pasos no se realizan en su cabalidad es porque se tienen problemas con el proceso de producción o se terminó el material y no hay una generación de información por el área logística o no existe revisión temprana por parte de mantenimiento en el daño de un equipo.

Es así como esta empresa requiere de una herramienta de inteligencia de negocios, con un sistema que les permita generar información descriptiva, del estado de las paras registradas en cada una de las etapas de la línea de producción. Además, un modelo predictivo que les permita generar información de posibles paras de línea de acuerdo con los atributos del sistema de base de datos almacenado, para ofrecer la productividad que se tenía esperada.

Descripción de los datos

Para realizar este proyecto se procedió a obtener la base de transaccional donde se almacenan todos los reportes de paras ingresados por pulsadores en cada una de las etapas de producción de la línea de producción de la empresa.

Luego de solicitar esa información se obtuvo una base de datos como la mostrada en la Figura 8.

Figura 8.

Base de datos inicial

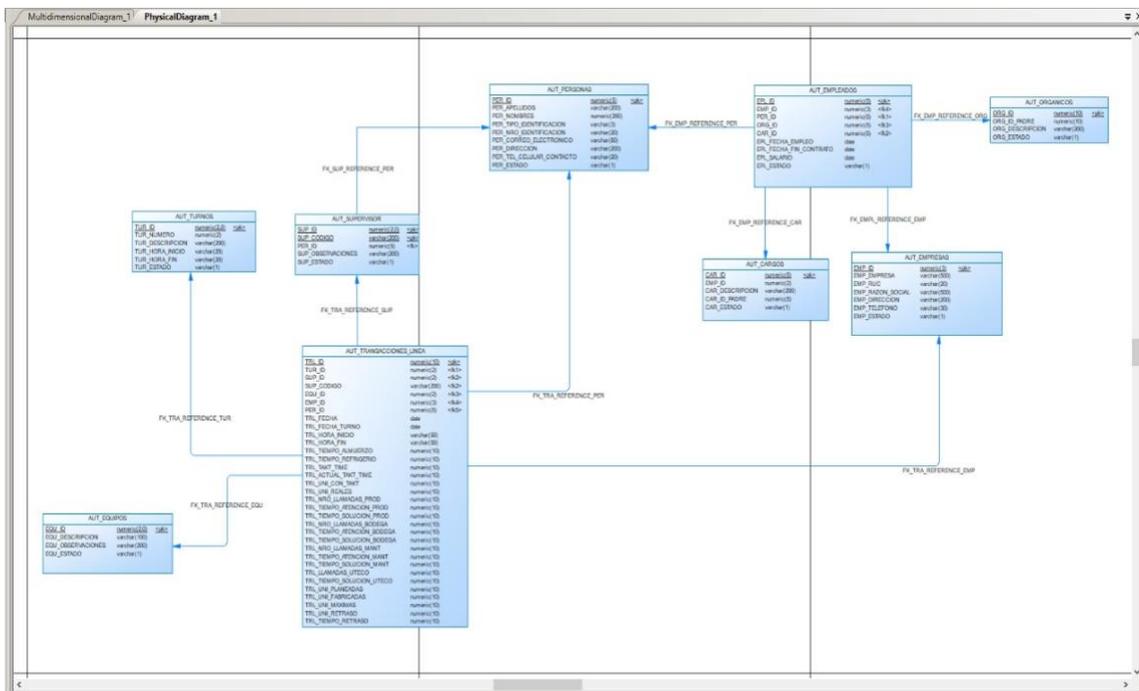
DW_E_ID	TSI_ID	DW_E_DATA_INDEX	DW_E_FECHA	DW_E_NCOLD	DW_E_NCOL1	DW_E_NCOL2	DW_E_NCOL3	DW_E_NCOL4	DW_E_NCOL5	DW_E_NCOL6	DW_E_NCOL7	DW_E_NCOL8	DW_E_NCOL9	DW_E_NCOL10
1	53259	6	857 09-10-2018 19:28:33 *	63	55	28	20	0	3	0	0	0	0	0
2	53260	6	858 09-10-2018 19:30:02 *	63	55	29	20	0	3	0	0	0	0	0
3	53261	6	859 09-10-2018 19:30:32 *	63	55	29	21	0	3	0	0	0	0	0
4	53262	6	860 09-10-2018 19:31:02 *	63	55	30	21	0	3	0	0	0	0	0
5	53263	6	861 09-10-2018 19:31:34 *	63	55	30	22	0	3	0	0	0	0	0
6	53264	6	862 09-10-2018 19:32:03 *	63	55	31	23	0	3	0	0	0	0	0
7	53265	6	863 09-10-2018 19:32:33 *	63	55	31	24	0	3	0	0	0	0	0
8	53266	6	864 09-10-2018 19:33:04 *	63	55	32	24	0	3	0	0	0	0	0
9	53267	6	865 09-10-2018 19:33:35 *	63	55	32	24	0	3	0	0	0	0	0
10	53268	6	866 09-10-2018 19:34:04 *	63	55	33	25	0	3	0	0	0	0	0
11	53269	6	867 09-10-2018 19:34:33 *	63	55	33	25	0	3	0	0	0	0	0
12	53270	6	868 09-10-2018 19:35:04 *	63	55	34	26	0	3	0	0	0	0	0
13	53271	6	869 09-10-2018 19:35:33 *	63	55	34	26	0	3	0	0	0	0	0
14	53272	6	870 09-10-2018 19:36:04 *	63	55	35	26	0	3	0	0	0	0	0
15	53273	6	871 09-10-2018 19:36:33 *	63	55	35	26	0	3	0	0	0	0	0
16	53274	6	872 09-10-2018 19:37:03 *	63	55	36	26	0	3	0	0	0	0	0
17	53275	6	873 09-10-2018 19:37:35 *	63	55	36	26	0	3	0	0	0	0	0
18	53276	6	874 09-10-2018 19:38:04 *	63	55	37	26	0	3	0	0	0	0	0
19	53277	6	875 09-10-2018 19:38:33 *	63	55	37	27	0	3	0	0	0	0	0
20	53278	6	876 09-10-2018 19:39:03 *	63	55	37	27	0	3	0	0	0	0	0
21	53279	6	877 09-10-2018 19:39:34 *	63	55	38	27	0	3	0	0	0	0	0
22	53280	6	878 09-10-2018 19:40:04 *	63	55	38	28	0	3	0	0	0	0	0
23	53281	6	879 09-10-2018 19:40:34 *	63	55	39	28	0	3	0	0	0	0	0
24	53282	6	880 09-10-2018 19:41:03 *	63	55	39	28	0	3	0	0	0	0	0
25	53283	6	881 09-10-2018 19:41:32 *	63	55	40	28	0	3	0	0	0	0	0
26	53284	6	882 09-10-2018 19:42:02 *	63	55	40	28	0	3	0	0	0	0	0
27	53285	6	883 09-10-2018 19:42:32 *	63	55	41	28	0	3	0	0	0	0	0
28	53286	6	884 09-10-2018 19:43:04 *	63	55	41	29	0	3	0	0	0	0	0
29	53287	6	885 09-10-2018 19:43:33 *	63	55	42	29	0	3	0	0	0	0	0
30	53288	6	886 09-10-2018 19:44:05 *	63	55	42	29	0	3	0	0	0	0	0
31	53289	6	887 09-10-2018 19:44:34 *	63	55	43	29	0	3	0	0	0	0	0
32	53410	6	1008 09-10-2018 20:45:03 *	63	55	66	45	0	3	0	0	0	1	43
33	53411	6	1009 09-10-2018 20:45:33 *	63	55	66	45	0	3	0	0	0	1	43
34	53412	6	1010 09-10-2018 20:46:04 *	63	55	66	45	0	3	0	0	0	1	43

Nota. Figura que muestra una captura de pantalla de los datos iniciales del proyecto.

Como se puede observar esta tabla no cuenta con ninguna estandarización, con nombre de atributos incomprensibles, con atributos no requeridos entre otras características. Es así como se comenzó identificando todos los atributos de la tabla para determinar los que serían utilizados en la creación del modelo, y por el nombre de estos fue necesario analizar el contenido para determinar de qué se trataba.

Con el análisis de la tabla original se procedió a realizar un modelo entidad relación para comprender de mejor manera la relación de los atributos obteniendo el mostrado en la Figura 9 desarrollado en la herramienta Power Designer.

Figura 9.
Modelo entidad relación



Nota. El gráfico presenta el modelo entidad relación diseñado para el proyecto.

A continuación, ahora si refiriéndose al caso de estudio se procedió a estructurar un modelo multidimensional (ver

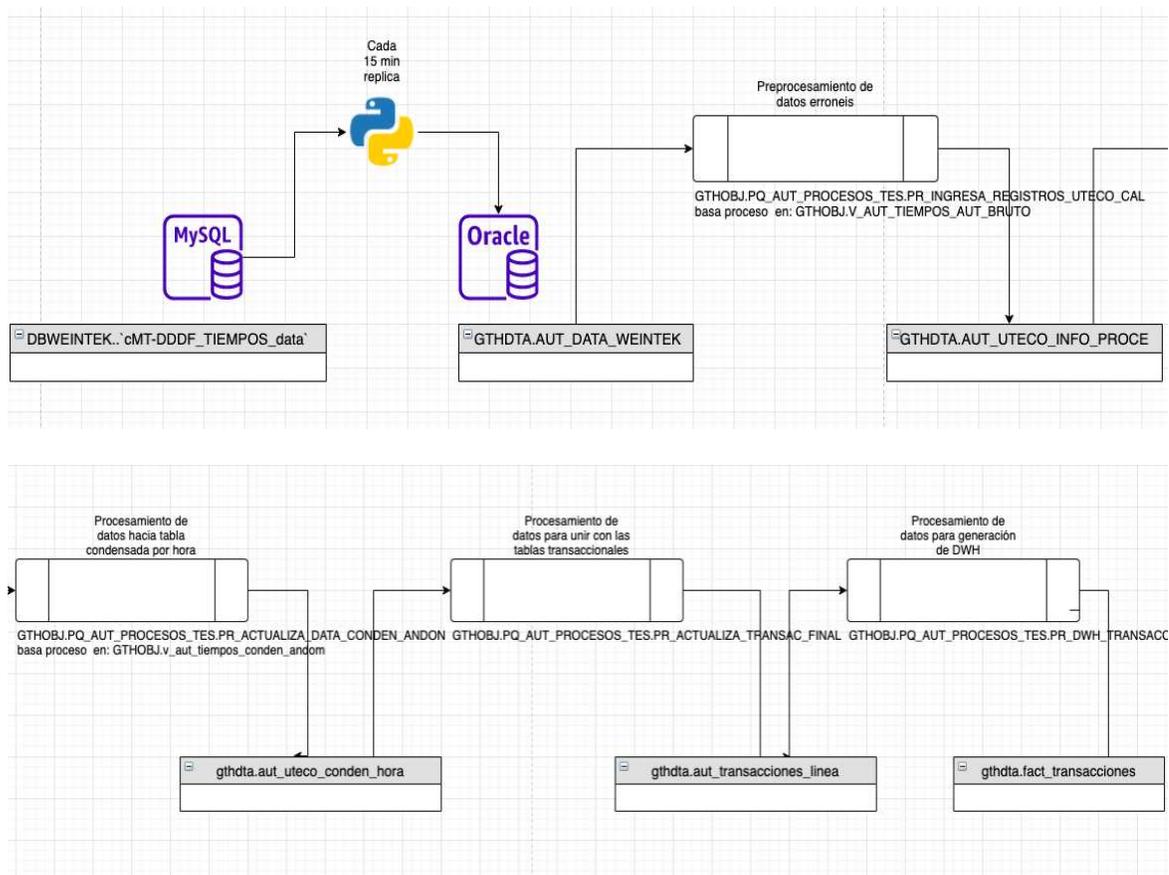
Figura 10) que será el punto de referencia para la creación de la bodega de datos.

Limpieza de los datos

El proceso de limpieza de datos siguió el proceso mostrado en la *Figura 11*, donde se detalla mediante un diagrama de bloques el procedimiento seguido para obtener la data final.

Figura 11.

Proceso limpieza de datos



Nota. La figura muestra el proceso de limpieza de datos ejecutada en el proyecto.

Debido a la ventaja de usar una base de datos más robusta y estable como lo es Oracle y a que la empresa caso de estudio posee las respectivas licencias para el manejo de este sistema de gestión de base de datos se procedió en primera instancia a realizar pasar la data a una base de datos de Oracle desde unas base de datos MySQL. Como lo indica la Figura 12, este proceso se lo realiza cada 15 min.

Figura 12.

Paso uno del proceso

A continuación, se procedió a realizar un preprocesamiento de los datos erróneos, tomando como datos de entrada los enviados por el proceso de réplica, mostrados en la Figura 14.

Figura 14.

Datos traídos desde MySQL

	DWE_ID	TSI_ID	DWE_DATA_INDEX	DWE_FECHA	DWE_NCOL0	DWE_NCOL1	DWE_NCOL2	DWE_NCOL3	DWE_NCOL4	DWE_NCOL5	DWE_NCOL6	DWE_NCOL7	DWE_NCOL8	DWE_NCOL9	DWE_NCOL10
1	53259	6	857	09-10-2018 19:29:33	63	55	28	20	0	3	0	0	0	0	0
2	53260	6	858	09-10-2018 19:30:02	63	55	29	20	0	3	0	0	0	0	0
3	53261	6	859	09-10-2018 19:30:32	63	55	29	21	0	3	0	0	0	0	0
4	53262	6	860	09-10-2018 19:31:02	63	55	30	21	0	3	0	0	0	0	0
5	53263	6	861	09-10-2018 19:31:34	63	55	30	22	0	3	0	0	0	0	0
6	53264	6	862	09-10-2018 19:32:03	63	55	31	23	0	3	0	0	0	0	0
7	53265	6	863	09-10-2018 19:32:33	63	55	31	24	0	3	0	0	0	0	0
8	53266	6	864	09-10-2018 19:33:04	63	55	32	24	0	3	0	0	0	0	0
9	53267	6	865	09-10-2018 19:33:35	63	55	32	24	0	3	0	0	0	0	0
10	53268	6	866	09-10-2018 19:34:04	63	55	33	25	0	3	0	0	0	0	0
11	53269	6	867	09-10-2018 19:34:33	63	55	33	25	0	3	0	0	0	0	0
12	53270	6	868	09-10-2018 19:35:04	63	55	34	26	0	3	0	0	0	0	0
13	53271	6	869	09-10-2018 19:35:33	63	55	34	26	0	3	0	0	0	0	0
14	53272	6	870	09-10-2018 19:36:04	63	55	35	26	0	3	0	0	0	0	0

Nota. Esta imagen muestra los datos que fueron extraídos.

El objeto V_AUT_TIEMPOS_AUT_BRUTO mostrado en la Figura 14 es una vista de la base de datos que toma la información en bruto de la tabla replicada a Oracle desde el MySQL, pero ya con la traducción de campos, para darle significado al sentido de la data. Adicional con un filtro que se basa en la columna.

Para comenzar con la depuración de los datos, se utiliza el procedimiento PR_INGRESA_REGISTROS_UTEICO que revisa en primera instancia las secuencias que deben instalar la tabla de Oracle GTHDTA.AUT_UTEICO_INFO_PROCE para instalar los registros y realizar los primeros cálculos y validaciones. Parte de la línea de código manejado para esta etapa se puede observar en la Figura 15.

Figura 15.

Código de depuración de la data, revisión instancias

```
IF i.takt_time IS NOT NULL AND i.actual_takt_time is not null AND i.uni_con_takt IS NOT NULL then
  ln_uni_maximas := round((i.takt_time*i.uni_con_takt)/i.actual_takt_time);
END IF;
IF ln_uni_maximas IS NOT NULL AND i.uni_reales IS NOT NULL THEN
  ln_uni_retraso := round(ln_uni_maximas - i.uni_reales);
END IF;
IF ln_uni_retraso IS NOT NULL AND i.actual_takt_time IS NOT NULL THEN
  ln_tiempo_retraso := round(ln_uni_retraso * i.actual_takt_time);
END IF;
```

Nota. Esta figura muestra el código realizado para la depuración de los datos extraídos mediante la revisión de instancias.

Una vez con los datos reales se procede a realizar agrupaciones por cada supervisor y equipo de trabajo, con esto se reduce la data y se tienen los condensados por hora. Seguidamente se tiene un objeto de base de datos en una vista llamada V_AUT_TIEMPOS_AUTOMATIZACION, en donde se realizaron las primeras depuraciones de cada registro por fechas y por valores cargados para revisar que la data esté correcta, esto se hace por equipos de trabajo, esto se puede visualizar en la Figura 16.

Figura 16.

Código depuración de la data, agrupaciones

```

CREATE OR REPLACE VIEW V_AUT_TIEMPOS_AUTOMATIZACION AS
SELECT dwe.dwe_id,
       dwe.tsl_id,
       dwe.fecha,
       supervisor,
       takt_time,
       actual_takt_time,
       h_inicio,
       h_fin,
       t_almuerzo,
       t_refrigerio,
       /*debido a data muy ruidosa se realizan las siguientes validaciones de turnos*/
       CASE
         WHEN trunc.fecha) <= to_date('24082018', 'DDMMYYYY') THEN
           2 --antes de esta fecha solo estaba un turno corriendo que era el dos
         WHEN trunc.fecha) > to_date('24082018', 'DDMMYYYY') AND
           trunc.fecha) <= to_date('04102018', 'DDMMYYYY') THEN --por supervisores se sabe que turno tuvieron en este rango de fechas
           CASE
             WHEN supervisor = 'A VIDA' THEN
               2
             WHEN supervisor = 'V BRIT' THEN
               3
             ELSE
               0
           END
         ELSE --fuera de estas fechas los datos en teoria van a venir limpios
           turno
       END turno,
       dwe.uip_fecha_turno fecha_turno,
       uni_con_takt,

```

```

SQL Output Statistics
END turno,
dwe.uip_fecha_turno fecha_turno,
uni_con_takt,
CASE
  WHEN nvl(dwe.uni_con_takt -
    (SELECT b.uni_con_takt
     FROM gthdta.aut_uteco_info_proce b
     WHERE b.uip_id = dwe.uip_id_ant),
    0) < 0 THEN
    dwe.uni_con_takt
  ELSE
    nvl(dwe.uni_con_takt -
    (SELECT b.uni_con_takt
     FROM gthdta.aut_uteco_info_proce b
     WHERE b.uip_id = dwe.uip_id_ant),
    0)
END uni_planeadas,
uni_reales,
CASE
  WHEN nvl(dwe.uni_reales -
    (SELECT b.uni_reales
     FROM gthdta.aut_uteco_info_proce b
     WHERE b.uip_id = dwe.uip_id_ant),
    0) < 0 THEN
    dwe.uni_reales
  ELSE
    nvl(dwe.uni_reales -
    (SELECT b.uni_reales
     FROM gthdta.aut_uteco_info_proce b
     WHERE b.uip_id = dwe.uip_id_ant),
    0)
END uni_fabricadas,

```

Nota. Esta figura muestra el código realizado para la depuración de los datos extraídos mediante las agrupaciones.

También se tiene un segundo objeto de base de datos que es una vista llamada V_AUT_TIEMPOS_CONDEN_ANDOM, en donde se generan las agrupaciones por equipos de trabajo por día y hora, para tener condensada la información que viene ya calculada del objeto anterior. Luego, una vez que se tiene condensada la información, se procede a la inserción en la tabla gthdta.aut_uteco_conden_hora, este proceso se realiza

solo con información de dos días para que la consulta no sea demasiado grande, teniendo un código como el mostrado en la Figura 17.

Figura 17.

Código de inserción data

```

PROCEDURE PR_ACTUALIZA_DATA_CONDEN_ANDON IS
BEGIN
DELETE FROM gthdta.aut_uteco_conden_hora
WHERE fecha >= trunc(sysdate - 2);

INSERT INTO gthdta.aut_uteco_conden_hora(h_inicio,h_fin,t_almuerzo,t_refrigerio,turno,supervisor,fecha,fecha_turno,anio,mes,dia,num_semana,
hora,dia_semana,takt_time,actual_takt_time,equipo,uni_con_takt,uni_reales,nro_llamadas_prod,tiempo_
tiempo_solucion_prod,nro_llamadas_bodega,tiempo_atencion_bodega,tiempo_solucion_bodega,nro_llamadas_mant,
tiempo_atencion_mant,tiempo_solucion_mant,llamadas_uteco,tiempo_solucion_uteco,nro_llamadas_prod_acum,
tiempo_atencion_prod_acum,tiempo_solucion_prod_acum,nro_llamadas_bodega_acum,tiempo_atencion_bodega_acum,
tiempo_solucion_bodega_acum,nro_llamadas_mant_acum,tiempo_atencion_mant_acum,tiempo_solucion_mant_acum,
llamadas_uteco_acum,tiempo_solucion_uteco_acum,unidades_retrazo_acum,tiempo_retrazo_acum,unidades_p
unidades_para_acum,tiempo_total_para_acum,tiempo_para_equipos_acum,uni_planeadas,uni_fabricadas,
uni_maximas,uni_retrazo,tiempo_retrazo)
SELECT h_inicio,h_fin,t_almuerzo,t_refrigerio,turno,supervisor,fecha,fecha_turno,anio,mes,dia,num_semana,
hora,dia_semana,takt_time,actual_takt_time,equipo,uni_con_takt,uni_reales,nro_llamadas_prod,tiempo_atencion_prod,
tiempo_solucion_prod,nro_llamadas_bodega,tiempo_atencion_bodega,tiempo_solucion_bodega,nro_llamadas_mant,
tiempo_atencion_mant,tiempo_solucion_mant,llamadas_uteco,tiempo_solucion_uteco,nro_llamadas_prod_acum,
tiempo_atencion_prod_acum,tiempo_solucion_prod_acum,nro_llamadas_bodega_acum,tiempo_atencion_bodega_acum,
tiempo_solucion_bodega_acum,nro_llamadas_mant_acum,tiempo_atencion_mant_acum,tiempo_solucion_mant_acum,
llamadas_uteco_acum,tiempo_solucion_uteco_acum,unidades_retrazo_acum,tiempo_retrazo_acum,unidades_maximas_acum,
unidades_para_acum,tiempo_total_para_acum,tiempo_para_equipos_acum,uni_planeadas,uni_fabricadas,
uni_maximas,uni_retrazo,tiempo_retrazo
FROM
gthobj.v_aut_tiempos_conden_andom
WHERE fecha in (trunc(sysdate - 2),trunc(sysdate - 1));
END;

```

Nota. En la figura se muestra parte del código de como se cargaron los datos sometidos al proceso ETL.

Una vez con la información correcta para los datos de la línea, se procede con la unificación del sistema transaccional con las tablas de maestros como son turnos, supervisor, equipos, y temas transaccionales. Luego se realiza el proceso de extracción de la información según el modelo entidad relación, también, se depura la misma y se transforma a el modelo multidimensional mediante la tabla GTHOBJ.PQ_AUT_PROCESOS_TES.PR_DWH_TRANSACCIONES.

Finalmente como parte de la extracción de datos se realizó primero una interacción con las tablas que arman tanto las dimensiones como la tabla de FACT. En las tablas de dimensiones se procede con la extracción de la información en bruto de las tablas considerando que en el caso que la tabla se encuentre completamente vacía y se insertan todos los registros de primera mano. Para esto se genera una secuencia, que es

la llave principal, se transforman los datos realizando revisión de espacios en blanco, valores nulos, así como la transformación a numérico o carácter dependiendo del tipo de dato a ser insertado, este procedimiento se lo puede observar en la Figura 18.

Figura 18.

Ejemplo código carga de datos a dimensión

```

INSERT /*+ APPEND */
INTO gthdta.dim_gestion_linea
  SELECT rownum + ln_secuencia_genera gel_id,
         nvl(to_number(TRIM(a.tur_numero)), 0) gel_numero_turno,
         nvl(to_char(TRIM(a.tur_descripcion), 'SIN/DESCRIPCION/TURNO') gel_descripcion_turno,
         nvl(to_char(TRIM(b.sup_codigo), 'SIN/CODIGO/SUPERVISOR') gel_codigo_supervisor,
         nvl(to_char(TRIM(b.per_observaciones), 'SIN/OBSERVACIONES') gel_observaciones_supervisor,
         nvl(to_char(TRIM(e.per_nombres || ' ' || e.per_apellidos),
         'SIN/NOMBRE') gel_nombre_supervisor,
         nvl(to_char(TRIM(c.equ_descripcion),
         'SIN/DESCRIPCION/EQUIPOS') gel_descripcion_equipos,
         a.tur_id,
         b.sup_id,
         c.equ_id,
         e.per_id
  FROM   gthdta.aut_turnos      a,
         gthdta.aut_supervisor b,
         gthdta.aut_equipos     c,
         gthdta.aut_personas    e
  WHERE  EXISTS (SELECT 'x'
                 FROM   gthdta.aut_transacciones_linea d
                 WHERE  d.tur_id = a.tur_id
                 AND    d.sup_id = b.sup_id
                 AND    d.equ_id = c.equ_id)
  AND    e.per_id = b.per_id
  AND    NOT EXISTS (SELECT 'X'
                    FROM   gthdta.dim_gestion_linea h
                    WHERE  h.tur_id = a.tur_id
                    AND    h.sup_id = b.sup_id
                    AND    h.equ_id = c.equ_id
                    AND    h.per_id = e.per_id);

```

Nota. Esta figura presenta parte del código que permitió cargar los datos a las respectivas dimensiones.

En el caso de que no sea la primera carga y los datos se encuentran ya insertados se procede a ejecutar el mismo SQL anterior, sin embargo, en la parte del “where” se tiene una sentencia que permite saber cuáles no se insertaron y se procede a insertar los mismos. Como se pueden realizar modificaciones en las tablas transaccionales, se

procede a realizar la actualización de las tablas de dimensiones por completo. Mediante comandos de población masiva mostrados en la Figura 19.

Figura 19.

Comandos de población masiva

```

-- Purpose : ESTE UPDATE SIRVE PARA ACTUALIZAR ABSOLUTAMENTE TODOS LOS DATOS DE GESTION DE LINEA
lv_proceso := 'ACTUALIZA DATOS DE GESTION DE LINEA|UPDATE';
OPEN c_gestion_linea;
LOOP
  FETCH c_gestion_linea BULK COLLECT
  INTO tn_gel_id,
      tn_numero_turno,
      ty_descripcion_turno,
      ty_codigo_supervisor,
      ty_observaciones_supervisor,
      ty_nombre_supervisor,
      ty_descripcion_equipos,
      tn_tur_id,
      tn_sup_id,
      tn_equ_id,
      tn_per_id LIMIT 10;
  EXIT WHEN tn_gel_id.count() = 0;

  FORALL i IN tn_gel_id.first .. tn_gel_id.last
  UPDATE gthdta.dim_gestion_linea a
  SET   gel_numero_turno           = tn_numero_turno(i),
        gel_descripcion_turno      = ty_descripcion_turno(i),
        gel_codigo_supervisor      = ty_codigo_supervisor(i),
        gel_observaciones_supervisor = ty_observaciones_supervisor(i),
        gel_nombre_supervisor      = ty_nombre_supervisor(i),
        gel_descripcion_equipos     = ty_descripcion_equipos(i)
  WHERE a.tur_id = tn_tur_id(i)
  AND   a.sup_id = tn_sup_id(i)
  AND   a.equ_id = tn_equ_id(i)
  AND   a.per_id = tn_per_id(i);
END LOOP;
CLOSE c_gestion_linea;

```

Nota. En la figura se presenta parte del código para actualizar todos los datos y cargar la información completa.

Al final se tiene un control completo de errores en el caso de que la población y se finaliza con la tabla de FACT realizando una población completa de datos realizado con el código de la

Figura 20.

Figura 20.

Código carga de datos a FACT

```

SQL      Output  Statistics
insert /*APPEND*/ into FACT_TRANSACCIONES
SELECT  a.trl_id,
        b.gel_id,
        c.emp_id,
        d.tie_id,
        a.emp_id emr_id,
        a.trl_fecha_turno,
        a.trl_hora_inicio,
        a.trl_hora_fin,
        a.trl_tiempo_almuerzo,
        a.trl_tiempo_refrigerio,
        a.trl_takt_time,
        a.trl_actual_takt_time,
        a.trl_uni_con_takt,
        a.trl_uni_reales,
        a.trl_nro_llamadas_prod,
        a.trl_tiempo_atencion_prod,
        a.trl_tiempo_solucion_prod,
        a.trl_nro_llamadas_bodega,
        a.trl_tiempo_atencion_bodega,
        a.trl_tiempo_solucion_bodega,
        a.trl_nro_llamadas_mant,
        a.trl_tiempo_atencion_mant,
        a.trl_tiempo_solucion_mant,
        a.trl_llamadas_uteco,
        a.trl_tiempo_solucion_uteco,
        a.trl_uni_planeadas,
        a.trl_uni_fabricadas,
        a.trl_uni_maximas,
        a.trl_uni_retrazo,
        a.trl_tiempo_retrazo
FROM    aut_transacciones_linea a,
        dim_gestion_linea      b,
        dim_empleados          c,
        dim_tiempos            d
WHERE   b.tur_id = a.tur_id
AND     b.sup_id = a.sup_id
AND     b.equ_id = a.equ_id
AND     c.per_id = a.per_id
AND     d.tie_id = a.trl_fecha

```

Nota. La figura muestra parte del código utilizado para cargar la información a la FACT.

Por último los datos de las dimensiones y de hecho quedaron almacenadas en la respectiva base de datos Oracle como los muestran las siguientes figuras.

Figura 21.

Datos cargados de la FACT

TRL_ID	GEL_ID	EMP_ID	TRL_ID	EMP_ID	TRL_FECHA_TURNOS	TRL_HORA_INICIO	TRL_HORA_FIN	TRL_TIEMPO_ALMUERZO	TRL_TIEMPO_REFRIGERIO	TRL_TAKT_TIME	TRL_ACTUAL_TAKT_TIME	TRL_UNI_CON_TAKT	TRL_UNI_REALIZADO
1	32040	32	4695	24-09-2018	4	19:03:05		0	0	72	62	213	
2	32050	32	4695	25-09-2018	4	19:03:05		0	0	72	62	430	
3	32051	32	4695	25-09-2018	4	11:03:15		1907	0	72	62	253	
4	32052	35	4695	26-09-2018	4	19:01:24		0	0	72	62	273	
5	32053	32	4695	25-09-2018	4	00:00:00		0	0	72	62	530	
6	32054	35	4695	26-09-2018	4	00:00:00		0	0	72	62	506	
7	32055	32	4695	27-09-2018	4	06:59:31		0	0	72	62	49	
8	32056	36	4695	22-04-2019	4	15:26:53		0	4050	60	50	398	
9	32057	32	4695	27-09-2018	4	06:59:31		2397	0	72	62	316	
10	32058	31	4695	22-04-2019	4	06:56:46	15:26:17	0	0	60	50	242	
11	32059	32	4695	29-09-2018	4	19:02:48		0	0	72	62	320	
12	32060	35	4695	28-09-2018	4	00:00:00		1817	0	72	62	324	
13	32061	35	4695	28-09-2018	4	00:00:00		1817	0	72	62	374	
14	32062	32	4695	29-09-2018	4	00:00:00		0	0	72	62	533	
15	32063	35	4695	01-10-2018	4	07:54:18		2275	0	72	62	520	
16	32064	35	4695	01-10-2018	4	07:54:18		2275	0	72	62	505	
17	32065	32	4695	29-09-2018	4	19:05:48		0	0	80	62	33	
18	32066	32	4695	29-09-2018	4	06:59:49		1962	0	72	62	222	
19	32067	35	4695	29-09-2018	4	06:59:49		2022	0	72	62	271	
20	32068	35	4695	29-09-2018	4	06:59:49		0	0	72	62	149	
21	32069	35	4695	02-10-2018	4	07:03:05		3135	2429	63	62	541	
22	32070	32	4695	02-10-2018	4	19:00:31		0	0	63	62	243	
23	32071	32	4695	03-10-2018	4	19:00:31		0	0	63	62	300	
24	32072	32	4695	03-10-2018	4	19:00:31		0	0	63	62	438	
25	32073	35	4695	02-10-2018	4	07:03:05		3135	2429	72	62	477	
26	32074	35	4695	02-10-2018	4	07:03:05		0	0	72	62	200	
27	32075	35	4695	02-10-2018	4	07:03:05		1955	0	72	62	223	
28	32076	35	4695	04-10-2018	4	10:42:57		2021	0	63	55	42	
29	32077	35	4695	04-10-2018	4	10:42:57		2453	0	63	55	92	
30	32078	35	4695	04-10-2018	4	07:05:14		0	0	63	62	172	
31	32079	32	4695	04-10-2018	4	18:58:17		0	0	63	62	399	
32	32080	32	4695	04-10-2018	4	00:00:00		0	0	63	62	621	
33	32081	35	4695	04-10-2018	4	10:42:57		0	0	63	62	36	
34	32082	32	4695	05-10-2018	4	18:58:17		0	0	63	62	556	
35	32083	26	4695	05-10-2018	4	07:51:37		3361	0	63	61	114	

Nota. La imagen muestra los datos cargados de la FACT en la base de datos Oracle.

Figura 22.

Datos cargados de la dimensión Gestión

GEL_ID	GEL_NUMERO_TURNOS	GEL_DESCRIPCION_TURNOS	GEL_CODIGO_SUPERVISOR	GEL_OBSERVACIONES_SUPERVISOR	GEL_NOMBRE_SUPERVISOR	GEL_DESCRIPCION_EQUIPOS
1	1	5 Turno Adicional Nocturno	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	VISION
2	2	2 Turno Normal Nocturno	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	VISION
3	3	3 Turno Adicional Mañana	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	VISION
4	4	1 Turno Normal Tarde	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	VISION
5	5	5 Turno Normal Mañana	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	VISION
6	6	2 Turno Normal Nocturno	A VIDA	SUPERVISOR TURNOS NOCTURNOS	AMELIA VALENTINA VIDAL SAQUINAULA	VISION
7	7	3 Turno Adicional Mañana	A VIDA	SUPERVISOR TURNOS NOCTURNOS	AMELIA VALENTINA VIDAL SAQUINAULA	VISION
8	8	5 Turno Adicional Nocturno	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	VISION
9	9	2 Turno Normal Nocturno	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	VISION
10	10	3 Turno Adicional Mañana	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	VISION
11	11	1 Turno Normal Tarde	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	VISION
12	12	0 Turno Normal Mañana	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	VISION
13	13	2 Turno Normal Nocturno	A VIDA	SUPERVISOR TURNOS NOCTURNOS	AMELIA VALENTINA VIDAL SAQUINAULA	VISION
14	14	3 Turno Adicional Mañana	A VIDA	SUPERVISOR TURNOS NOCTURNOS	AMELIA VALENTINA VIDAL SAQUINAULA	VISION
15	15	5 Turno Adicional Nocturno	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	PILOTOS
16	16	2 Turno Normal Nocturno	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	PILOTOS
17	17	3 Turno Adicional Mañana	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	PILOTOS
18	18	1 Turno Normal Tarde	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	PILOTOS
19	19	0 Turno Normal Mañana	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	PILOTOS
20	20	2 Turno Normal Nocturno	A VIDA	SUPERVISOR TURNOS NOCTURNOS	AMELIA VALENTINA VIDAL SAQUINAULA	PILOTOS
21	21	3 Turno Adicional Mañana	A VIDA	SUPERVISOR TURNOS NOCTURNOS	AMELIA VALENTINA VIDAL SAQUINAULA	PILOTOS
22	22	5 Turno Adicional Nocturno	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	DREAM
23	23	2 Turno Normal Nocturno	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	DREAM
24	24	3 Turno Adicional Mañana	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	DREAM
25	25	1 Turno Normal Tarde	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	DREAM
26	26	0 Turno Normal Mañana	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	DREAM
27	27	2 Turno Normal Nocturno	A VIDA	SUPERVISOR TURNOS NOCTURNOS	AMELIA VALENTINA VIDAL SAQUINAULA	DREAM
28	28	3 Turno Adicional Mañana	A VIDA	SUPERVISOR TURNOS NOCTURNOS	AMELIA VALENTINA VIDAL SAQUINAULA	DREAM
29	29	5 Turno Adicional Nocturno	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	IMPECABLES
30	30	2 Turno Normal Nocturno	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	IMPECABLES
31	31	3 Turno Adicional Mañana	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	IMPECABLES
32	32	1 Turno Normal Tarde	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	IMPECABLES
33	33	0 Turno Normal Mañana	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	IMPECABLES
34	34	2 Turno Normal Nocturno	A VIDA	SUPERVISOR TURNOS NOCTURNOS	AMELIA VALENTINA VIDAL SAQUINAULA	IMPECABLES
35	35	3 Turno Adicional Mañana	A VIDA	SUPERVISOR TURNOS NOCTURNOS	AMELIA VALENTINA VIDAL SAQUINAULA	IMPECABLES
36	36	5 Turno Adicional Nocturno	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	HIBRIDOS
37	37	2 Turno Normal Nocturno	V BRIT	SUPERVISOR TURNOS DIA	VICENTE GUILLERMO BRITO ESCOBAR	HIBRIDOS

Nota. La imagen muestra los datos cargados de la dimensión Gestión en la base de datos Oracle.

Figura 23.

Datos cargados de la dimensión Empleados

EMP_ID	EMP_NOMBRE	EMP_TIPO_IDENTIFICACION	EMP_NRO_IDENTIFICACION	EMP_FECHA_EMPLEO	EMP_FECHA_FIN_CONTRATO	EMP_SALARIO	EMP_DESCRIPCION_CARGO
1	MARCO HUMBERTO ORELLANA OLIVARES	CED	0704364058	27-11-2012	04-09-2040	385	COBRADOR VERIFICADOR MIXTO
2	HERNAN RUBEN GRANIZO CONSTANTE	CED	1204769630	27-11-2012	04-09-2040	385	COBRADOR VERIFICADOR MIXTO
3	PABLO DANLO RAMOS TACO	CED	0921172854	06-12-2012	05-12-2013	324	COBRADOR - VERIFICADOR
4	VIRGINIA KERLIZ MOREIRA COX	CED	1205828567	13-12-2012	12-12-2013	324	VENDEDOR MINOREO
5	MARLON MANUEL CURILLO RAMON	CED	0706642840	13-12-2012	12-12-2013	324	COBRADOR - VERIFICADOR
6	PABLO JIMMY BURGOS SALTOS	CED	1307798163	13-12-2012	21-09-2040	411	COBRADOR VERIFICADOR
7	LUIS ALBERTO RIVAS MALLA	CED	0104232491	20-12-2012	19-12-2013	650	AUDITOR INTERNO
8	LADY PAMELA GONGORA COCHANCELA	CED	0921219101	16-11-2012	31-12-2012	298	PROMOTOR DE VENTAS
9	JOSÉ ERNESTO SANGUOZZA GUERRERO	CED	1714551767	23-08-2016	31-05-2294	396	COBRADOR VERIFICADOR CORRIENTE
10	MARCO RAFAEL RIVERA FLORES	CED	0302438387	15-11-2012	14-11-2013	298	VENDEDOR MINOREO
11	PEDRO EVERALDO SOLORZANO MENDOZA	CED	0912107406	15-11-2012	23-08-2040	411	COBRADOR VERIFICADOR
12	SEGUNDO ANTONIO DAVILA RAMOS	CED	0906423918	15-11-2012	23-08-2040	411	COBRADOR VERIFICADOR CORRIENTE
13	CARLOS MANUEL ESTEVEZ PACHECO	CED	1205573718	27-11-2012	26-11-2013	324	COBRADOR - VERIFICADOR
14	CARLOS GUSTAVO GUEVARA RODRIGUEZ	CED	1205713280	18-12-2012	26-09-2040	411	COBRADOR VERIFICADOR
15	MICHELE VERONICA CASTILLO MELO	CED	0921018388	21-12-2013	09-01-2293	720	ASISTENTE ADMINISTRATIVO COMERCIAL
16	ERIKA VANESSA DOMINGUEZ MOSQUERA	CED	0705375608	16-11-2012	31-12-2012	298	PROMOTOR DE VENTAS
17	LUIS FERNANDO QUIROGA QUIROGA	CED	0104750864	15-11-2012	09-01-2293	515	AUXILIAR DE LOGISTICA
18	WILMER GONZALO VILLAMAR PAZ	CED	1310212145	22-11-2012	09-03-2293	520	ASISTENTE SERVICIO TECNICO
19	DIEGO SEBASTIAN NAUTA TAPIA	CED	0103805093	01-03-2013	09-12-2040	742	ESPECIALISTA DE SOPORTE MX
20	DIEGO SEBASTIAN NAUTA TAPIA	CED	0103805093	14-02-2017	29-11-2290	1050	ESPECIALISTA EN TELECOMUNICACIONES Y SERVIDORES
21	JOHNNY VINICIO ANDRADE ARMIJOS	CED	0104838487	06-12-2012	05-12-2013	325	OPERADOR PLASTICOS MANUALES
22	MIGUEL FERNANDO LLESCAS GUILLEN	CED	0104493841	06-12-2013	14-09-2291	1100	TECNICO DE MANTENIMIENTO ELECTRICO
23	EDUARDO FABRICIO PAREDES ONTANEDA	CED	1714091855	06-12-2012	14-09-2040	366	COBRADOR VERIFICADOR MIXTO
24	AMBROSIO ALFREDO VASCONEZ GUZMAN	CED	1717704884	01-07-2015	09-04-2293	500	ASISTENTE DE BOEAGA
25	WILLIAM ALEXANDER PINOS ALVARADO	CED	0106693869	11-12-2012	10-12-2013	325	OPERADOR ENSAMBLE FI LINDO
26	JUAN CARLOS CAMRHO ESPAÑA	CED	1205903259	04-12-2012	12-09-2040	411	COBRADOR VERIFICADOR
27	HENRY GERMAN MUÑOZ MOROCHO	CED	0103660627	11-12-2013	19-09-2291	1100	TECNICO DE MANTENIMIENTO MECANICO
28	ALEX JAVIER ESPINOZA CIGUENZA	CED	0923703417	20-12-2012	28-09-2040	366	COBRADOR VERIFICADOR MIXTO
29	DIEGO GEOVANNY FEJOO NASPUCHA	CED	0105512040	22-11-2012	09-01-2293	425	RECIPIOR/DESPACHADOR
30	ALEXI FERNANDEZ CERVANTES CEVALLOS	CED	1309982369	04-12-2012	03-12-2013	324	COBRADOR - VERIFICADOR
31	FERNANDO PATRICIO ORELLANA ROSERO	CED	0603621906	04-12-2012	03-12-2013	324	COBRADOR - VERIFICADOR
32	BYRON FABIAN CUENCA PERALTA	CED	1103957229	02-10-2012	01-10-2013	324	BOLEGUERO
33	JORGE EDUARDO CALLE VERDUGO	CED	0105003562	02-10-2012	01-10-2013	324	PROMOTOR DE MARCA
34	GLORIA MELISSA CASTRO VALAREZO	CED	2400089965	22-04-2013	31-05-2013	324	PROMOTOR DE VENTAS
35	DIANA ELIZABETH SAN PEDRO MONSERRATE	CED	0922330121	05-11-2012	31-12-2012	298	PROMOTOR DE VENTAS
36	GABRIELA JENIFFER RUEDA OROZZO	CED	0804281194	05-11-2012	31-12-2012	298	PROMOTOR DE VENTAS

Nota. La imagen muestra los datos cargados de la dimensión Empleados en la base de datos Oracle.

Figura 24.

Datos cargados de la dimensión Empresa

EMR_ID	EMR_EMPRESA	EMR_RUC	EMR_RAZON_SOCIAL	EMR_DIRECCION	EMR_TELEFONO
1	INDURAMA	0190061264001	INDUGLOB S.A.		
2	BLANCANDINA		BLANCANDINA		
3	MARCIMEX	0190001490001	MARCIMEX S.A.		
4	SERVANDINA		SERVANDINA		
5	MIGRACION				
6	ENERCH	0190400400001	COMPANIA DE TRANSPORTES ENERCH S.A.		
7	TERCERIZADORAS		TERCERIZADORAS		
8	CUARMUEBLES		CUARMUEBLES S.A.		
9	CYBERSUR		CYBERSUR		
10	BLANCANDINA NICARAGUA	0190061264001	BLANCANDINA NICARAGUA		
11	BLANCANDINA COLOMBIA	0190061264001	BLANCANDINA COLOMBIA		
12	BLANCANDINA GUATEMALA	0190061264001	BLANCANDINA GUATEMALA		
13	BLANCANDINA EL SALVADOR	0190061264001	BLANCANDINA EL SALVADOR		
14	BLANCANDINA HONDURAS	0190061264001	BLANCANDINA HONDURAS		
15	BLANCANDINA COSTA RICA	0190061264001	BLANCANDINA COSTA RICA		
16	BLANCANDINA REP. DOMINICANA	0190061264001	BLANCANDINA REP. DOMINICANA		
17	BLANCANDINA URUGUAY	0190061264001	BLANCANDINA URUGUAY		
18	CONANSA		CONANSA		
19	SEGPSOSES CALITDA.		SEGPSOSES CALITDA.		
20	AUSTROMAR		AUSTROMAR		
21	AUSTROMAR - TECNOACES		TECNOACES		
22	CUARMUEBLES - TECNOACES		TECNOACES		
23	ENERCH - TECNOACES		TECNOACES		
24	IGM	0992767550001	IGM		
25	INDUGLOB ARANDIANO		ARANDIANO		
26	INDUGLOB CUENKCLEAN		CUENKCLEAN		
27	SERVICO		SERVICO		
28	SERVANDINA - TECNOACES		TECNOACES		
29	INNOVA		INNOVA		
30	GLOBAL		GLOBAL		
31	BLANCANDINA PANAMA	0190061264001	BLANCANDINA PANAMA		
32	ELECTROANDINA		ELECTROANDINA		
33	LIMPORT		LIMPORT		
34	BLANCANDINA BOLIVIA	0190061264001	INDUGLOB S.A.	A	7
35	BLANCANDINA PAISES CARIBE	0190061264001	INDUGLOB S.A.		
36	CYBERSUR TECNOACES		TECNOACES		

Nota. La imagen muestra los datos cargados de la dimensión Empresa en la base de datos Oracle.

Figura 25.

Datos cargados de la dimensión Tiempo

TIE_ID	TME_MES	TTR_TRIMESTRE	TAN_ANIO	TIE_FECHA_JULIANA	TIE_FECHA_ANTERIOR	TIE_FECHA_ANIO_ANTERIOR	TIE_FECHA_ACTUAL	TIE_NOMBRE_DIA	TIE_MES_ACTUAL	TIE_FECHA_FIN_MES	TME_MES_ACTUAL_FIN
1	28-12-1993	199312	1993	93362	*		19-03-2020	*	202003	31-12-1993	202003
2	30-12-1993	199312	1993	93364	*		19-03-2020	*	202003	31-12-1993	202003
3	28-03-1990	199003	1990	90087	*		19-03-2020	*	202003	31-03-1990	202003
4	23-06-1997	199706	1997	97174	*		19-03-2020	*	202003	30-06-1997	202003
5	29-06-1997	199706	1997	97180	*		19-03-2020	*	202003	30-06-1997	202003
6	01-07-1997	199707	1997	97182	*		19-03-2020	*	202003	31-07-1997	202003
7	04-07-1997	199707	1997	97185	*		19-03-2020	*	202003	31-07-1997	202003
8	07-07-1997	199707	1997	97188	*		19-03-2020	*	202003	31-07-1997	202003
9	09-07-1997	199707	1997	97190	*		19-03-2020	*	202003	31-07-1997	202003
10	10-07-1997	199707	1997	97191	*		19-03-2020	*	202003	31-07-1997	202003
11	18-07-1997	199707	1997	97199	*		19-03-2020	*	202003	31-07-1997	202003
12	21-07-1997	199707	1997	97202	*		19-03-2020	*	202003	31-07-1997	202003
13	22-07-1997	199707	1997	97203	*		19-03-2020	*	202003	31-07-1997	202003
14	26-07-1997	199707	1997	97207	*		19-03-2020	*	202003	31-07-1997	202003
15	27-07-1997	199707	1997	97208	*		19-03-2020	*	202003	31-07-1997	202003
16	28-07-1997	199707	1997	97209	*		19-03-2020	*	202003	31-07-1997	202003
17	29-07-1997	199707	1997	97210	*		19-03-2020	*	202003	31-07-1997	202003
18	02-08-1997	199708	1997	97214	*		19-03-2020	*	202003	31-08-1997	202003
19	03-08-1997	199708	1997	97215	*		19-03-2020	*	202003	31-08-1997	202003
20	04-08-1997	199708	1997	97216	*		19-03-2020	*	202003	31-08-1997	202003
21	05-08-1997	199708	1997	97217	*		19-03-2020	*	202003	31-08-1997	202003
22	09-03-1979	197903	1979	79068	*		19-03-2020	*	202003	31-03-1979	202003
23	12-03-1979	197903	1979	79071	*		19-03-2020	*	202003	31-03-1979	202003
24	14-03-1979	197903	1979	79073	*		19-03-2020	*	202003	31-03-1979	202003
25	15-03-1979	197903	1979	79074	*		19-03-2020	*	202003	31-03-1979	202003
26	16-03-1979	197903	1979	79075	*		19-03-2020	*	202003	31-03-1979	202003
27	17-03-1979	197903	1979	79076	*		19-03-2020	*	202003	31-03-1979	202003
28	19-03-1979	197903	1979	79078	*		19-03-2020	*	202003	31-03-1979	202003
29	21-03-1979	197903	1979	79080	*		19-03-2020	*	202003	31-03-1979	202003
30	25-03-1979	197903	1979	79084	*		19-03-2020	*	202003	31-03-1979	202003
31	26-03-1979	197903	1979	79085	*		19-03-2020	*	202003	31-03-1979	202003
32	28-03-1979	197903	1979	79087	*		19-03-2020	*	202003	31-03-1979	202003
33	29-03-1979	197903	1979	79088	*		19-03-2020	*	202003	31-03-1979	202003
34	30-03-1979	197903	1979	79089	*		19-03-2020	*	202003	31-03-1979	202003
35	01-04-1979	197904	1979	79091	*		19-03-2020	*	202003	30-04-1979	202003
36	04-04-1979	197904	1979	79094	*		19-03-2020	*	202003	30-04-1979	202003

Nota. La imagen muestra los datos cargados de la dimensión Tiempo en la base de datos Oracle.

Elección herramientas y algoritmos

Para la elección de las herramientas con las que se iban a trabajar, se consideró como selección principal, las provistas por la empresa caso de estudio, ya que de esta manera le interacción con los resultados serían más fáciles de usar y confiables.

Además, se realizó un análisis de la importancia de cada de las herramientas, que denotó que las mismas satisfacían los requerimientos del proyecto.

Oracle

Oracle es una herramienta que ha mejorado significativamente la capacidad de gestión de proyectos y reducción de costos operativos en algunas grandes empresas. Al permitir integrar todas actividades comerciales con un seguimiento preciso de los costos, tomando decisiones estratégicas, informadas y ofrecidas con un mayor valor a los clientes (Oracle, 2020).

Negocios de moda mediante comercio electrónico requirió una estrategia digital sólida para satisfacer las necesidades de los clientes, donde utilizando Oracle todos estos requerimientos fueron solventados (Oracle, 2020).

Mantenimiento, progreso de la versión, incluso administración de procesadores: todo es más rápido con Oracle. Además, las aplicaciones dentro del tablero, se puede hacer con el mismo trabajo pero en una décima parte del tiempo que solía tomar. Oracle es fácil de administrar y, sobre todo, rápido de administrar (Oracle, 2020).

Estas son algunas descripciones dadas de Oracle por empresas casos de éxito, que han mejorado sus sistemas de gestión de datos gracias a Oracle, por lo tanto, el uso de esta herramienta dio un plus al desarrollo del proyecto.

Tableau

La analítica de datos para una empresa puede significar el medio de acceso a la información concreta para la toma de decisiones y estrategias de negocios en base a criterios rigurosos de una manera fácil.

Actualmente, existen potentes herramientas de inteligencia de negocios (Business Intelligence o BI) y visualización de datos para ayudar a una organización a ilustrar sus ideas, visualizar sus datos, compartir sus análisis con los clientes y la comunidad global en general.

Es así que para llevar a cabo estos análisis existen en el mercado algunas herramientas de BI entre las que destaca Tableau, liderando por siete años el Cuadrante Mágico de Gartner. Aunque, no todas las herramientas enfocadas a negocio son iguales. Tampoco los líderes designados por Gartner deberían ser la solución perfecta para una organización.

Desde el 2010 Tableau ya aparecía en la sección de retadores (challengers) en el cuadrante mágico de Gartner. Con los años fue ganando popularidad hasta que en 2014 se colocó líder en el top de herramientas de Business Intelligence. Desde entonces únicamente ha sido en 2019 cuando ha sido superado por Power Bi de Microsoft como se lo puede observar hasta el último informe mostrado en la Figura 26.

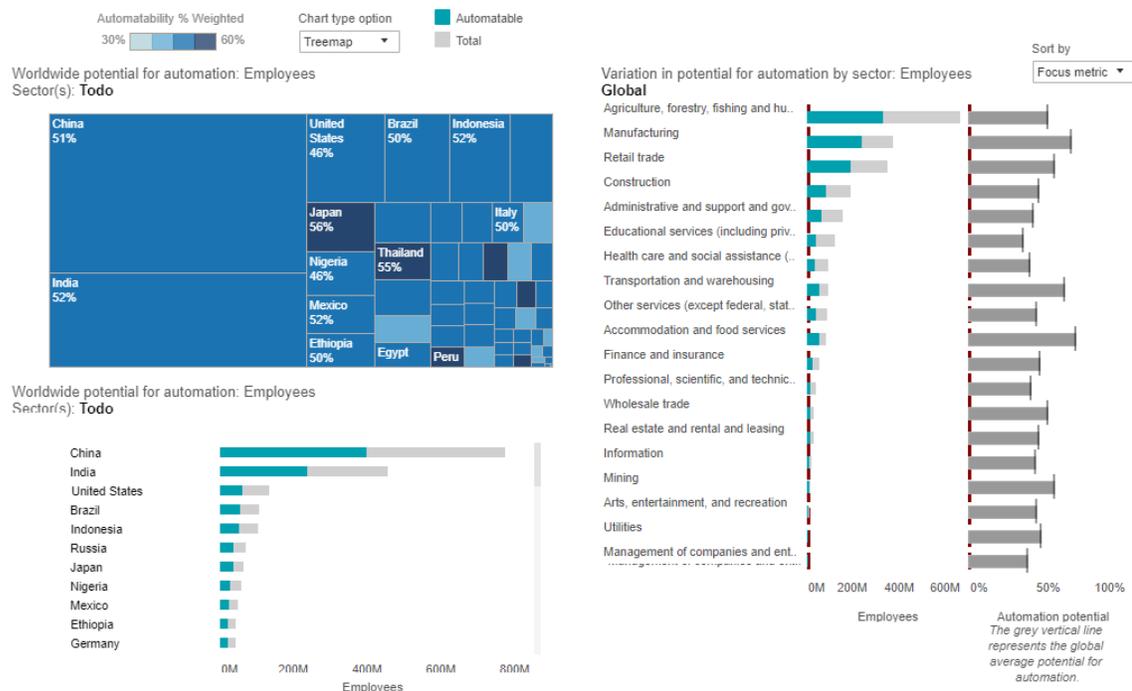
Figura 26.

Cuadrante Mágico de Gartner 2020 herramientas BI.



Nota. La figura muestra las herramientas distribuidas dentro de los cuadrantes de Gartner. Tomado de (García-Calvo Fernández, 2020).

Entre las características consideradas para clasificar a Tableau dentro del cuadrante mágico de Gartner en el grupo de los líderes de las herramientas de BI, está su enfoque al análisis de datos a través de medios de visualización como se lo observa en la *Figura 27*.

Figura 27.**Interfaz Tableau**

Nota. La figura muestra la interfaz gráfica que presenta Tableau para el análisis de datos Tomado de (Russo, 2020). Russo, M. (2020, febrero).

Además, se consideró servicios que Tableau brinda como: capacidad de colaboración en tiempo real, varias formas de compartir los informes, capacidad de compartir de manera gratuita (ciertas limitaciones), soporte para conexión a más de 30 tipos de fuentes de datos, mezcla fuentes de datos, soporte para cubos OLAP, integración con mapeo de drivers listos para muchas bases de datos, liderazgo en los esfuerzos de construcción de la comunidad (varios videos de capacitación, blogs, foros o participación en redes sociales). Esta magnitud de opciones permite una gran flexibilidad y elimina muchas restricciones (Russo, 2020).

Lo más destacado de la herramienta es su potente facilidad de uso, considerada la más fácil de usar de esta lista. Ninguna otra solución de BI puede hacerlo más fácil que Tableau ya que otorga a los usuarios un poder asombroso. Es lo suficientemente fácil

para un usuario de negocios informal y, a la vez, tan poderosa como las herramientas de desarrollo (Russo, 2020).

En la Tabla 3, se puede observar las ventajas y desventajas que ofrece la herramienta al usuario final.

Tabla 3

Pros & Contras de Tableau

PROS	CONTRAS
<ul style="list-style-type: none"> • Interfaz de usuario intuitiva y atractiva. • Integración perfecta con muchas plataformas modernas de Big Data desde Hadoop hasta BigQuery. • Proporciona una extensa lista de conexiones de datos nativas que permiten una fácil integración con datos de muchas fuentes. • Capacidad de respuesta puesto Tableau es compatible con plataformas móviles. • Poderosa colaboración comunitaria • Desarrollo constante de nuevas actualizaciones, se lanzan regularmente y son fáciles de instalar. • Atención al Cliente confiable. • Una amplia biblioteca de materiales de vídeo sobre la herramienta, cursos en línea, blogs de aprendizaje 	<ul style="list-style-type: none"> • Se requiere preparación de datos inicial. • Las características pueden parecer demasiado especializadas y restrictivas ya que Tableau está diseñada para un uso amplio. • Aunque es excelente para fines analíticos no puede reemplazar las aplicaciones de informes financieros. • Brinda la capacidad de establecer seguridad de “nivel bajo” en el nivel de datos pero la implementa de una manera un poco precaria. • No existe un concepto de versiones con el servidor Tableau. • Requiere alguna consultoría especializada de TI.

Nota. La tabla presenta las ventajas y desventajas de utilizar la herramienta Tableau. Tomado de (Russo, 2020).

Tableau tiene diferentes planes de licencias similares a otras soluciones de BI, para fines de este proyecto se usará la versión Tableau Desktop. Es una versión de escritorio destinada a cualquier usuario individual. El uso personal implica que está destinado a un desarrollador individual y admite 6 fuentes de datos, mientras que el uso empresarial significa que está destinado a las necesidades comerciales y permite hasta 44 fuentes de datos.

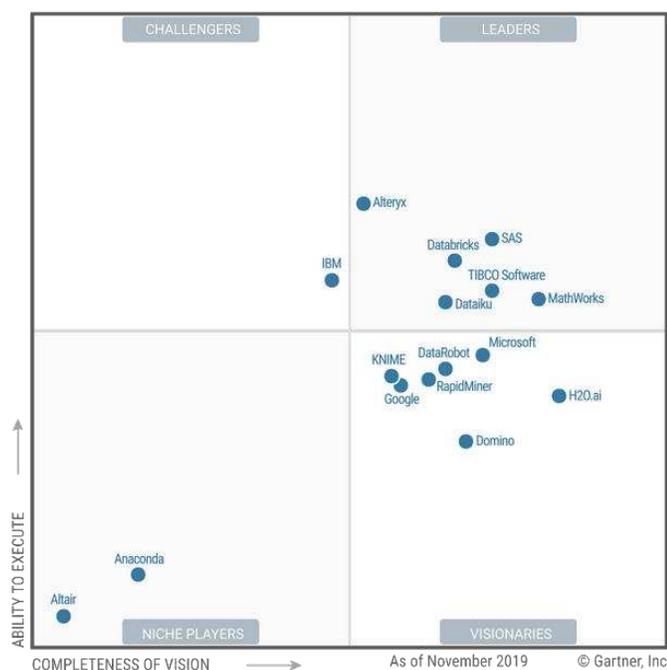
Rapidminer

De acuerdo con el cuadrante mágico de Gartner mostrado en la Figura 28, Rapidminer se encuentra en el cuadrante de los visionarios. Las fortalezas por las que Rapidminer fue seleccionado son la facilidad para administrar su flujo de ciencia de datos de extremo a extremo, ofrece un programa de certificación a través de su Academia, flexibilidad, agilidad y escalabilidad moderna.

Rapidminer también permite la contenerización con Docker y Kubernetes para ejecutar y escalar modelos de forma transparente. En su nueva solución Model Ops automatizada, RapidMiner permite el cumplimiento y la auditoría de modelos tanto en las instalaciones como en la nube. Esta gobernanza cubre no solo el linaje de datos, sino también la capacitación del modelo, el uso del modelo, el reciclaje del modelo, el control de versiones del modelo y el acceso al modelo.

Figura 28.

Cuadrante Mágico de Gartner para herramientas de Minería de Datos



Nota. La figura muestra las herramientas de minería de datos distribuidas en los cuadrantes de Gartner. Tomado de (Gartner, 2020).

Las desventajas encontradas son el crecimiento relativo lento y mayor competencia debido parcialmente al resultado de una estrategia comercial que prioriza la rentabilidad sobre el crecimiento. Por cuanto Rapidminer se considera una herramienta potente y lo único que lo desvaloriza es una falta de competitividad ante otras herramientas.

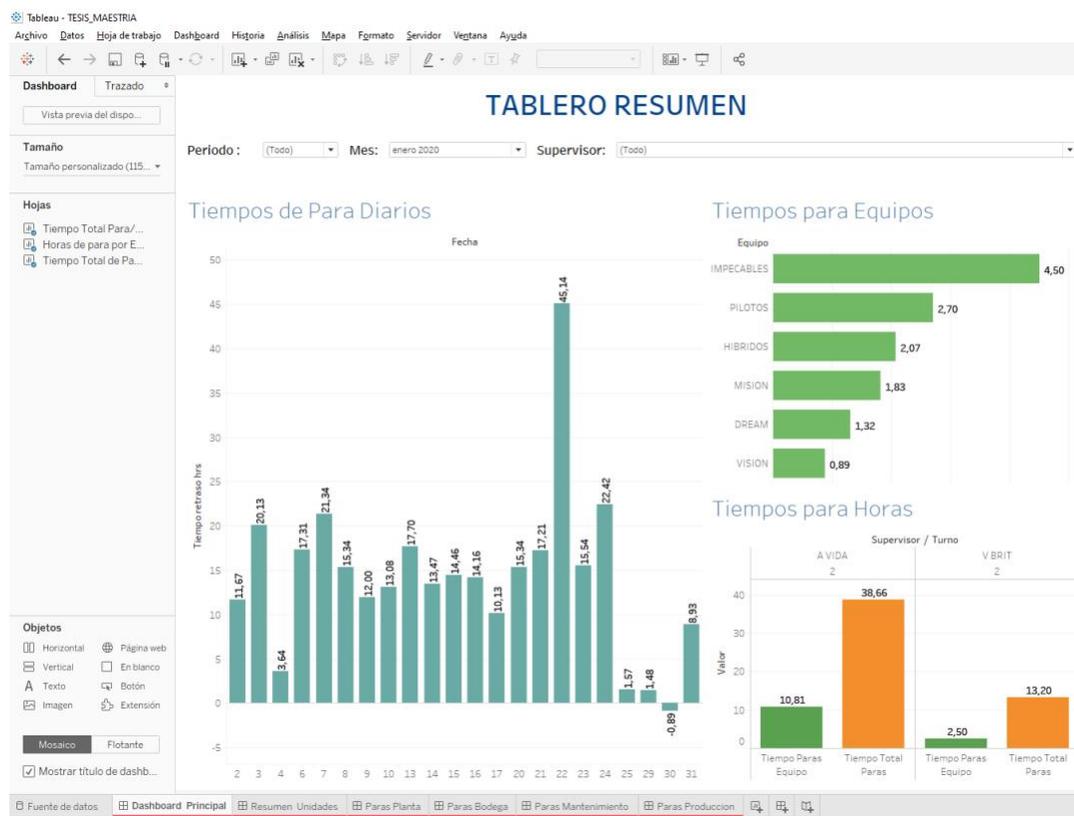
Dashboard descriptivos

Con la data ya analizada y limpia se procedió realizar un conjunto dashboard descriptivos que permitieron observar el comportamiento de los datos en base al propósito del proyecto. Es así como utilizando la herramienta Tableau (ver

Figura 29) se elaboró los informes mostrados a continuación.

Figura 29.

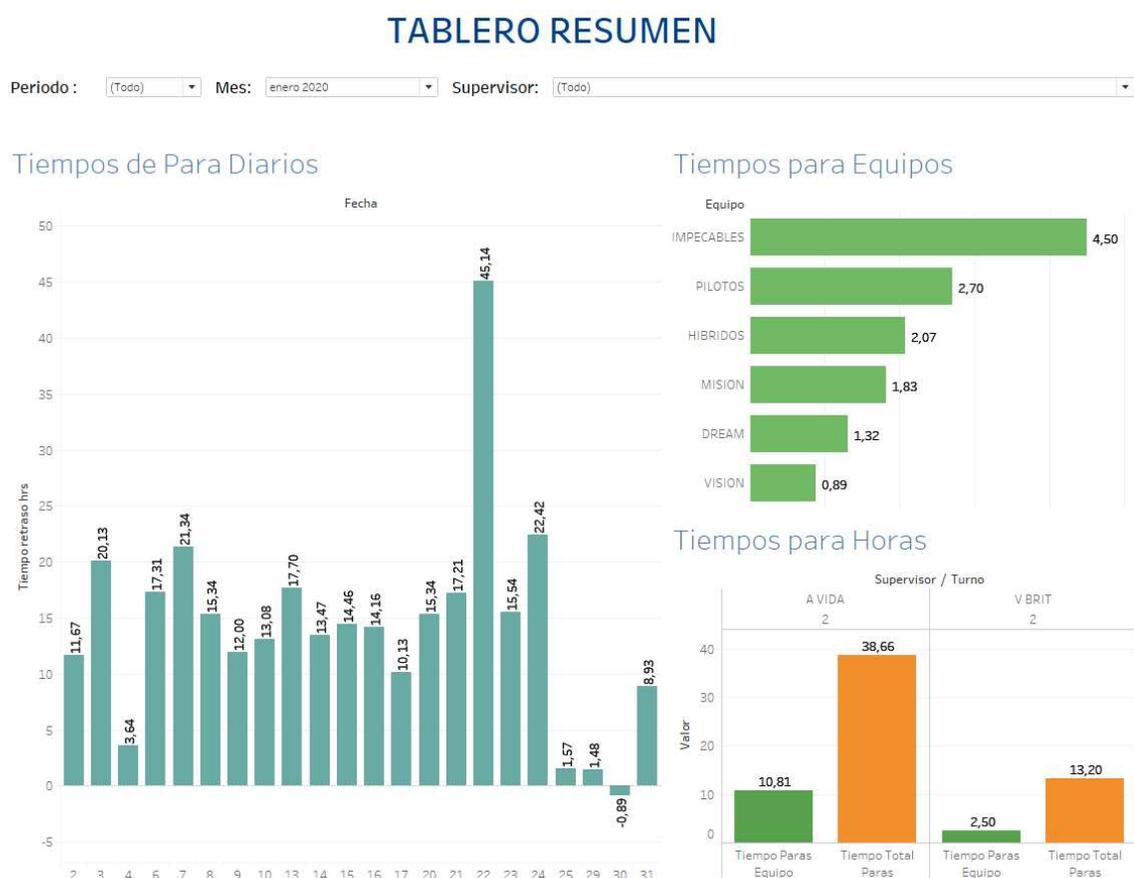
Dashboard en herramienta Tableau



Nota. La figura muestra el ejemplo de un dashboard creado en la herramienta Tableau.

El primer informe creado, permite observar un resumen general de la paras registrados en las etapas de la línea de producción, permitiendo visualizar el comportamiento de los datos obtenidos del sistema de la empresa (Andon), como se lo observa en la

Figura 30.

Figura 30.*Informe descriptivo general*

Nota. La figura muestra un dashboard creado en la herramienta Tableau que muestra de manera general los datos.

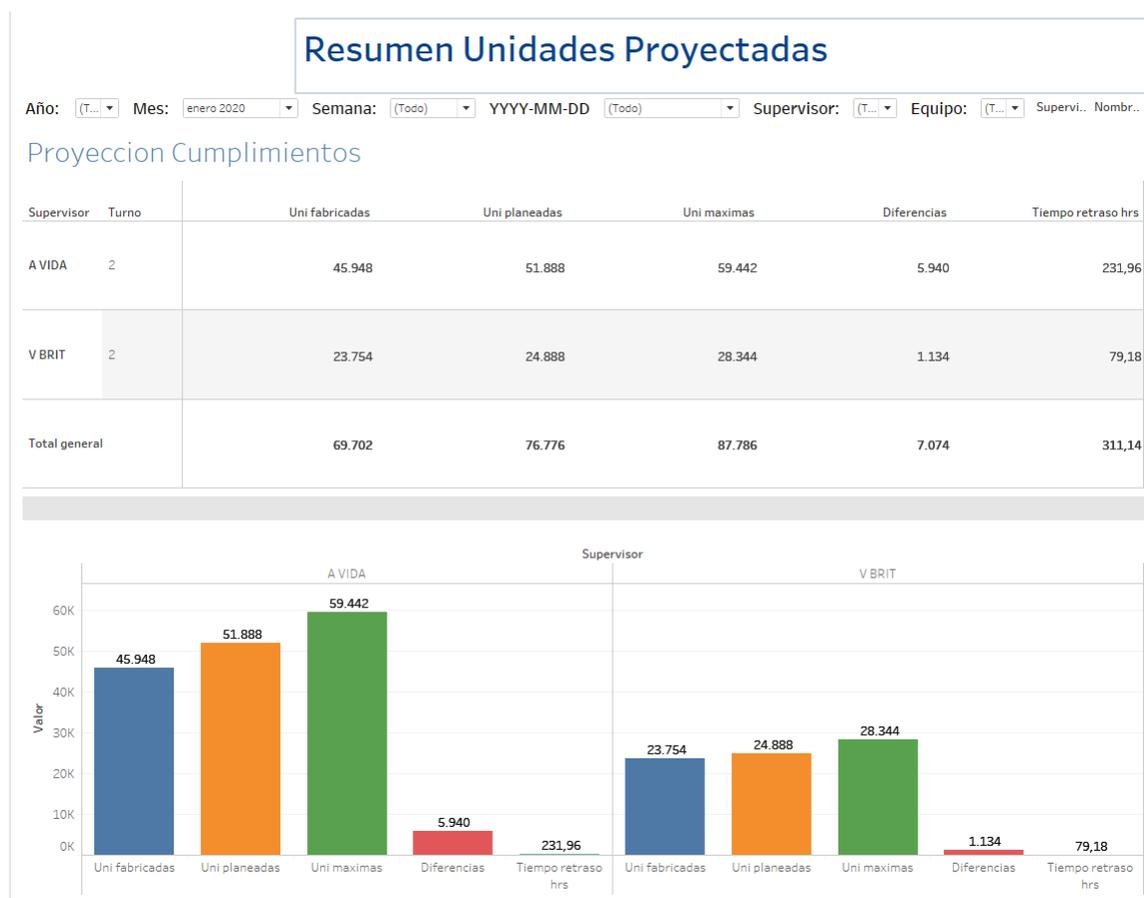
Este dashboard tiene como filtros año, mes y supervisor que permiten determinar el comportamiento de las paras dependiendo del año o mes cuando se produjo o del supervisor que estuvo a cargo. El informe indica de manera gráfica las paras totales por día. Además, el otro gráfico del informe permite observar las paras por equipos de trabajo y las paras causadas por otras causas. Se entiende por paras al tiempo que la empresa dejó de producir, expresado en horas. Con esto informe se permite tener un descriptivo

para revisar en donde ocurren los fallos en la línea y realizar las correcciones correspondientes.

El segundo informe presentado en la Figura 31 muestra las paradas en la línea de producción, mostrando en primera instancia el número de unidades producidas vs el número de unidades reales, consideradas como el número máximo de unidades que se esperaba producir si no se hubieran presentado las paras.

Figura 31.

Informe descriptivo Paradas

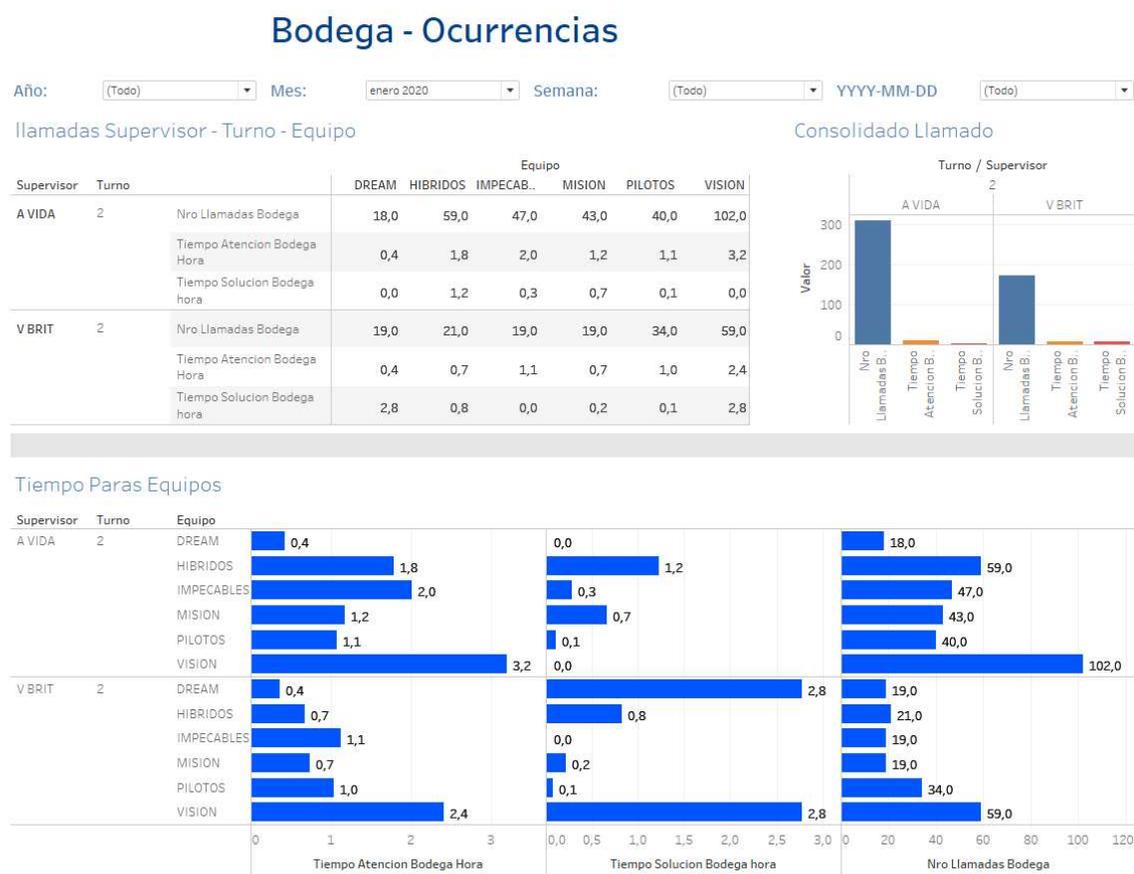


Nota. La figura muestra un dashboard creado en la herramienta Tableau que muestra las paras en línea de producción de la empresa caso de estudio.

En el segundo gráfico presenta el comportamiento de las unidades paradas de los equipos vs el número total de paras de la línea independientemente de las paras de equipos y que se pudieron dar por otras causas. El tercer informe de la Figura 32 hace referencia a las paras de la planta debido a llamados a la bodega.

Figura 32.

Informe descriptivo Llamadas Bodega



Nota. La figura muestra un dashboard creado en la herramienta Tableau que muestra el comportamiento de las unidades paradas de los equipos vs el número total de paras de la línea.

Mediante los 4 tableros mostrados presenta en primera instancia las bodegas por equipos, donde se muestra el número de llamadas a bodega por problemas de inventario para seguir produciendo esto por supervisor y equipos. El segundo tiene el tiempo total

de bodega que indica el tiempo en horas de paras ocasionas por cada llamada por el supervisor. El tercero es bodega por equipos G, donde se muestra el número de llamadas a bodega por problemas de inventario para seguir produciendo, esto por supervisor y equipos con un entorno gráfico. El cuarto tiene el Tiempo Total Bodega G que indica en tiempo de horas cuanto de paras ocasiono cada llamada por supervisor a nivel gráfico.

Los siguientes informes son parecidos al tercero orientados a paras de la planta por llamados a mantenimiento, por problemas de la máquina o paras de la planta por llamadas a producción, debido al mal manejo del equipo, los mismos se pueden observar en la Figura 33 y Figura 34.

Figura 33.

Informe descriptivo Llamadas Mantenimiento



Nota. La figura muestra un dashboard creado en la herramienta Tableau que muestra las paras de la planta por llamados a mantenimiento, por problemas de la máquina.

Figura 34.

Informe descriptivo Llamadas Producción



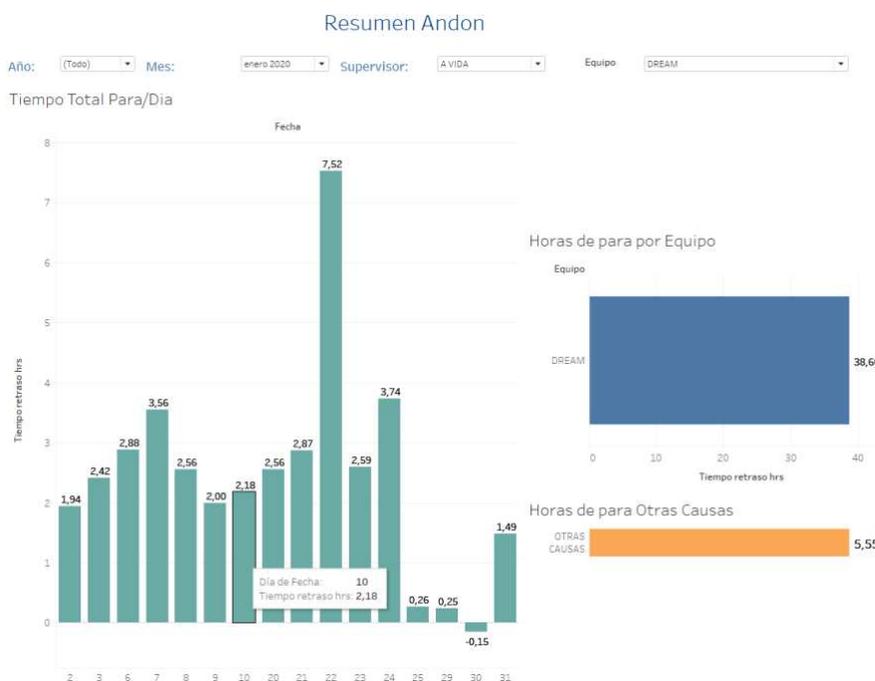
Nota. La figura muestra un dashboard creado en la herramienta Tableau que muestra las paradas de la planta por llamadas a producción, debido al mal manejo del equipo.

Creación del Modelo

Para la creación del modelo de minería de datos se utilizó la característica Auto Model de Rapidminer, que permite analizar un conjunto de datos utilizando varios algoritmos a la vez y en poco tiempo. Además, esta herramienta permite adecuar la data a ser utilizada en la creación del modelo, para obtener resultados óptimos. En el primer paso se selecciona la base de datos con la que se desea trabajar en el modelo, a continuación, fue necesario seleccionar el tipo de proceso a realizarse, siendo para este

proyecto, una predicción. Acompañada de esta característica es necesario indicar el atributo a ser predecido, que para el caso de estudio se consideró al tiempo de para, puesto se desea determinar los patrones de comportamiento que interfieren en este tiempo. A continuación, Rapidminer mostrará los valores que este atributo podría predecir, para en el siguiente paso, determinar la lista las variables de entrada que alimentarán de mejor manera al modelo a crear. Es importante recalcar que la herramienta permite el usuario configurar las características con las que quiere trabajar en todas las fases de este proyecto. Finalmente, la herramienta permite escoger los algoritmos con los que se desea trabajar, seleccionando en este caso los 3 por defecto que presentan los siguientes resultados.

Para analizar los resultados obtenidos en cada modelo se hará referencia al primer informe descriptivo mostrado en la figura, con datos de enero del 2020 con las variables solo de un equipo mostrado en la Figura 35.

Figura 35.*Ejemplo para modelo*

Nota. La figura muestra un dashboard creado en la herramienta Tableau que muestra un ejemplo con datos del enero de 2020.

Algoritmo lineal

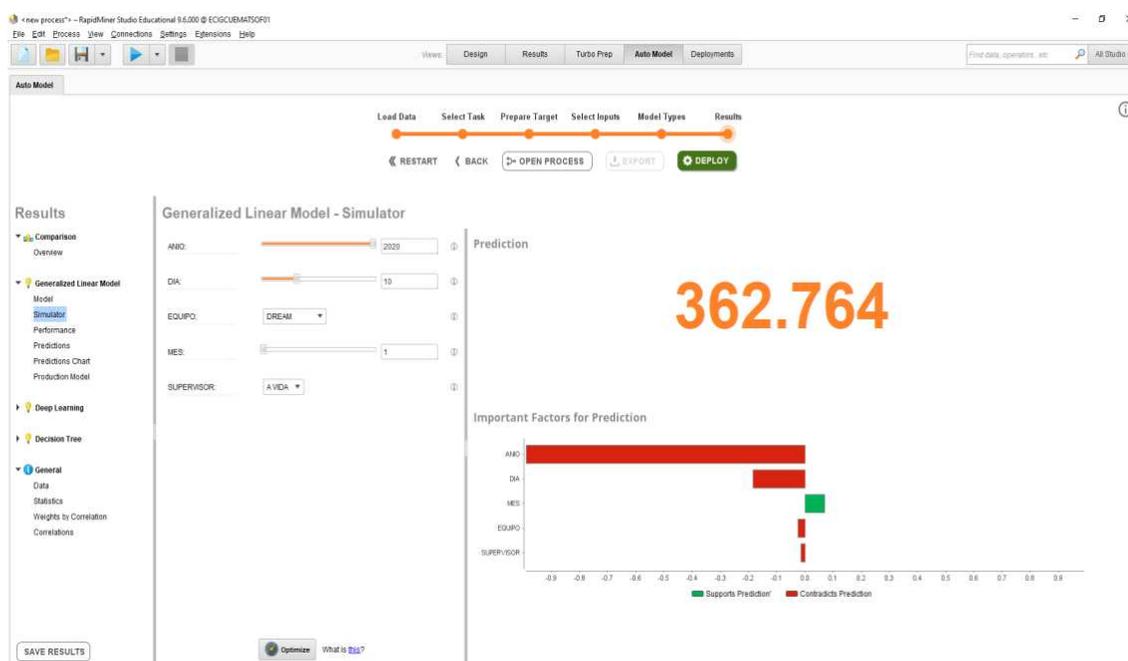
Algoritmo que permite representar la relación entre dos (o más) variables a través de una expresión lógico-matemática que, aparte de resumir cómo es esa relación. Este algoritmo va a permitir realizar predicciones de los valores que tomará una de las dos variables a partir de los valores de la otra utilizando la estadística. Este algoritmo es uno de los más utilizados a la hora de predecir los valores de una variable cuantitativa a partir de los valores de otras variables.

En base a los resultados obtenidos mediante este algoritmo mostrados en la *Figura 36*, se determina que el año es una de las variables que más determinará el cambio de tiempo de para, mientras que le mes no será determinante para este valor.

Para el ejemplo propuesto se puede observar que para los datos antes propuestos, se tendrá una para de 362764 segundos que darían 0,10 horas.

Figura 36.

Modelo con algoritmo Lineal



Nota. La figura muestra la simulación de la predicción usando el modelo con algoritmo lineal.

Algoritmo *Deep Learning*

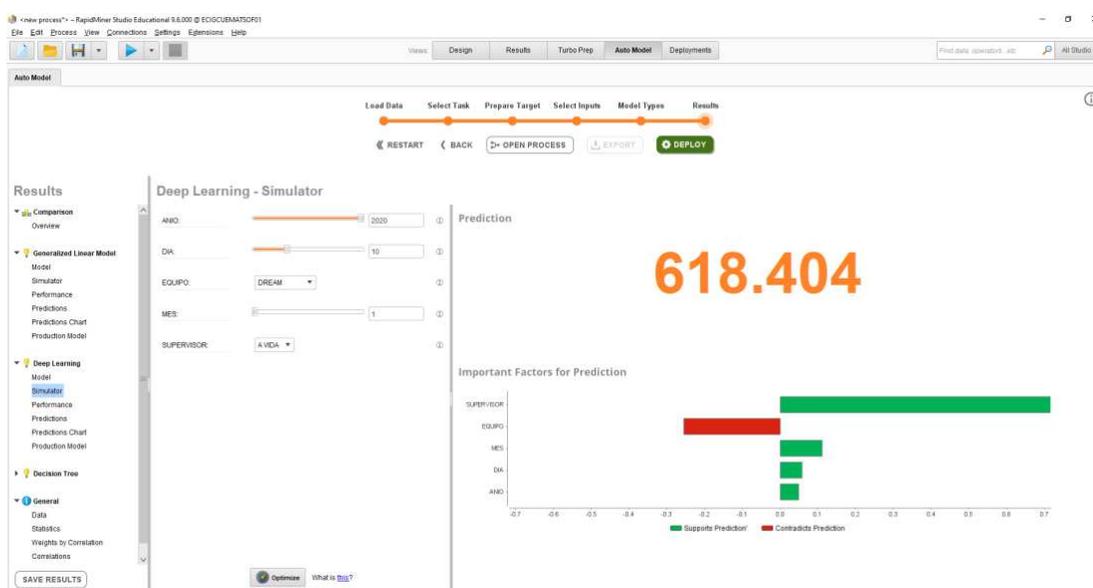
Deep Learning es una técnica de aprendizaje automático que enseña a los ordenadores a hacer lo que resulta natural para las personas: aprender mediante ejemplos. Con el *Deep Learning*, se puede crear modelos que aprende en base a ejemplos tomados en las fases de entrenamiento de la data. Los modelos se entrenan mediante un amplio conjunto de datos etiquetados y arquitecturas de redes neuronales que contienen muchas capas.

Analizando los resultados obtenidos en Rapidminer utilizando este algoritmo mostrado en la Figura 37, se puede observar cómo en este caso el atributo

único y determinante para que el tiempo de para cambie es el equipo, mientras que los otros atributos no contribuyen en el cambio de este tiempo, principalmente no existirá cambios en el tiempo si se cambia el supervisor. De acuerdo, al ejemplo planteado para los mismos datos el tiempo de para sería de 618404 segundos, que equivalen a 0,18 horas.

Figura 37.

Modelo con algoritmo Deep Learning



Nota. La figura muestra la simulación de la predicción usando el modelo con algoritmo *Deep learning*.

Algoritmo Decision Tree

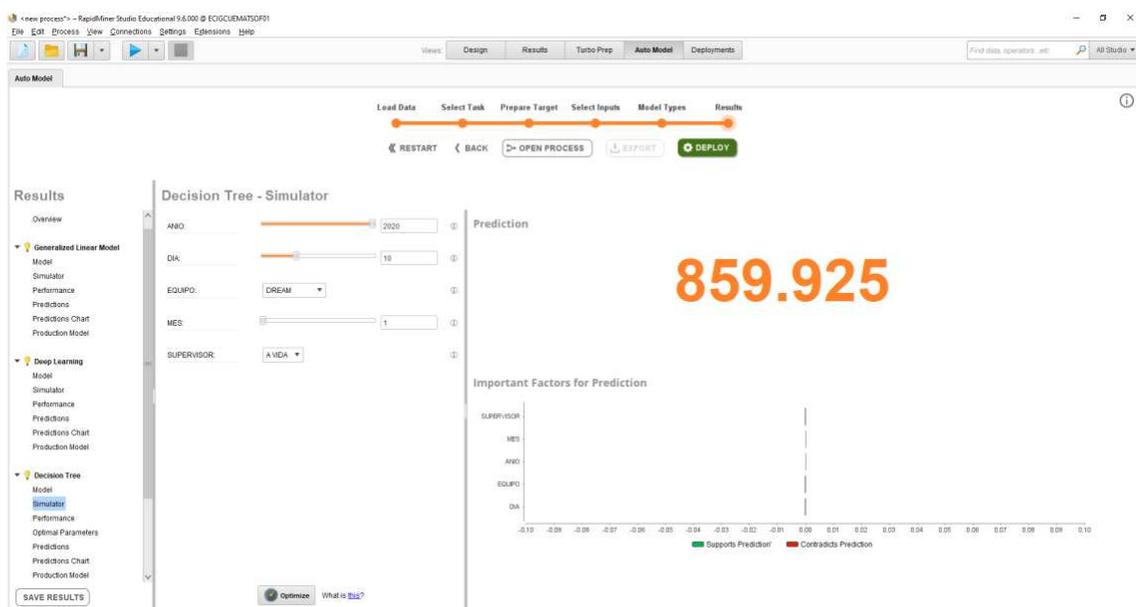
Es un tipo de algoritmo de aprendizaje supervisado que se utiliza principalmente en problemas de clasificación, aunque funciona para variables de entrada y salida categóricas como continuas. En esta técnica, se divide la data en dos o más conjuntos homogéneos, basados en el diferenciador más significativos en las variables de entrada. El árbol de decisión identifica la variable más significativa y su valor proporciona los

mejores conjuntos homogéneos de población. Todas las variables de entrada y todos los puntos de división posibles se evalúan y se elige la que tenga mejor resultado.

De acuerdo, a este algoritmo se puede observar en la Figura 38, que ninguna de las variables de entrada consideradas ocasionarán cambios en los tiempos de para y para el ejemplo propuesto la predicción con los datos de entrada propuestos darían un tiempo de para de 0,23 horas.

Figura 38.

Modelo con algoritmo Decision Tree



Nota. La figura muestra la simulación de la predicción usando el modelo con algoritmo decision tree.

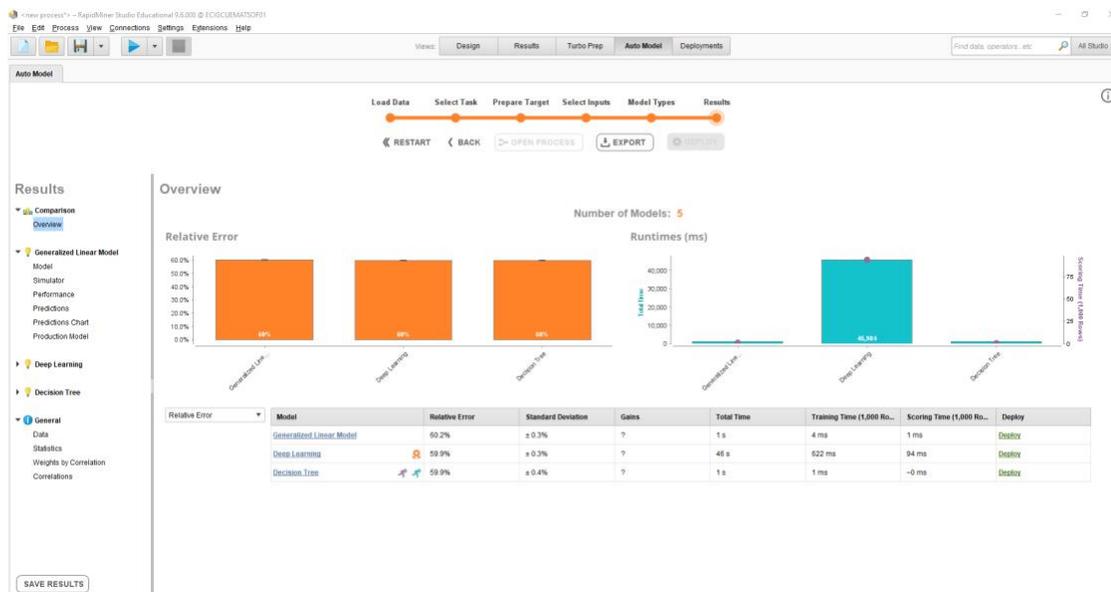
Evaluación del modelo

Para evaluar los modelos creados se puede observar el comportamiento de los modelos de acuerdo a cada uno de los algoritmos, se debe considerar que Rapidminer utiliza una validación simple donde divide el conjunto de datos en dos un conjunto de entrenamiento y uno pequeño de prueba, los de prueba no se deben usar en el entrenamiento, se usan para determinar la confianza o error del modelo.

En base a los resultados obtenidos y mostrados en la Figura 39, se puede determinar que el algoritmo que más confianza presenta el algoritmo lineal.

Figura 39.

Resultados evaluación modelos



Nota. La figura muestra la evaluación de los modelos implementados en los datos analizados.

Capítulo V. Conclusiones y Recomendaciones

Conclusiones

La revisión inicial de la literatura, permitió conocer y entender de mejor manera el giro del negocio. Además, permitió determinar las características de las herramientas a utilizar puesto que al utilizarse herramientas dadas por la empresa es necesario conocer las mismas.

Mediante la recolección y el análisis de la información existente en la empresa sobre la línea de producción de refrigeración, se identificó valores concurrentes que ayudaron a determinar patrones de comportamiento. De esta manera, se comprendió la necesidad de pronosticar los patrones de comportamiento en el proceso de producción. También, se observó como alarmas de llamados a mantenimiento, bodega y producción influyen en los tiempos de retraso en la producción. Este proceso también interfiere en la calidad de los productos en sus líneas de fabricación.

La facilidad de usar Rapidminer está en que permite realizar la comparación entre algunos algoritmos de minería de datos, para de esta manera con los mismos datos identificar el algoritmo que mejor se acopla al caso de estudio y el que brinda mayor confianza. En este caso se pudo observar como el algoritmo de decisión tree presenta un error absoluto menor que lo definiría como el algoritmo más adecuado para el estudio realizado.

El modelo de minería de datos creado se lo verificó con tres algoritmos determinando algunos patrones que afectan la baja producción y calidad de los productos por tiempos de para.

Evaluar el modelo analítico – predictivo a través de la herramienta Rapidminer permitió identificar de una manera rápida el algoritmo que refleja mejor nivel de confianza de acuerdo a los datos analizados.

Utilizar una metodología al analizar los datos permitió organizar el flujo de trabajo de una manera eficiente, es decir, considerando los pasos a seguir adecuadamente para que los datos obtenidos reflejen un resultados descriptivo y predictivo de los patrones de interés en el caso de estudio.

Los tiempos de retardo originados por llamados a mantenimiento son los que mayor influencia tiene en las paras de la línea de producción.

De acuerdo al algoritmo Deep Learning la variación de equipo influirá mayormente en los tiempos de para, mientras que para el algoritmo lineal el atributo que más influye es el año de análisis.

Recomendaciones

Es relevante que los informes descriptivos sean desarrollados con el apoyo de algún trabajador de la empresa, para que de esta manera éstos reflejen los datos que ellos requieren.

Utilizar una herramienta visual para mostrar los datos facilita la comprensión de los mismos y que sirva de mejor manera en la toma de decisiones.

El modelo de minería de datos podría ser implementados en una aplicación de visualización para poder ser entregado a la empresa.

La interacción entre la data y los empleados es importante para obtener resultados adecuados, puesto que si los datos son ingresados adecuadamente las respuestas será más confiables.

Bibliografía

Acosta, V. M. (2019, mayo 17). Diferencia entre Data Mart y Data warehouse. ¡Descúbrela! *Canal Informática y TICS*. <https://revistadigital.inesem.es/informatica-y-tics/diferencia-entre-data-mart-y-data-warehouse/>

Bello, O., & Leonardo, F. (2016). *Clasificación de historias provenientes de familiares de víctimas y sobrevivientes del conflicto sociopolítico de Colombia*. <http://repository.javeriana.edu.co/handle/10554/21198>

Bordera, M. (2018, julio 5). *¿Tableau? Las 9 ventajas de Tableau vs. BI Tradicional | Clavei*. Clavei | Expertos en Transformación Digital. <https://www.clavei.es/blog/tableau-las-9-ventajas-de-tableau-vs-bi-tradicional/>

Cabezas, E., Andrade, D., & Torres, J. (2018). *Introduccion a la Metodologia de la investigacion científica.pdf*. <http://repositorio.espe.edu.ec/jspui/bitstream/21000/15424/1/Introduccion%20a%20la%20Metodologia%20de%20la%20investigacion%20cientifica.pdf>

Chen, X., Ye, Y., Williams, G., & Xu, X. (2007). A Survey of Open Source Data Mining Systems. En T. Washio, Z.-H. Zhou, J. Z. Huang, X. Hu, J. Li, C. Xie, J. He, D. Zou, K.-C. Li, & M. M. Freire (Eds.), *Emerging Technologies in Knowledge Discovery and Data Mining* (Vol. 4819, pp. 3-14). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-77018-3_2

De Benito Crosetti, B., & Salinas Ibáñez, J. M. (2016). La Investigación Basada en Diseño en Tecnología Educativa. *Revista Interuniversitaria de Investigación en Tecnología Educativa*. <https://doi.org/10.6018/riite2016/260631>

Dean, J. (s. f.). *Big Data, Data Mining, and Machine Learning*. 289.

Easterday, M. W., Rees Lewis, D. G., & Gerber, E. M. (2018). The logic of design

research. *Learning: Research and Practice*, 4(2), 131-160.

<https://doi.org/10.1080/23735082.2017.1286367>

Ecured. (2019, julio). *PowerDesigner—EcuRed*.

<https://www.ecured.cu/PowerDesigner>

Fontaines, T., Maza, J., & Pirela, J. (2020). Tendencias en investigación (Primera).

<https://d1wqtxts1xzle7.cloudfront.net/63601649/Tendencias20200611-22099>

[ok1e25.pdf?1591914361=&response-content](https://d1wqtxts1xzle7.cloudfront.net/63601649/Tendencias20200611-22099-ok1e25.pdf?1591914361=&response-content)

[disposition=inline%3B+filename%3DLa_matriz_de_consistencia_en_el_proyecto.pdf&Expires=1598079999&Signature=RUYdWGG8gRVP546yyMIQXH-](https://d1wqtxts1xzle7.cloudfront.net/63601649/Tendencias20200611-22099-ok1e25.pdf?1591914361=&response-content-disposition=inline%3B+filename%3DLa_matriz_de_consistencia_en_el_proyecto.pdf&Expires=1598079999&Signature=RUYdWGG8gRVP546yyMIQXH-)

[xpires=1598079999&Signature=RUYdWGG8gRVP546yyMIQXH-](https://d1wqtxts1xzle7.cloudfront.net/63601649/Tendencias20200611-22099-ok1e25.pdf?1591914361=&response-content-disposition=inline%3B+filename%3DLa_matriz_de_consistencia_en_el_proyecto.pdf&Expires=1598079999&Signature=RUYdWGG8gRVP546yyMIQXH-)

[U12cA2AA71p076SSnNxbsJVcH3FKJ7--N8gyyz5qtijB2rFIPfkLDAqCyS6Om-](https://d1wqtxts1xzle7.cloudfront.net/63601649/Tendencias20200611-22099-ok1e25.pdf?1591914361=&response-content-disposition=inline%3B+filename%3DLa_matriz_de_consistencia_en_el_proyecto.pdf&Expires=1598079999&Signature=RUYdWGG8gRVP546yyMIQXH-)

[mW4bUUKIRGaBCdn5r98n968vhBf2ioAOJ3DuExhPLxUwwaup0IC7mJ6NiTiDNmeOCn](https://d1wqtxts1xzle7.cloudfront.net/63601649/Tendencias20200611-22099-ok1e25.pdf?1591914361=&response-content-disposition=inline%3B+filename%3DLa_matriz_de_consistencia_en_el_proyecto.pdf&Expires=1598079999&Signature=RUYdWGG8gRVP546yyMIQXH-)

[Y6q3IaQlaGNicc2BT-](https://d1wqtxts1xzle7.cloudfront.net/63601649/Tendencias20200611-22099-ok1e25.pdf?1591914361=&response-content-disposition=inline%3B+filename%3DLa_matriz_de_consistencia_en_el_proyecto.pdf&Expires=1598079999&Signature=RUYdWGG8gRVP546yyMIQXH-)

[ySZkHjgZKMHFp4PiJI4Ve7I3xwm87S0~6FKeNp0VVCiY9ybwOosQOY-](https://d1wqtxts1xzle7.cloudfront.net/63601649/Tendencias20200611-22099-ok1e25.pdf?1591914361=&response-content-disposition=inline%3B+filename%3DLa_matriz_de_consistencia_en_el_proyecto.pdf&Expires=1598079999&Signature=RUYdWGG8gRVP546yyMIQXH-)

[CFWj6pnuF~cL~M2I6zOP-ZcFE-](https://d1wqtxts1xzle7.cloudfront.net/63601649/Tendencias20200611-22099-ok1e25.pdf?1591914361=&response-content-disposition=inline%3B+filename%3DLa_matriz_de_consistencia_en_el_proyecto.pdf&Expires=1598079999&Signature=RUYdWGG8gRVP546yyMIQXH-)

[3wwJq63q5vk~SIRdb1PwVB6Q86pQxE9AR18LTVFtTVBvOThURaiTkTaYw8UYujWOP](https://d1wqtxts1xzle7.cloudfront.net/63601649/Tendencias20200611-22099-ok1e25.pdf?1591914361=&response-content-disposition=inline%3B+filename%3DLa_matriz_de_consistencia_en_el_proyecto.pdf&Expires=1598079999&Signature=RUYdWGG8gRVP546yyMIQXH-)

[YB4MIHTtCpiEMPej7GsmDJu0FIA__&Key-Pair-](https://d1wqtxts1xzle7.cloudfront.net/63601649/Tendencias20200611-22099-ok1e25.pdf?1591914361=&response-content-disposition=inline%3B+filename%3DLa_matriz_de_consistencia_en_el_proyecto.pdf&Expires=1598079999&Signature=RUYdWGG8gRVP546yyMIQXH-)

[Id=APKAJLOHF5GGSLRBV4ZA#page=31](https://d1wqtxts1xzle7.cloudfront.net/63601649/Tendencias20200611-22099-ok1e25.pdf?1591914361=&response-content-disposition=inline%3B+filename%3DLa_matriz_de_consistencia_en_el_proyecto.pdf&Expires=1598079999&Signature=RUYdWGG8gRVP546yyMIQXH-)

Gallagher, C., Madden, M. G., & D'Arcy, B. (2015). *A Bayesian Classification Approach to Improving Performance for a Real-World Sales Forecasting Application*. 475-480. <https://doi.org/10.1109/ICMLA.2015.150>

García-Calvo Fernández, J. A. (2020, febrero 24). *Business Intelligence: Cuadrante mágico de Gartner 2020*. <https://www.inforges.es/post/cuadrante-magico-de-gartner-2020-business-intelligence>

Gartner. (2020, febrero). *Gartner Reprint*.

<https://www.gartner.com/doc/reprints?id=1->

1YCTPMUL&ct=200213&st=sb&utm_medium=Website+

Gil, J. (2020, febrero 5). Tableau: La importancia del análisis de datos. *Blog de Hiberus Tecnología*. <https://www.hiberus.com/crecemos-contigo/tableau-la-importancia-del-analisis-de-datos/>

Gobi, N., & Rathinavelu, A. (2018). Analyzing cloud based reviews for product ranking using feature based clustering algorithm. *Cluster Computing*. <https://doi.org/10.1007/s10586-018-1996-3>

González, R. A., & Quimbaya, A. P. (2012). *LA INVESTIGACIÓN CIENTÍFICA BASADA EN EL DISEÑO COMO EJE DE PROYECTOS DE INVESTIGACIÓN EN INGENIERÍA*. 10.

Gonzalez, R., & Pomares Quimbaya, A. (2012). *La investigación científica basada en el diseño como eje de proyectos de investigación en ingeniería*.

Gorakala, S. K., & Usuelli, M. (2015). *Building a Recommendation System with R*. Packt Publishing.

Greg, D. (2018). *Los mejores 10 algoritmos en minería de datos*. Techlandia. https://techlandia.com/mejores-10-algoritmos-mineria-datos-info_295108/

Kelchner Luanne. (2018, febrero 1). *Descripciones del trabajo de un gerente de manufactura | Cuida tu dinero* [Descripciones del trabajo de un gerente de manufactura]. Cuida tu dinero. <https://www.cuidatudinero.com/13100722/descripciones-del-trabajo-de-un-gerente-de-manufactura>

Manasson, A. (2020, febrero 9). *A Not So Secret Way Of Becoming A Better Data Scientist*. Medium. <https://towardsdatascience.com/why-using-crisp-dm-will-make-you-a-better-data-scientist-66efe5b72686>

Masip, D. (2002, julio 19). *Qué es Oracle*. <https://desarrolloweb.com/articulos/840.php>

Mendoza Mendoza, J. A., Calambás Marín, D. A., Malaver Moreno, J. A., & Ruiz Rodríguez, L. A. (2019). *Portal web para la consulta y análisis de datos abiertos relacionados con la contratación pública: ADACOP*. <http://repository.javeriana.edu.co/handle/10554/43716>

Neteris. (2019, julio 1). *Tableau Software—Herramienta de Visualización de Datos*. Neteris. <https://neteris.com/software/tableau-software-visualizacion-datos>

Oracle. (2020). *¿Está preparado para unificar sus bases de datos en nubes de bases de datos?* <https://www.oracle.com/es/database/>

Pagano, A., & Buitron, V. (2009). *Reorganización de las trayectorias escolares de los alumnos con sobreedad en el nivel primario: Ciudad de Buenos Aires (Argentina): programas de aceleración*. Fundación Iberoamericana para la Educación, la Ciencia y la Cultura.

Palacios, H. J. G., Toledo, R. A. J., Pantoja, G. A. H., & Navarro, Á. A. M. (2017). A comparative between CRISP-DM and SEMMA through the construction of a MODIS repository for studies of land use and cover change. *Advances in Science, Technology and Engineering Systems Journal*, 2(3), 598-604. <https://doi.org/10.25046/aj020376>

Ramirez, J. (2019). *SISTEMA EXPERTO BASADO EN ONTOLOGÍA PARA LA DETECCIÓN DE FRAUDE EN TARJETAS DE CRÉDITO*. <https://repository.javeriana.edu.co/bitstream/handle/10554/46120/Documento.pdf?sequence=1&isAllowed=y>

Romero, J. (2019, junio 11). *Aprende la Técnicas y algoritmos de Minería de Datos actuales. Posicionamiento Web, más tráfico, más clientes*. <https://jorgeromero.net/tecnicas-y-algoritmos-de-mineria-de-datos/>

Russo, M. (2020, febrero). *Comparativa de plataformas Business Intelligence y análisis de datos: Principales herramientas*.

<https://www.paradigmadigital.com/techbiz/comparativa-plataformas-business-intelligence-principales-herramientas/>

Sánchez, R. C., Nahón, A. E., & Ramírez, M. T. G. (s. f.). *APLICACIÓN DEL MÉTODO DE INVESTIGACIÓN BASADA EN EL DISEÑO DE LA CREACIÓN DEL CENTRO DE INVESTIGACIÓN EN INNOVACIÓN Y TECNOLOGÍA EDUCATIVA*. 13.

Senplades. (2012). *Transformación de la matriz productiva*. SENPLADES. http://www.planificacion.gob.ec/wp-content/uploads/downloads/2013/01/matriz_productiva_WEBtodo.pdf

Usmani, Z. A., Manchekar, S., Malim, T., & Mir, A. (2017). A predictive approach for improving the sales of products in e-commerce. *Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2017 Third International Conference on*, 188–192.