



Universidad Internacional de La Rioja
Escuela Superior de Ingeniería y Tecnología

Maestría en Inteligencia Artificial

**Nueva Metodología CSKT para mejorar
los proyectos de implementación de
Machine Learning en Ingeniería
Industrial**

| | |
|---------------------------------------|-------------------------------|
| Trabajo de innovación presentado por: | • José Antonio Ogosi Auqui |
| Fecha: | Lunes 28 de setiembre de 2022 |

Índice de contenidos

| | | |
|------|--|-----|
| 1. | Introducción | 6 |
| 1.1. | Justificación del Trabajo | 7 |
| 1.2. | Estructura en secciones..... | 10 |
| 2. | Objetivos | 10 |
| 3. | Desarrollo conceptual | 12 |
| 3.1. | Empatizar..... | 23 |
| 3.2. | Definir..... | 33 |
| 3.3. | Investigación de antecedentes | 36 |
| 3.4. | Idear | 39 |
| 3.5. | Prototipar | 43 |
| 3.6. | Selección de prototipos y criterio para hacerlo (evaluar)..... | 50 |
| 4. | Metodología | 52 |
| 4.1. | Scrum..... | 52 |
| 4.2. | Producto Mínimo Viable (MVP) | 65 |
| 5. | Implementación de la propuesta | 68 |
| 5.1. | Planificación y estimación | 68 |
| 5.2. | Implementación | 69 |
| 5.3. | Despliegue | 88 |
| 5.4. | Mantenimiento | 96 |
| 6. | Validación y diseño experimental | 100 |
| 7. | Resultados y discusión | 117 |
| 7.1. | Metodología propuesta | 117 |

| | |
|---|-----|
| 7.2. Desarrollo de las encuestas..... | 118 |
| 8. Conclusiones y trabajo futuro | 119 |
| Referencias bibliográficas..... | 121 |
| Anexo A. Actualización del Sprint Burn Down Chart..... | 124 |
| Anexo B. Página web | 125 |
| Anexo C. Cronograma del proyecto de investigación | 126 |
| Anexo D. Evidencia de encuestas | 127 |
| Anexo E. Evidencia de video | 128 |

Índice de tablas

| | |
|--|----|
| Tabla 1 <i>Caso del Negocio</i> | 52 |
| Tabla 2 <i>Visión</i> | 53 |
| Tabla 3 <i>Involucrados del proyecto</i> | 54 |
| Tabla 4 <i>Product Backlog Item</i> | 55 |
| Tabla 5 <i>Puntos</i> | 59 |
| Tabla 6 <i>Tareas</i> | 60 |
| Tabla 7 <i>Burndown Chart</i> | 63 |
| Tabla 8 <i>Daily Scrum</i> | 63 |
| Tabla 9 <i>Cuadro de flujo para la implementación de la nueva metodología CSKT</i> | 65 |

Índice de figuras

| | |
|--|-----|
| Fig. 1. Factores de cambios en proyectos de Minería de datos | 8 |
| Fig. 2. Metodología CRISP-Dm | 11 |
| Fig. 3. Metodología SEMMA | 11 |
| Fig. 4. Metodología KDD | 12 |
| Fig. 5. Metodología TDSP | 12 |
| Fig. 6. Modelo del proceso CRISP-DM | 14 |
| Fig. 7. Proceso de KDD por Timarán-Pereira [11] | 17 |
| Fig. 8. Ciclo de vida de ciencia de datos [12] | 21 |
| Fig. 9. Burn Down Chart | 64 |
| Fig. 10. Diagrama de flujo para la implementación de la metodología CSKT | 67 |
| Fig. 11. Metodología CSKT | 68 |
| Fig. 12. Planificación del Proyecto - Metodología CSKT | 69 |
| Fig. 11. Modelo CANVAS de la metodología propuesta | 117 |
| Fig. 11. Grado de satisfacción | 118 |

1. Introducción

A través de los años el uso de la tecnología nos ha permitido poder recolectar grandes cantidades de datos que han permitido el desarrollo de la Minería de datos, disciplina que se define como el proceso de extraer conocimiento útil, comprensible y novedoso de grandes volúmenes de datos, siendo su principal objetivo encontrar información oculta o implícita, que no es posible obtener mediante métodos estadísticos convencionales [1]. <https://core.ac.uk/download/pdf/301040544.pdf>

Para las organizaciones la implementación de la minería de datos representa un reto de desarrollo e implementación el poder considerar todas las variables con las que cuenta un proyecto de Minería de datos y la manera en que se deben de trabajar en conjunto con el equipo desarrollador del proyecto para que su curso sea exitoso por lo que se han propuesto diversas metodologías que proporcionan una serie de pasos y tareas a realizar cuyo objetivo ha sido poder conceptualizar todas variables y poder cumplir con las metas del proyecto y la organización.

Según [2] una metodología hace referencia al conjunto de procedimientos basados en principios lógicos, utilizados para alcanzar una gama de objetivos que rigen una investigación científica. <https://www.ecured.cu/Metodolog%C3%ADa>. En términos de Minería de datos, las metodologías nos permiten una adecuada planeación que promueva resultados efectivos y en algunos casos de herramientas de software ayudan a que las metodologías puedan ser implementadas.

Cada día se deben aplicar miles de procesos de Machine Learning con las metodologías ya existentes como CRISP-DM, SEMMA, KDD y TDSP que más adelante serán detalladas en este documento. Todas estas metodologías son buenas para procesos basados en sistemas, sin embargo, existen otros campos de la ingeniería como la Ingeniería Industrial que se encuentra muy relacionada a la optimización de procesos y que podría resultar beneficiada al adecuar alguna metodología a sus necesidades. A través de la consulta de la página de Alicia, que es una base de datos que contiene tesis peruanas se notó que no hay trabajos de Machine Learning desarrollados por tesis de esta carrera a pesar de

poder aplicar sus conocimientos en el desarrollo de Machine Learning, por lo que al encontrar esta necesidad se plantea el desarrollo de este trabajo:

- El objetivo es diseñar una nueva metodología basada en CRISP-DM, SEMMA, KDD y TDSP para hacer posible la implementación de proyectos de mejora de procesos en Ingeniería de Industrial y que los próximos trabajos de tesis de estudiantes en Ingeniería de Industrial consideren el tema de Machine Learning para la optimización de procesos en los proyectos que consideren apropiados.
- El trabajo se desarrolla en 8 partes en las que se presenta el tema para que el lector comprenda la necesidad de la investigación, se define además los objetivos de la necesidad encontrada y luego se diseña el marco conceptual para comprender las diferentes vertientes teóricas de la naturaleza de propuesta de la investigación. Esta propuesta, una vez definida la metodología sobre la que se realizará la investigación, es guiada por la presentación propuesta, la cual se valida con la implementación y el diseño de la investigación, y esta investigación determina las conclusiones del diseño de la nueva metodología y finalmente se brindan las recomendaciones para trabajos futuros en Ingeniería Industrial.

1..1. Justificación del Trabajo

Las metas y objetivos del proyecto de minería de datos deben ser claros y específicos; debe ser completamente entendido por todos los participantes involucrados. Estas metas y objetivos también deben ser dinámicos ya que deben corresponder a las necesidades y requerimientos de la empresa, planta u organización. Los objetivos deben revisarse y actualizarse periódicamente y deben establecerse a intervalos consistentes con la perspectiva de la empresa u organización; de lo contrario, los datos, reglas, formularios e informes organizados y definidos por el proyecto quedarán desactualizados e inútiles. Hay presente la necesidad de objetivos claros pero dinámicos que se alineen con la evolución de productos, empresas, organizaciones y procesos. Nuevos mercados, nuevos clientes, nuevos productos y procesos pueden requerir diferentes datos, diferentes tareas o diferentes herramientas. Cualquier proyecto de minería de datos que intente lograr resultados útiles en estas circunstancias debe tener en cuenta los cambios (consulte la Figura 1) por Rohanzadeh [1].

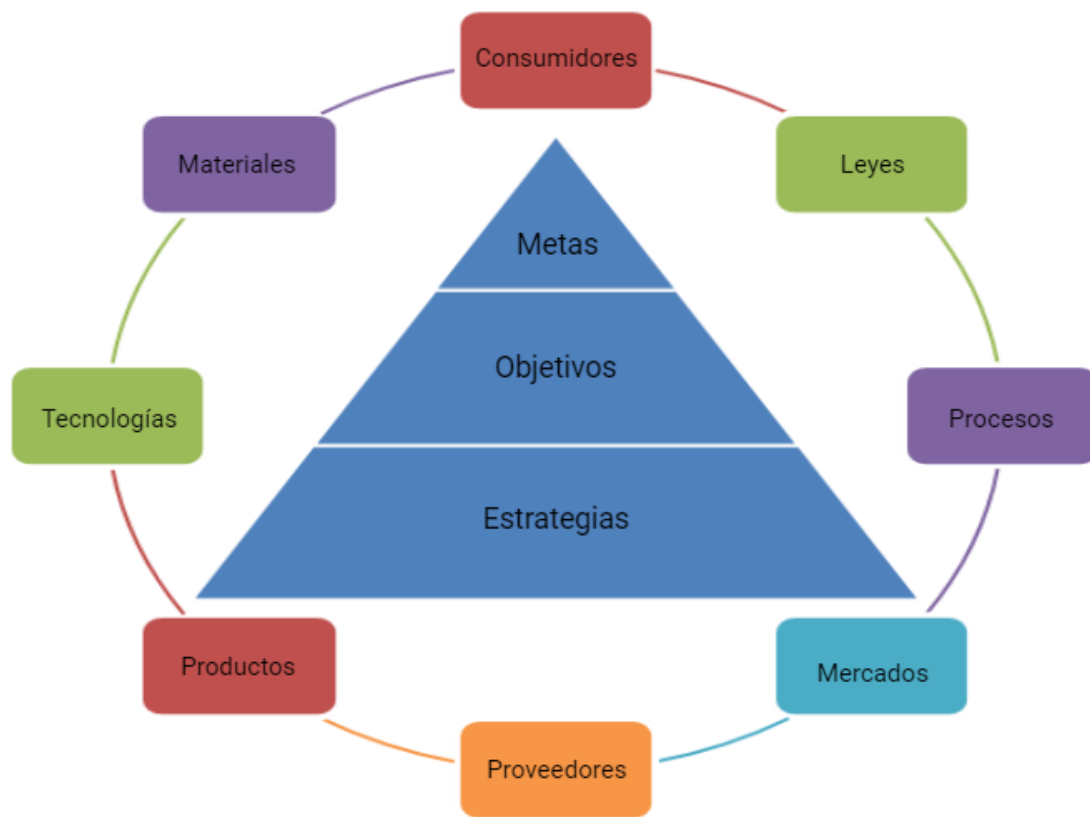


Fig. 1. Factores de cambios en proyectos de Minería de datos

Se han introducido métodos de minería de datos para proporcionar una visión más completa del proceso de descubrimiento de conocimiento, más allá de la aplicación de algoritmos estadísticos o aprendizaje automático. Su objetivo es definir un flujo de trabajo común, comenzando con la identificación de preguntas relacionadas, que conduce al procesamiento específico de datos sin procesar y, a menudo, no estructurados, y finalmente con el propósito de descubrir nueva información. Aunque los algoritmos de modelado son un componente importante de cualquier proceso de descubrimiento de conocimiento, su selección y aplicación consume solo una fracción del tiempo total involucrado. Por otro lado, la recopilación, limpieza y preparación manual de datos a menudo requiere una cantidad significativa de tiempo por Huber [2].

Hay tantas opciones, tareas, técnicas, herramientas, formatos y métodos para la minería de datos que a los ingenieros industriales les resulta muy difícil diseñar e implementar proyectos. Aunque las metodologías ya existen, están diseñadas para paquetes de software específicos. La mayoría de estos métodos utilizan métodos estadísticos

tradicionales. No está claro si este enfoque de minería de datos es suficiente para obtener la amplia gama de datos necesarios para las aplicaciones de ingeniería industrial. Por lo tanto, se necesita un método de minería de datos para cumplir con los requisitos específicos de la ingeniería industrial. Tal metodología ayudará a los ingenieros industriales a seleccionar herramientas de minería de datos apropiadas e implementar proyectos de minería de datos desde una perspectiva de sistemas por Solarte [3].

El hecho es que todos los ingenieros industriales pueden ser Data Scientist porque el estándar para dicha carrera es el dominio de las matemáticas y la estadística, pero hay un exceso control sobre las actividades a causa de que hace falta una formación en informática, lo que hace difícil que puedan aplicar Machine Learning para optimizar procesos sin conocer de:

- SQL
- R
- Python
- Tableau
- Power BI
- Excel avanzado

Sin estas habilidades no es posible que un aprender de las herramientas de un Data Scientist sin un poco de conocimiento en programación por Rodríguez [4].

Es imperativo tratar este problema porque el ingeniero industrial tiene potencial para diseñar procesos y sistemas que mejoren la calidad y la productividad, pero el ingeniero sin datos no es ingeniero solo se necesita aplicar los conceptos de las diferentes metodologías CRISP-DM, SEMMA, KDD y TDSP de acuerdo a las características de la disciplina y permitir que sea posible que los ingenieros industriales interactúen con otras disciplinas que combinan el conocimiento de los datos y la integración del conocimiento, esto último es algo que da buenas ideas para diseñar una metodología concerniente a los procesos en que se ocupa el ingeniero industrial como el diseño, mejora e instalación de sistemas integrados de personas, materiales, información, equipos y energía, aplicando Machine learning por Darwish [5].

1..2. Estructura en secciones

El trabajo se desarrolla en 8 partes en las que se presenta el tema para que el lector comprenda la necesidad de la investigación, se define además los objetivos de la necesidad encontrada y luego se diseña el marco conceptual para comprender las diferentes vertientes teóricas de la naturaleza de propuesta de la investigación. Esta propuesta, una vez definida la metodología sobre la que se realizará la investigación, es guiada por la presentación propuesta, la cual se valida con la implementación y el diseño de la investigación, y esta investigación determina las conclusiones del diseño de la nueva metodología y finalmente se brindan las recomendaciones para trabajos futuros en Ingeniería Industrial.

2. Objetivos

Para realizar este apartado, nos apoyamos en los principios MELDS.

Mindset: Fomentar una Mentalidad correcta.

Experimentation: Provomer la Experimentación.

Leadership: Apoyar el Liderazgo activamente implicado.

Data: Incorporar Datos a la estrategia.

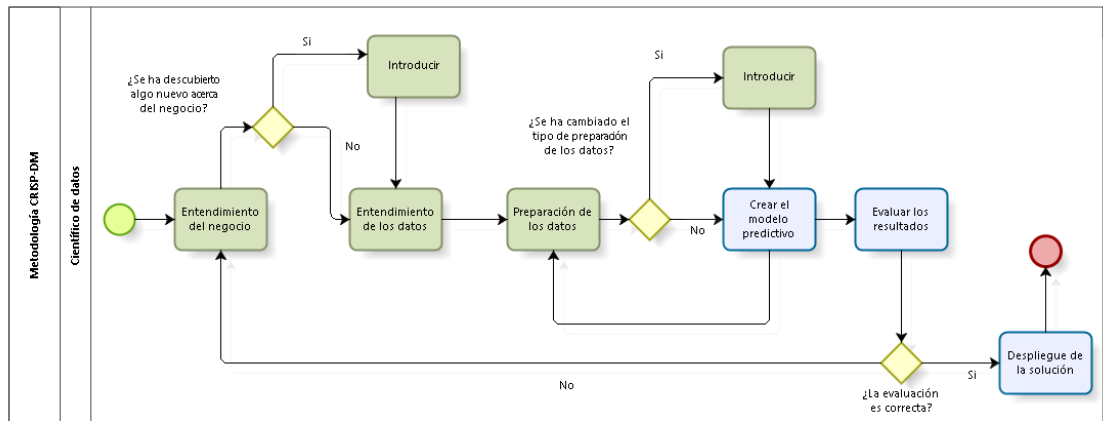
Skills: Cultivar nuevas Habilidades.

Tras aplicar los principios MELDS el Objetivo General es

- Diseñar y desarrollar una nueva metodología llamada CSKT que se **ajuste** a la implementación de proyectos de Machine Learning desarrollados en el campo de la Ingeniería Industrial.

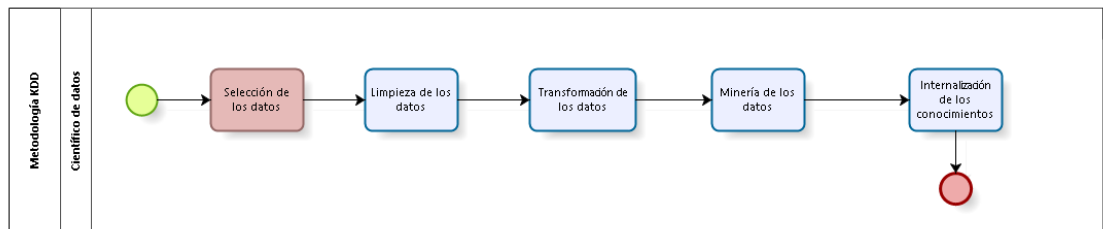
Para ello el despliegue de la metodología se mostrará en un diagrama de procesos en Bizagi, con las tareas que se van a adaptar a nuestra nueva metodología,

tomando solo aquellas tareas que forman parte de las metodologías ya existentes (ver Anexo 1) que son adaptables a las necesidades que tienen los proyectos de Ingeniería Industrial. Cada fase de nuestra nueva metodología tendrá sus especificaciones hasta llegar a la última fase, porque el fin, es que el científico de datos, al probar esta metodología, con los pasos claros de esta, logre implementar su Machine Learning a su proceso.



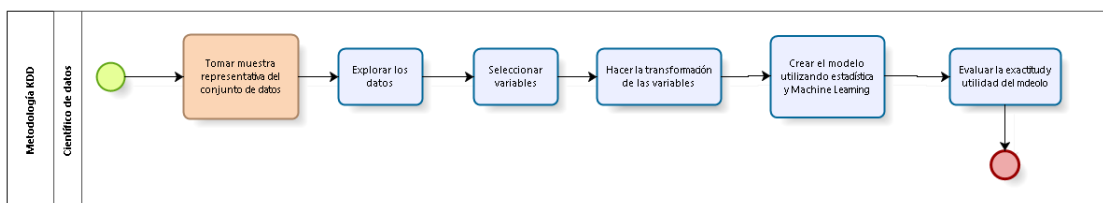
Powered by bizagi Modeler

Fig. 2. Metodología CRISP-Dm



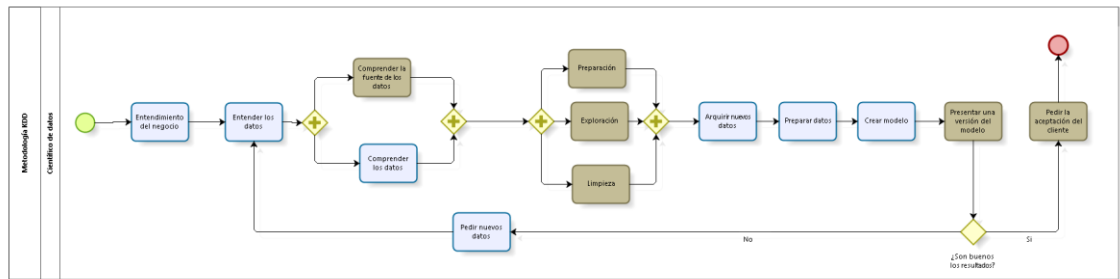
Powered by bizagi Modeler

Fig. 3. Metodología SEMMA



Powered by bizagi Modeler

Fig. 4. Metodología KDD



Presented by
bizagi
Innovation

Fig. 5. Metodología TDSP

La Propuesta se va a evaluar con un test de ANOVA realizado al sector de ingeniería industrial.

3. Desarrollo conceptual

Para realizar este apartado, nos venimos apoyando en el proceso Design Thinking.

- Empatizar
- Definir
- Investigación de antecedentes
- Idear
- Prototipar: Cual será el prototipo que vas a utilizar
- Selección de prototipos y criterio para hacerlo

Si se diseña la nueva metodología CSKT para Machine Learning, posiblemente mejorarán los proyectos de optimización de procesos en el área de Ingeniería Industrial; para confirmar esta hipótesis se ha revisado los antecedentes de solución más relevantes como el trabajo realizado por (1) los datos han de analizarse con la necesidad de un ritmo sin precedentes. Las tiendas también han formado parte de los patrones de comportamiento de los clientes, logrando avances en la construcción de robots autónomos y la predicción de fallas. Esto ha llevado a que se logren resultados precisos, dichos avances incluyen la detectabilidad de valores atípicos, el manejo de datos

faltantes, la formatización, integración y normalización de datos. Es concreto para los autores, eliminar ambigüedades e inconsistencias previo procesamiento, también para la minería de datos, es por ello que se logra proponer un marco llamado APREP-DM, también aplicable al análisis de datos y el uso de sensores de transmisión de datos. Prueba de este estudio es el poder analizar datos de sensores utilizando APREP-DM por Nagashima [6]. Otro punto inherente es la reingeniería de procesos para mejorar el proceso de negocio. Se pone énfasis en los marcos disponibles, también en la automatización de procesos comerciales. Esto indica que se han realizado esfuerzos para aplicar el aprendizaje automático en la gestión y mejora de los procesos comerciales en general. En concreto los autores se ocupan del descubrimiento de procesos, la predicción del comportamiento de procesos, la mejora de procesos y la mejora de proyectos. La investigación muestra que la ampliación de las aplicaciones para la reingeniería de procesos comerciales es muy prometedora. El estudio propuso un modelo de aprendizaje automático para automatizar la reingeniería de procesos comerciales, inspirado en los principios de Lean Six Sigma para eliminar el desperdicio y la divergencia en los procesos comerciales. Esta aplicación es recomendada especialmente para identificar las características de aprendizaje automático por Al-Anqoudi [7]. En el mismo sentido se reconoce como el sector de Ingeniería industrial ha ido avanzando en su desarrollo, especialmente ante sistemas en las industrias manufactureras más evolucionadas, pero todavía están en albores en los sectores industriales tradicionales, como los paneles de madera, la fabricación de láminas, el procesamiento de pulpa y papel y la producción de paneles compuestos. Para ello se recurre al marco PREFAB para minimizar los productos rechazando utilizando inteligencia artificial. Este marco también se construyó por medio de la investigación en acción de una fábrica de paneles a base de madera. Todo este marco es entregado para proporcionar soporte a las decisiones que se han producido en el taller y para respaldar la ingeniería de procesos/productos por Dias [8].

A partir de este análisis, para diseñar la solución basada en las metodologías CRISP-DM, SEMMA, KDD y TDSP nos apoyamos del proceso Design Thinking.

- **CRISP – DM**

- o CRISP-DM son las siglas de Cross-Industry Standard Process for Data Mining que consiste en una metodología para el desarrollo de proyectos de Minado de Datos (DM). Dicha metodología se realizó para catalogar y guiar los pasos más comunes en el desarrollo de proyectos de Minado de Datos.
- o Su primera versión se publicó en 1999, desarrollada por un equipo de experimentados ingenieros en Minado de Datos, dando como resultado una guía de 6 fases divididas en 4 niveles de abstracción organizados de forma jerárquica en tareas y considerada como el estándar para el desarrollo de proyectos y minado de datos, hasta la fecha es la metodología mayormente utilizada.

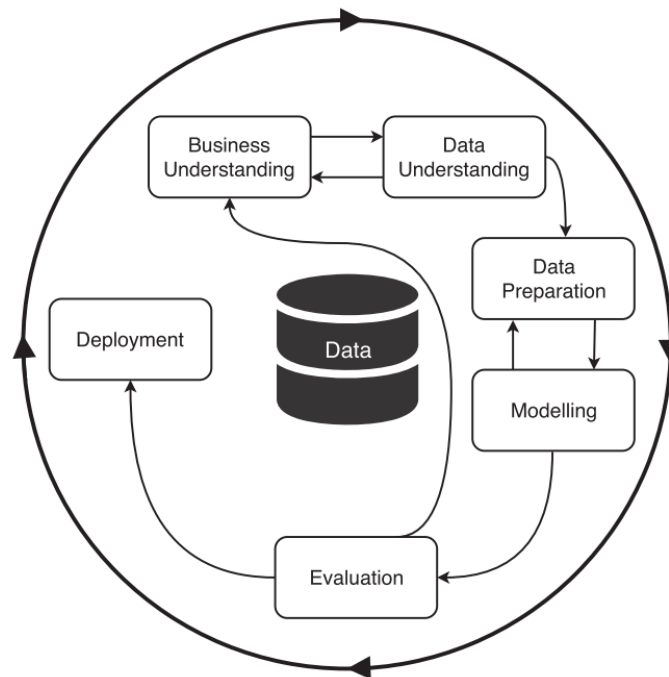


Fig. 6. Modelo del proceso CRISP-DM

Las fases que la constituyen son las siguientes:

1. Comprensión del negocio

Esta fase tiene como premisa la pregunta ¿Qué necesita el negocio? Y se basa en entender los objetivos y los

requerimientos del proyecto con el fin de definir el problema a resolver y diseñar la planificación preliminar para que los objetivos se cumplan.

2. Comprensión de los datos

Después, se comienza con la recolección inicial de los datos con el propósito de familiarizarse con los mismos, identificando problemas de calidad asociados a ellos e información adicional relevante para la formulación de las primeras hipótesis.

3. Preparación de los datos

La fase de preparación de los datos abarca todas aquellas actividades destinadas a la construcción del conjunto de datos finales. Las tareas de esta fase pueden ser ejecutadas varias veces, sin un orden predefinido. Las mismas incluyen la selección de tablas, registros y atributos, así como también la transformación y limpieza de datos para que puedan ser tratados por las herramientas de modelado.

4. Modelado

En esta fase, se seleccionan y aplican las técnicas de modelado más apropiadas para el proyecto en cuestión, calibrando sus parámetros a valores óptimos. Básicamente, existen varias técnicas para un mismo tipo de problemas en proyectos de Explotación de Datos. Algunas de ellas, demandan requerimientos específicos sobre los datos que se van a procesar, por tal motivo muchas veces es necesario volver a la fase de preparación de los datos antes de avanzar con el modelado de los mismos.

5. Evaluación

Esta fase, involucra la evaluación del modelo y revisión de los pasos ejecutados en relación a los objetivos del negocio, y busca determinar si hay alguna razón de negocio para el cual el modelo es deficiente, asegurándonos de esta forma, alcanzar los objetivos inicialmente propuestos. Al final de esta fase, se debe tener una decisión sobre el uso de los resultados alcanzados.

6. Implementación o despliegue

La fase de despliegue o implementación, dependiendo de los requisitos del proyecto, puede ser tan sencilla como la generación de un simple reporte o tan compleja como la implementación de un proceso de Explotación de Datos repetible en toda la empresa. En determinadas ocasiones, el propio cliente es quien lleva a cabo los pasos concretos de la implementación y no el propio analista de datos; lo cual permite al cliente conocer de manera anticipada qué acciones son requeridas con el fin de hacer uso del modelo creado.

- **KDD**

El Descubrimiento de conocimiento en bases de datos (KDD por sus siglas en inglés Knowledge Discovery in Databases) consiste en un proceso automático que a su vez es interactivo e iterativo que consiste en extraer patrones en forma de reglas o funciones, a partir de datos, para que el usuario los analice.

El proceso KDD se resume en las siguientes etapas:

- Selección
- Preprocesamiento/limpieza
- Transformación/reducción
- Minería de datos (data mining)
- Interpretación/ evaluación

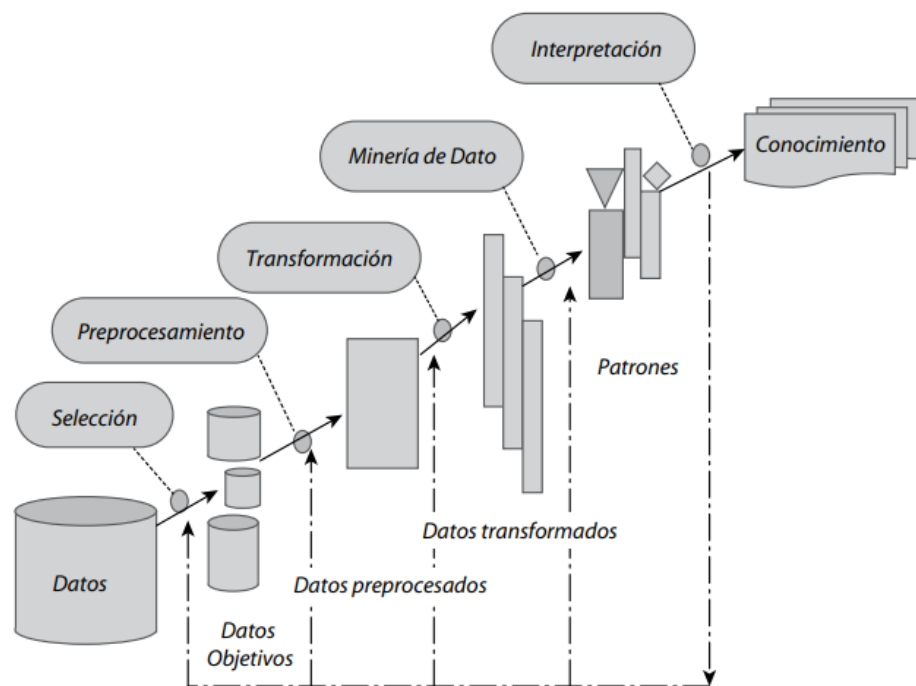


Fig. 7. Proceso de KDD por Timaran-Pereira [11]

- **Etapa de seleccion**

En la etapa de seleccion, una vez identificado el conocimiento relevante y prioritario y definidas las metas del proceso kdd, desde el punto de vista del usuario final, se crea un conjunto de datos objetivo, seleccionando todo el conjunto de datos o una muestra representativa de este, sobre el cual se realiza el proceso de descubrimiento. La seleccion de los datos vara de acuerdo con los objetivos del negocio. Etapa de preprocesamiento/limpieza En la etapa de preprocesamiento/limpieza (data cleaning) se analiza la calidad de los datos, se aplican operaciones basicas como la remocion de datos ruidosos, se seleccionan estrategias para el manejo de datos desconocidos (missing y empty), datos nulos, datos duplicados y tecnicas estadsticas para su reemplazo. En esta etapa, es de suma importancia la interaccion con el

usuario o analista. Los datos ruidosos (noisy data) son valores que están significativamente fuera del rango de valores esperados; se deben principalmente a errores humanos, a cambios en el sistema, a información no disponible a tiempo y a fuentes heterogéneas de datos. Los datos desconocidos empty son aquellos a los cuales no les corresponde un valor en el mundo real y los missing son aquellos que tienen un valor que no fue capturado. Los datos nulos son datos desconocidos que son permitidos por los sistemas gestores de bases de datos relacionales (sgbdr). En el proceso de limpieza todos estos valores se ignoran, se reemplazan por un valor por omisión, o por el valor más cercano, es decir, se usan métricas de tipo estadístico como media, moda, mínimo y máximo para reemplazarlos.

- **Etapas de transformación/reducción**

En la etapa de transformación/reducción de datos, se buscan características útiles para representar los datos dependiendo de la meta del proceso. Se utilizan métodos de reducción de dimensiones o de transformación para disminuir el número efectivo de variables bajo consideración o para encontrar representaciones invariantes de los datos (Fayyad et al., 1996). Los métodos de reducción de dimensiones pueden simplificar una tabla de una base de datos horizontal o verticalmente. La reducción horizontal implica la eliminación de tuplas idénticas como producto de la sustitución del valor de un atributo por otro de alto nivel, en una jerarquía definida de valores categóricos o por la discretización de valores continuos (por ejemplo, edad por un rango de edades). La reducción vertical implica la eliminación de atributos que son insignificantes o redundantes con respecto al problema, como la eliminación de

llaves, la eliminación de columnas que dependen funcionalmente (por ejemplo, edad y fecha de nacimiento). Se utilizan técnicas de reducción como agregaciones, compresión de datos, histogramas, segmentación, discretización basada en entropía, muestreo, entre otras (Han y Kamber, 2001).

Etapa de minería de datos El objetivo de la etapa minería de datos es la búsqueda y descubrimiento de patrones insospechados y de interés, aplicando tareas de descubrimiento como clasificación (Quinlan, 1986) (Wang, Iyer y Scott, 1998), clustering (Ng y Han, 1994), (Zhang, Ramakrishnan, Livny, 1996), patrones secuenciales (Agrawal y Srikant, 1995) y asociaciones (Agrawal y Srikant, 1994), (Srikant y Agrawal, 1996), entre otras. Las técnicas de minería de datos crean modelos que son predictivos o descriptivos. Los modelos predictivos pretenden estimar valores futuros o desconocidos de variables de interés, que se denominan variables objetivo, dependientes o clases, usando otras variables denominadas independientes o predictivas, como por ejemplo predecir para nuevos clientes si son buenos o malos basados en su estado civil, edad, género y profesión, o determinar para nuevos estudiantes si desertan o no en función de su zona de procedencia, facultad, estrato, género, edad y promedio de notas. Entre las tareas predictivas están la clasificación y la regresión. Los modelos descriptivos identifican patrones que explican o resumen los datos; sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos, como identificar grupos de personas con gustos similares o identificar patrones de compra de clientes en una determinada zona de la ciudad. Entre las tareas descriptivas se cuentan las reglas de asociación, los patrones secuenciales, los clustering y las correlaciones. Por lo

tanto, la escogencia de un algoritmo de minería de datos incluye la selección de los métodos por aplicar en la búsqueda de patrones en los datos, así como la decisión sobre los modelos y los parámetros más apropiados, dependiendo del tipo de datos (categóricos, numéricos) por utilizar.

- **Etapas de interpretación/evaluación de datos**

En la etapa de interpretación/evaluación, se interpretan los patrones descubiertos y posiblemente se retorna a las anteriores etapas para posteriores iteraciones. Esta etapa puede incluir la visualización de los patrones extraídos, la remoción de los patrones redundantes o irrelevantes y la traducción de los patrones útiles en términos que sean entendibles para el usuario. Por otra parte, se consolida el conocimiento descubierto para incorporarlo en otro sistema para posteriores acciones o, simplemente, para documentarlo y reportarlo a las partes interesadas; también para verificar y resolver conflictos potenciales con el conocimiento previamente descubierto.

- **TDSP**

El proceso de ciencia de datos en equipo (TDSP por sus siglas en inglés, Team Data Science Process) es una metodología de ciencia de datos ágil e iterativa, desarrollada por Microsoft para proporcionar soluciones de análisis predictivo y aplicaciones inteligentes de manera eficiente por Timarán-Pereira [11].

- **Componentes principales del TDSP**

- Una definición de ciclo de vida de la ciencia de datos
- Una estructura de proyecto estandarizada
- Infraestructura y recursos recomendados para proyectos de ciencia de datos

- Herramientas y utilidades recomendadas para la ejecución de proyectos
 - **Ciclo de vida de ciencia de datos**

Se proporciona un ciclo de vida para estructurar el desarrollo de los proyectos de ciencia de datos y se describen todos los pasos que siguen los proyectos.

Este ciclo de vida ha sido diseñado para proyectos de ciencia de datos que se enviarán como parte de aplicaciones inteligentes las cuales implementan modelos de aprendizaje o inteligencia artificial de máquina para realizar un análisis predictivo.

De forma iterativa se presentan las principales fases por las que pasa un proyecto:

 - Conocimiento del negocio
 - Adquisición y comprensión de los datos
 - Modelado
 - Implementación

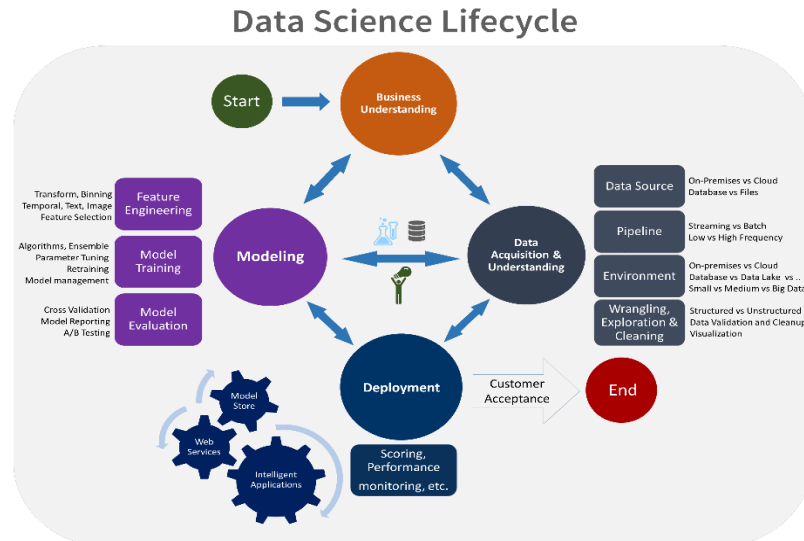


Fig. 8. Ciclo de vida de ciencia de datos [12]

Dentro del Ciclo de vida del proceso de ciencia de datos en equipo se describen los objetivos, las tareas y los artefactos de documentación

de cada fase del ciclo de vida de TDSP. Estas tareas y artefactos están asociados con roles de proyecto:

- Arquitecto de soluciones
- Jefe de proyecto
- Ingeniero de datos
- Científico de datos
- Desarrollador de aplicaciones
- Responsable de Proyecto

○ **Estructura del proyecto estandarizada**

Cuando todos los proyectos comparten una estructura de directorio y usan plantillas para los documentos de proyecto, resulta fácil para los miembros del equipo encontrar información sobre sus proyectos.

Consiste en la búsqueda de una estructura de directorio que use plantillas para todos los documentos del proyecto, con el fin de encontrar la información de los proyectos.

○ **Infraestructura y recursos para los proyectos de ciencia de datos**

TDSP proporciona recomendaciones para administrar análisis compartido e infraestructura de almacenamiento, por ejemplo:

- sistemas de archivos en la nube para almacenar conjuntos de datos
- databases
- clústeres de macrodatos (SQL o Spark)
- servicio de aprendizaje automático

○ **Herramientas y utilidades para la ejecución de proyectos**

En la mayoría de las organizaciones la introducción de procesos presenta ciertos desafíos. Las herramientas proporcionadas para implementar el proceso y el ciclo de vida de ciencia de datos ayudan a reducir las barreras a su adopción y la normalizan. TDSP proporciona un conjunto inicial de herramientas y scripts para impulsar la adopción de TDSP dentro de un equipo. También ayuda a automatizar algunas de las tareas comunes del ciclo de vida de ciencia de datos, como la exploración de datos y el modelado de línea de base. Existe una estructura bien definida que se proporciona a los individuos para que contribuyan con herramientas y utilidades compartidas al repositorio de código compartido de su equipo. Estos recursos se pueden aprovechar luego en otros proyectos dentro del equipo o en la organización.

3..1. Empatizar

Vamos a poner en práctica lo planificado que es una ficha de observación básica a las tesis de alumnos de la carrera de Ingeniería Industrial, a continuación, la siguiente ficha permitirá orientar en el recojo de información y aspectos que se consideren importantes para la siguiente fase.

FICHA DE OBSERVACIÓN BÁSICA N° 1

- I. Datos generales:
 - 1.1. Título de tesis: MODELO DE PRONÓSTICO DE DEMANDA DE EFECTIVO PARA LAS OFICINAS DE UNA ENTIDAD BANCARIA EN UNA CIUDAD DEL INTERIOR DEL PAIS
 - 1.2. Grado académico: Ingeniería Industrial
 - 1.3. Autor: Anggela del Rosario Zavala Jacinto
 - 1.4. Fecha: Julio del 2013
 - 1.5. Institución académica: Pontifica Universidad Católica del Perú

1.6. Liga: <http://hdl.handle.net/20.500.12404/4816>

II. Aspectos:

| Pregunta | Respuesta |
|--|---|
| ¿Función de proceso, por ejemplo, recursos humanos, finanzas, cadena de suministro y producción? | Pronóstico de demanda de efectivo |
| ¿Procesa versiones de la historia; la versión del proceso abre otra dimensión del modelo de predicción del diseño del proceso alternativo.? | Sí, el modelo es aplicado para que se adapte al modelo centralizado que tiene el banco, dando como un nuevo proceso mejorado que puede ser aplicado en las sucursales que se encuentren al interior del país, con motivo de ser replicado y así el banco pueda seguir extendiéndose alrededor del país. |
| ¿Desempeña la actividad contra sus KPIs: malo, bueno O excelente. Este atributo puede ser más preciso cuando los datos de KPI están disponibles? | Bueno, cumple con KPI, sin embargo es necesario la aplicación del pronóstico para conocer qué tan exitoso será. |
| ¿Valor del costo de operación de la actividad (¿aceptable? Sí o no). identificar las fuentes de datos y los conjuntos de datos requeridos. Ayuda a identificar la siguiente lista de atributos (características de | Sí, los datos necesarios de entrada para la red neuronal se tienen a través del análisis de datos del banco y los datos que formaron parte para la construcción del modelo: fecha, |

| | | |
|---|---|--|
| <p>aprendizaje automático) como entradas para el modelo de aprendizaje automático?</p> | <p>día de la semana, mes del año, montos de efectivo para el abastecimiento de los cajeros. Estos datos determinaron el requerimiento de efectivo de cada oficina.</p> | |
| <p>¿Nivel de satisfacción del cliente (de uno a cinco); implementa la satisfacción del cliente en cada paso.?</p> | <p>El modelo pronóstico que se propone tiene un porcentaje de error promedio de 6.82% para las oficinas mixtas y de un 8.02% para las oficinas captadoras, lo cual al ser aplicado el modelo debería generar una satisfacción al cliente de 5, sin embargo, no se tienen datos que corroboren esta información.</p> | |
| <p>¿Hace un cuello de botella? (Si o no).</p> | <p>No</p> | |
| <p>¿Hay suficientes recursos para realizar la tarea? (Si o no).</p> | <p>Sí</p> | |
| <p>¿Está digitalizado? (Si o no).</p> | <p>No</p> | |
| <p>¿Se puede automatizar? (Si o no).</p> | <p>Sí</p> | |
| <p>¿Es un paso de verificación del cumplimiento de la auditoría? (Si o no).</p> | <p>No</p> | |

| | |
|--|---|
| <p>¿Está siendo realizado por el mismo actor que el paso anterior en el proceso? (Si o no).</p> | <p>No</p> |
| <p>¿Valor de la actividad para el proceso (1–10)? Si es menos de tres, ¿debería eliminarse?</p> | <p>10, en este caso la actividad está fuertemente ligada con el objetivo del banco para la atracción y preferencia de los clientes por lo que el aseguramiento que brinda la aplicación de esta actividad (la disposición de dinero en efectivo para todos los trámites que se realicen) garantiza la preferencia de clientes o atracción de posibles clientes.</p> |
| <p>¿Cuán relevante es la actividad para los objetivos del proceso? (Relevante o irrelevante).</p> | <p>Relevante</p> |
| <p>¿Es un servicio prestado a los clientes de paso? (Si o no).</p> | <p>No</p> |
| <p>¿Es una actividad interna o requiere aportes externos? Interno se entiende en el sentido de la organización, específicamente, el departamento de funciones.</p> | <p>Actividad interna</p> |

| | |
|--|--|
| Nivel de partes interesadas. (VIP, propietario, consumidor, empleado y sistema). | Banco, empleados de este y clientes (con el fin de generar la preferencia del banco entre sus clientes). |
| ¿Es de empresa a empresa o de empresa a cliente? | De empresa a empresa |

FICHA DE OBSERVACIÓN BÁSICA N° 2

III. Datos generales:

- 3.1. Título de tesis: Desarrollo de un modelo predictivo para un sistema de manufactura orientado a la industria 4.0
- 3.2. Grado académico: Maestría en Ingeniería Industrial
- 3.3. Autor: Ing. Lina María Zúñiga Quiñones
- 3.4. Fecha: 2021
- 3.5. Institución académica: Pontificia Universidad Javeriana de Colombia
- 3.6. Liga: <https://repository.javeriana.edu.co/handle/10554/55631>

IV. Aspectos:

| Pregunta | Respuesta |
|--|---|
| ¿Función de proceso, por ejemplo, recursos humanos, finanzas, cadena de suministro y producción? | Producción |
| ¿Procesa versiones de la historia; la versión del proceso abre otra | El modelo es predictivo y el autor considera que deben realizarse |

| | |
|--|---|
| <p>dimensión del modelo de predicción del diseño del proceso alternativo.?</p> | <p>más consideraciones para poder aplicar las variaciones que podrían surgir en el modelo, es decir, que cambios habría si se elabora otro producto.</p> |
| <p>¿Desempeña la actividad contra sus KPIs: malo, bueno O excelente. Este atributo puede ser más preciso cuando los datos de KPI están disponibles?</p> | <p>Los KPI's no se encuentran disponibles, sin embargo en la literatura que revisó el autor considera los KPI's dentro de la predicción analizada.</p> |
| <p>¿Valor del costo de operación de la actividad (¿aceptable? Sí o no). identificar las fuentes de datos y los conjuntos de datos requeridos. Ayuda a identificar la siguiente lista de atributos (características de aprendizaje automático) como entradas para el modelo de aprendizaje automático?</p> | <p>No se mencionan costos de operación, sin embargo se muestra cómo es que fue implementado el sistema para monitoreo predictivo.</p> |
| <p>¿Nivel de satisfacción del cliente (de uno a cinco); implementa la satisfacción del cliente en cada paso.?</p> | <p>Se desconoce, el modelo propuesto indica que la predicción realizada ayuda a que la persona que va a manufacturar el producto puede darse una idea del tiempo que se tomaría en realizar el proceso y el material necesario.</p> |
| <p>¿Hace un cuello de botella? (Si o no).</p> | <p>No</p> |

| | |
|--|--|
| ¿Hay suficientes recursos para realizar la tarea? (Si o no). | Sí, fue implementado el sistema. |
| ¿Está digitalizado? (Si o no). | Sí |
| ¿Se puede automatizar? (Si o no). | Sí |
| ¿Es un paso de verificación del cumplimiento de la auditoría? (Si o no). | No, no se menciona alguna auditoría. |
| ¿Está siendo realizado por el mismo actor que el paso anterior en el proceso? (Si o no). | No |
| ¿Valor de la actividad para el proceso (1–10)? Si es menos de tres, ¿debería eliminarse? | 10, la propuesta del proceso según los datos que indica el autor debería ser indispensable para la planeación de una célula de manufactura, considerando los ajustes del producto a desarrollar, la cantidad, el tiempo y los materiales (considerados en el desarrollo del modelo). |
| ¿Cuán relevante es la actividad para los objetivos del proceso? (Relevante o irrelevante). | Relevante |
| ¿Es un servicio prestado a los clientes de paso? (Si o no). | Sí |

| | |
|--|--|
| <p>¿Es una actividad interna o requiere aportes externos? Interno se entiende en el sentido de la organización, específicamente, el departamento de funciones.</p> | <p>Es una actividad interna</p> |
| <p>Nivel de partes interesadas. (VIP, propietario, consumidor, empleado y sistema).</p> | <p>Propietario (el laboratorio), el empleado al tener las herramientas necesarias para desempeñar su trabajo de manera más efectiva y por último el sistema que se verá beneficiado.</p> |
| <p>¿Es de empresa a empresa o de empresa a cliente?</p> | <p>De empresa a empresa, la propuesta la hace una estudiante de la misma Facultad de Ingeniería.</p> |

FICHA DE OBSERVACIÓN BÁSICA N° 3

- V. Datos generales:
- 5.1. Título de tesis: Desarrollo de un modelo para la predicción del precio del cobre empleando herramientas de Machine Learning
 - 5.2. Grado académico: Ingeniera Industrial
 - 5.3. Autor: Almudena Fosca Gamarra
 - 5.4. Fecha: Diciembre del 2020

5.5. Institución académica: Pontificia Universidad Católica Del Perú

5.6. Liga: <http://hdl.handle.net/20.500.12404/17730>

VI. Aspectos:

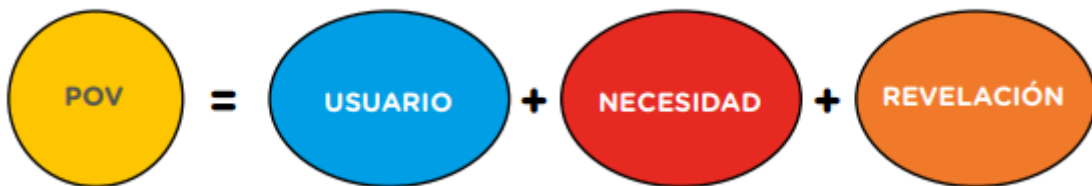
| Pregunta | Respuesta |
|--|--|
| ¿Función de proceso, por ejemplo, recursos humanos, finanzas, cadena de suministro y producción? | Dentro del departamento de planeación y aplicado en la cadena de suministro |
| ¿Procesa versiones de la historia; la versión del proceso abre otra dimensión del modelo de predicción del diseño del proceso alternativo.? | Sí, la investigación realizada abre la posibilidad de tener varios modelos. |
| ¿Desempeña la actividad contra sus KPIs: malo, bueno O excelente. Este atributo puede ser más preciso cuando los datos de KPI están disponibles? | No hay disponibilidad de KPI's, sin embargo en las conclusiones el autor describe como la herramienta estadística ARIMA, que fue utilizada, es la óptima para predecir estadísticamente. |
| ¿Valor del costo de operación de la actividad (¿aceptable? Sí o no). identificar las fuentes de datos y los conjuntos de datos requeridos. Ayuda a identificar la siguiente lista de atributos (características de aprendizaje automático) como entradas para el modelo de aprendizaje automático? | No se mencionan valores de costos, sin embargo, conforme a los resultados y por la naturaleza del proyecto investigado, la relación costo beneficio sería bastante buena al tener un buen sistema de predicción. |

| | |
|--|--|
| ¿Nivel de satisfacción del cliente (de uno a cinco); implementa la satisfacción del cliente en cada paso.? | No es mencionado en la investigación. |
| ¿Hace un cuello de botella? (Si o no). | No |
| ¿Hay suficientes recursos para realizar la tarea? (Si o no). | Sí |
| ¿Está digitalizado? (Si o no). | No |
| ¿Se puede automatizar? (Si o no). | Sí |
| ¿Es un paso de verificación del cumplimiento de la auditoría? (Si o no). | No |
| ¿Está siendo realizado por el mismo actor que el paso anterior en el proceso? (Si o no). | No |
| ¿Valor de la actividad para el proceso (1–10)? Si es menos de tres, ¿debería eliminarse? | 10, considerando que el modelo predictivo serviría para aumentar productividad y reducir costos. |
| ¿Cuán relevante es la actividad para los objetivos del proceso? (Relevante o irrelevante). | Relevante |
| ¿Es un servicio prestado a los clientes de paso? (Si o no). | No |
| ¿Es una actividad interna o requiere aportes externos? Interno se entiende | Sí |

| | |
|--|---|
| en el sentido de la organización, específicamente, el departamento de funciones. | |
| Nivel de partes interesadas. (VIP, propietario, consumidor, empleado y sistema). | El propietario y el departamento de planeación al tener un sistema que puede predecir y por lo tanto orientar sobre la manera en que se pueden seguir los procesos. |
| ¿Es de empresa a empresa o de empresa a cliente? | Empresa a empresa |

3..2. Definir

Recordando la información que obtuvimos utilizando la fase: **Empatizar**, de la metodología Design Thinking ahora desarrollaremos la fase: **Definir**, donde pasaremos a analizar y sintetizar la información recolectada a fin de llegar a conclusiones sobre las necesidades aplicando la técnica **Punto de vista** o **POV**.



Esta técnica consiste en definir el problema jugando, considerando que el usuario es el de la tesis y la necesidad viene de las respuestas brindadas en la fase: **Empatizar** y el Insight o visión viene de las preguntas de lectura crítica de las tesis, para lo cual se consideraron estas características plasmadas en nuestro **POV** siguiente:

| USUARIO | + | NECESIDAD | + | INSIGHT |
|---------|-----------------|--|---------------|---|
| Anggela | necesita | La aplicación del pronóstico | porque | Quiere el pronóstico de la demanda efectiva |
| Anggela | necesita | Corrobar la información de satisfacción del cliente | porque | Quiere obtener información sobre la eficacia de su aplicación |
| Anggela | necesita | Digitalizar sus KPIs | porque | Quiere buscarlos a partir de la información que tiene disponible |
| Anggela | necesita | automatizar sus tareas | porque | Quiere ayudar en la optimización laboral |
| Anggela | necesita | Elaborar una lista de verificación para sus procesos | porque | Quiere detallar la comprobación de sus resultados. |
| Anggela | necesita | Realizar el proceso de auditoria | porque | Quiere promover la mejora continua de los procesos de la organización |
| Anggela | necesita | Prestar servicio a los clientes | porque | quiere conocer el comportamiento de los clientes |
| Lina | necesita | Establecer un número de indicadores clave | porque | Quiere brindar una imagen clara de sus logros esperados |

| | | | | |
|----------|-----------------|---|---------------|---|
| Lina | necesita | mencionar los costos de operación | porque | Quiere analizar la importancia de los costos de producción |
| Lina | necesita | Medir la satisfacción del cliente | porque | quiere cumplir las expectativas por la falta de personalización |
| Lina | necesita | Automatizar tareas sencillas de manufactura | porque | Quiere utilizar mecanismos simples para la planificación de tareas |
| Lina | necesita | Comprobar de forma exhaustiva los procesos | porque | Quiere generar evidencia que proporcione mayor fiabilidad a la empresa |
| Lina | necesita | Realizar una auditoria | porque | Quiere poner a disposición de la empresa los balances de la situación |
| Almudena | necesita | Establecer un número de indicadores clave | porque | quiere explicar los precios del cobre |
| Almudena | necesita | Desarrollar técnicas de análisis de costos | porque | Quiere identificar los elementos que intervendrán en el modelo predictivo |
| Almudena | necesita | Cuantificar la satisfacción del cliente | porque | Quiere diseñar un servicio que incluya parámetros de calidad |

| | | | | |
|----------|-----------------|--|---------------|--|
| Almudena | necesita | digitalizar el proceso de predicción del precio de cobre | porque | Quiere ampliar la información sobre el precio del cobre |
| Almudena | necesita | Construir el modelo de predicción del precio de cobre | porque | Quiere brindar información sobre los futuros precios del cobre |
| Almudena | necesita | Validar el modelo de predicción de precio del cobre | porque | Quiere verificar si el modelo hace buenas predicciones |
| Almudena | necesita | Formar un equipo de auditoria | porque | Quiere ejecutar instrumentos de capacitación en la empresa |
| Almudena | necesita | Fijar un precio por cada predicción realizada | porque | Quiere crear un modelo de mercado para realizar un análisis de las predicciones de demanda |

3..3. Investigación de antecedentes

MODELO DE PRONÓSTICO DE DEMANDA DE EFECTIVO PARA LAS OFICINAS DE UNA ENTIDAD BANCARIA EN UNA CIUDAD DEL INTERIOR DEL PAIS

Resumen

El presente trabajo trata sobre el análisis de la demanda de efectivo en las oficinas de provincia de una entidad bancaria tomando como base la metodología que se sigue actualmente a fin de identificar oportunidades de mejora que permitan

tener resultados óptimos en los indicadores establecidos. Este análisis está centrado en la evaluación de la exactitud y precisión que son necesarias en el cálculo de los pronósticos de demanda para cada tipo de oficina existente, de manera que se puedan aplicar los modelos y la metodología que permitan obtener información más cercana a la realidad. En los primeros capítulos se definen los conceptos y etapas del proceso de extracción del conocimiento, para luego, por medio de estos, describir y explicar las redes neuronales, método elegido para la solución de la problemática expuesta. Asimismo, se detalla la situación actual del proceso que es objeto de análisis, identificando los factores que serían relevantes para la ejecución del modelo. En los capítulos posteriores, se procedió con el análisis de los datos obtenidos, de modo que se puedan identificar las variables relevantes que se incluirían dentro del modelo. De igual forma, se describe la lógica y los parámetros a tomar en cuenta para el mismo, y que dan origen a las diferentes etapas de prueba y error para llegar a los resultados esperados. Para ello, se recopilaron los datos necesarios de los movimientos en las oficinas los cuales sirvieron como datos de entrada para las redes neuronales generadas. El software empleado fue NeuralTools 5.7, debido a los beneficios que éste proporciona, desde el manejo de datos en hojas de cálculo, las cuales utilizan la interfaz de MS Excel, hasta el uso de herramientas y aplicaciones dinámicas que permiten manipular los datos de forma más sencilla optimizando tiempos. Finalmente, se procedió con la evaluación de los resultados obtenidos, haciendo la comparación de los valores predichos con los resultados reales, logrando alcanzar un porcentaje de exactitud de 93.18% para los pronósticos de una oficina mixta y de 91.98% para los de una oficina captadora, generando una expectativa de ahorro de S/. 25,000.00 anuales.

Desarrollo de un modelo predictivo para un sistema de manufactura orientado a la industria 4.0

Resumen

El monitoreo predictivo de los procesos de manufactura en cualquier organización tiene como objetivo pronosticar el comportamiento del sistema y de esta forma tomar decisiones de manera proactiva que mejoren su eficiencia. En el pasado se han propuesto varias técnicas de monitoreo predictivo, en su mayoría basadas en el análisis de datos históricos relacionados con el proceso mediante la utilización de técnicas de diferentes disciplinas como la Minería de Procesos. Este proyecto consiste en el desarrollo de un modelo predictivo para un sistema de manufactura orientado a la industria 4.0. El sistema simulado está basado en el sistema de producción que se encuentra instalado en los laboratorios de aprendizaje de manufactura de la Pontificia Universidad Javeriana. La metodología utilizada para desarrollar el modelo consta de 6 etapas cuya definición parte de la revisión literaria de trabajos similares desarrollados y se basa principalmente en las etapas de la guía de trabajo para proyectos de minería de datos CRISP-DM (Cross-Industry Standard Process for Data Mining). A través de las herramientas de Minería de Procesos brindadas por el software Celonis se realizó un análisis descriptivo del sistema y finalmente, por medio de un código desarrollado por Nirdizati Org en Python, se creó el modelo predictivo. Este modelo brinda información sobre el tiempo restante para que se finalice un producto que está en proceso. El conocimiento de esta información en momentos oportunos puede apalancar la toma de buenas decisiones para optimizar el desempeño del sistema de manufactura.

Desarrollo de un modelo para la predicción del precio del cobre empleando herramientas de Machine Learning

Resumen

A lo largo del presente trabajo de investigación se exploró el uso de herramientas de inteligencia artificial (Machine Learning) en la predicción del precio de cobre. Este proyecto de investigación se desarrolla dentro de las actividades del Grupo

de investigación en finanzas aplicadas (GIFA) de la PUCP. En el cual, a partir de conocimientos interdisciplinarios se busca explotar metodologías de Inteligencia Artificial aplicando Machine Learning en el ámbito de inversión financiera. En el capítulo 1, se exponen trabajos previos para el pronóstico de acciones, índices bursátiles y commodities, pudiendo comparar y contrastar los resultados obtenidos al aplicar diversos algoritmos. De esta forma, se emplean los estudios previos presentados como base para la ejecución y selección del modelo en esta tesis. Asimismo, de manera más detallada se presentan los factores más importantes en el comportamiento del precio de un commodity. En el capítulo 2, se estructura la metodología a emplear en el desarrollo de la investigación. Se especifica el tipo de investigación y diseño, así como las métricas de evaluación a emplear. El tercer capítulo corresponde al modelo de predicción basado en herramientas estadísticas tradicionales. Se presenta la metodología Box Jenkins como punto de partida para la ejecución del modelo ARIMA, posteriormente se evalúan los resultados obtenidos con este. A partir del cuarto capítulo se introducen los conceptos de Machine Learning. Inicialmente se presenta un flujograma base para la elaboración de un algoritmo, y con este se estructuran dos modelos: regresión lineal y SVR. A lo largo de este capítulo se construyen ambos algoritmos de manera básica, desde la categorización del problema hasta la validación, según el flujo de procesos presentado. El quinto capítulo tiene como objetivo evaluar la importancia de realizar un análisis de selección de atributos con el fin de mejorar el modelo. De esta forma, se utilizan dos algoritmos de selección y posteriormente se comparan los resultados obtenidos. El sexto y último capítulo del cuerpo de esta investigación se centra en optimizar el modelo SVR a través de la implementación de algoritmos de selección de hiperparámetros.

3.4. Idear

Se desea saber cómo formular una idea para cada necesidad encontrada en la fase: **Definir**, en relación con los tesisistas y las revelaciones encontradas de la siguiente manera:

| PREGUNTA INICIAL GENERAL | + | USUARIO | RESUELVE LA NECESIDAD | + | INSIGHT O RELEVACIÓN |
|---------------------------|-------|-------------|--|----------|---|
| ¿Cómo podríamos nosotros? | hacer | que Anggela | aplique el pronóstico | para que | pronostique la demanda efectiva. |
| ¿Cómo podríamos nosotros? | Hacer | que Anggela | corrobore la información de satisfacción del cliente | para que | obtenga información sobre la eficacia de su aplicación. |
| ¿Cómo podríamos nosotros? | Hacer | que Anggela | digitalice sus KPIs | para que | busque a partir de la información que tiene disponible. |
| ¿Cómo podríamos nosotros? | Hacer | que Anggela | automatice sus tareas | para que | ayude en la optimización laboral. |
| ¿Cómo podríamos nosotros? | Hacer | que Anggela | elabore una lista de verificación para sus procesos | para que | detalle la comprobación de sus resultados. |
| ¿Cómo podríamos nosotros? | Hacer | que Anggela | realice el proceso de auditoría | para que | promueva la mejora continua de los procesos. |

| | | | | | |
|---------------------------|-------|-------------|--|----------|--|
| | | | | | de la organización |
| ¿Cómo podríamos nosotros? | Hacer | Que Anggela | Preste servicio a los clientes | Para que | Conozca el comportamiento de los clientes |
| ¿Cómo podríamos nosotros? | Hacer | Que Lina | Establezca un número de indicadores clave | Para que | Brinde una imagen clara de sus logros esperados |
| ¿Cómo podríamos nosotros? | Hacer | Que Lina | Mencione los costos de operación | Para que | Analice la importancia de los costos de producción |
| ¿Cómo podríamos nosotros? | Hacer | Que Lina | Mida la satisfacción del cliente | Para que | Cumpla las expectativas por la falta de personalización |
| ¿Cómo podríamos nosotros? | Hacer | Que Lina | Automatice tareas sencillas de manufactura | Para que | Utilice mecanismos simples para la planificación de tareas |
| ¿Cómo podríamos nosotros? | Hacer | Que Lina | Compruebe de forma | Para que | Genere evidencia que proporcione |

| | | | | | |
|---------------------------|-------|--------------|---|----------|--|
| | | | efectiva los procesos | | mayor fiabilidad a la empresa |
| ¿Cómo podríamos nosotros? | Hacer | Que Lina | Realice una auditoria | Para que | Ponga a disposición de la empresa los balances de la situación |
| ¿Cómo podríamos nosotros? | Hacer | Que Almudena | Establezca un número de indicadores clave | Para que | Explique los precios del cobre |
| ¿Cómo podríamos nosotros? | Hacer | Que Almudena | Desarrolle técnicas de análisis de costos | Para que | Identifique los elementos que intervendrán en el modelo predictivo |
| ¿Cómo podríamos nosotros? | Hacer | Que Almudena | Cuantifique la satisfacción del cliente | Para que | Diseñe un servicio que incluya parámetros de calidad |
| ¿Cómo podríamos nosotros? | Hacer | Que Almudena | digitalice el proceso de predicción del precio de cobre | Para que | Amplie la información sobre el precio del cobre |

| | | | | | |
|---------------------------|-------|--------------|---|----------|--|
| ¿Cómo podríamos nosotros? | Hacer | Que Almudena | Construya el modelo de predicción del precio de cobre | Para que | Brinde información sobre futuros precios del cobre |
| ¿Cómo podríamos nosotros? | Hacer | Que Almudena | Valide el modelo de predicción de precio del cobre | Para que | Verifique si el modelo hace buenas predicciones |
| ¿Cómo podríamos nosotros? | Hacer | Que Almudena | Forme un equipo de auditoría | Para que | Ejecute instrumentos de capacitación en la empresa |
| ¿Cómo podríamos nosotros? | Hacer | Que Almudena | Fije un precio por cada predicción realizada | Para que | Cree un modelo de mercado para realizar un análisis de las predicciones de demanda |

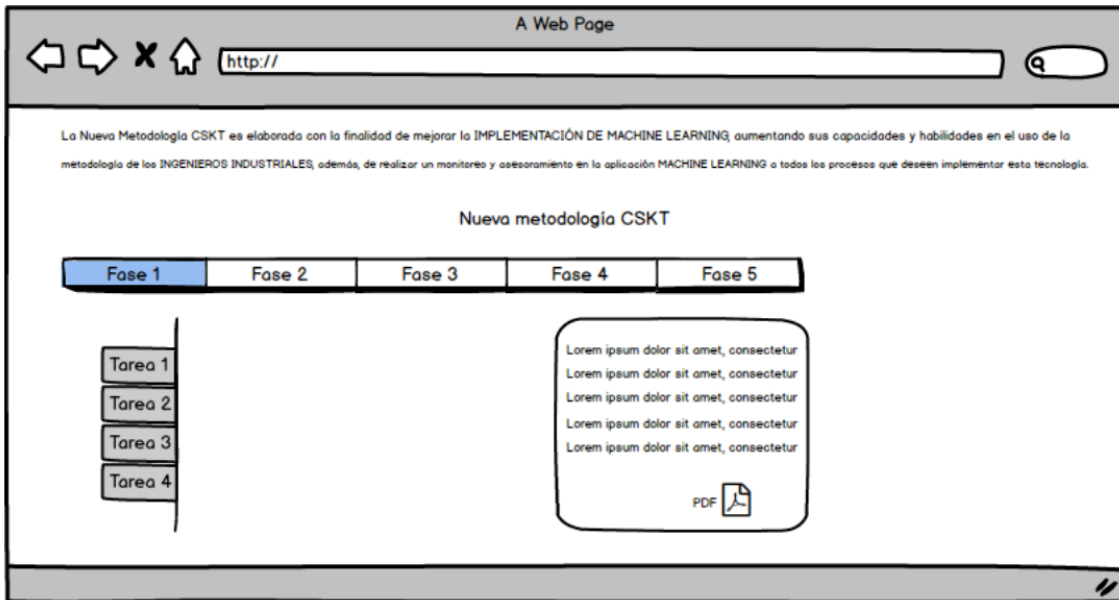
3..5. Prototipar

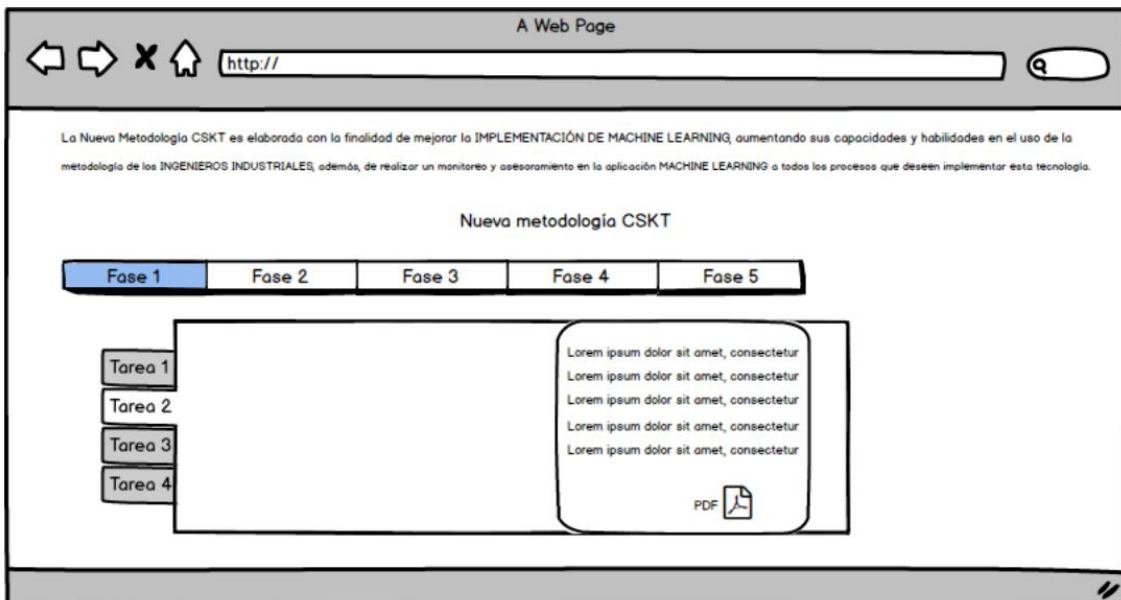
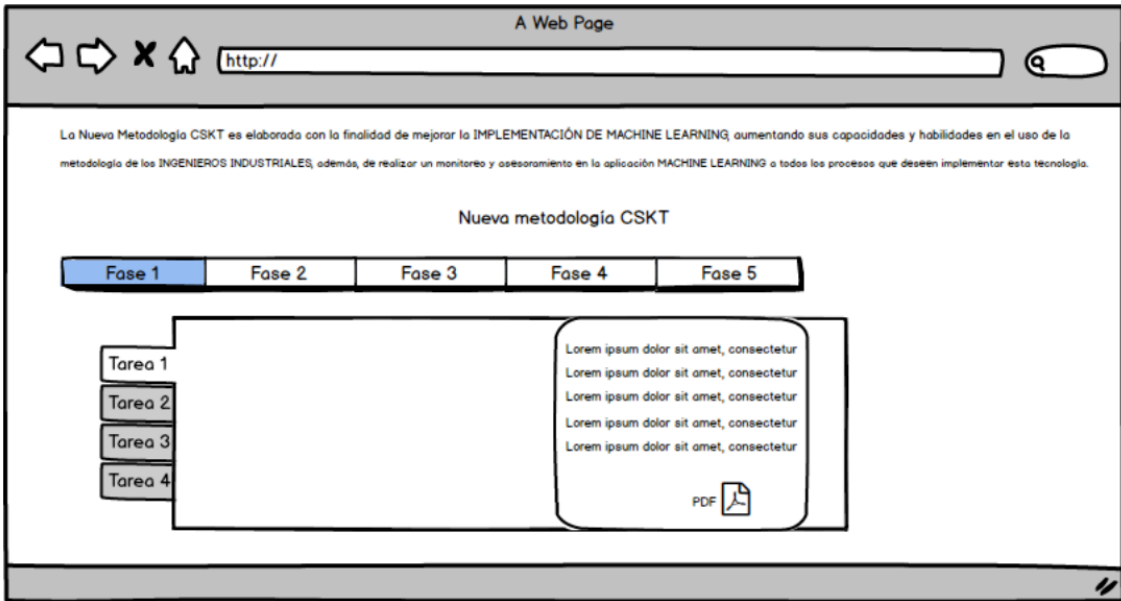
Cual será el prototipo que vas a utilizar

| PREPROTOTIPO | |
|--------------|------------------------|
| NOMBRE | Nueva metodología CSKT |

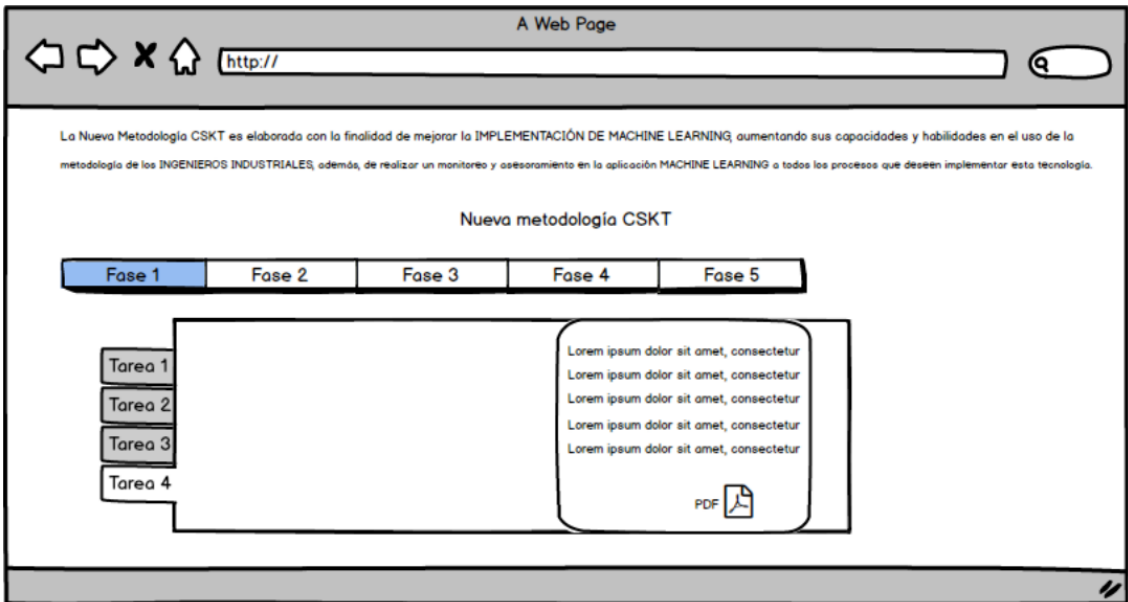
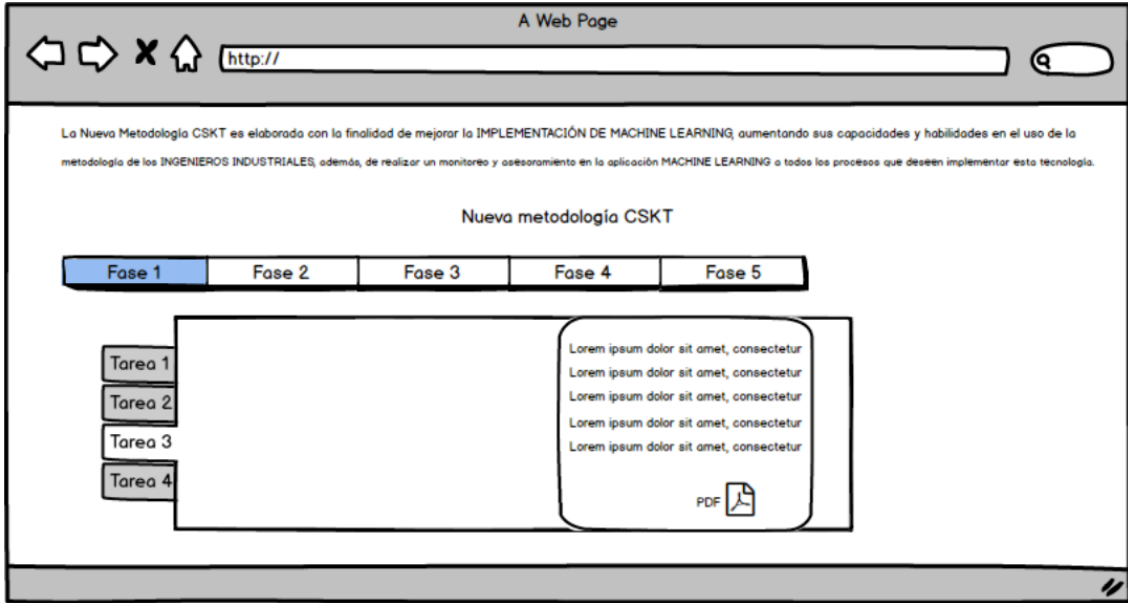
| | | |
|-----------------------------------|---|--------------------|
| DESCRIPCIÓN | Este prototipo pretende compartir un PDF con la documentación relacionada a cada fase de la nueva metodología que se está desarrollando, en esta página se recabará el número de descargas que se realiza para medir los resultados. | |
| ESQUEMA | Diagrama de flujo de las fases para la implementación de la metodología CSKT | |
| ELEMENTOS CLAVE | FASE I: MODELAMIENTO EMPRESARIAL FASE II: DEFINICIÓN DEL PROCESO FASE III: EJECUCIÓN DE MODELO DE MACHINE LEARNING FASE IV: PRESENTACIÓN DE LA SOLUCIÓN FASE V: MEDICIÓN DE RESULTADOS Y CIERRE | |
| COSTES DE PRODUCCIÓN | Descripción | Costo total |
| | Registrar el nombre del dominio | 40-60 Soles |
| | Contratar el hosting | 80-120 Soles |
| | Costes básicos para la página web | 120-180 Soles |
| TÉCNICAS DE PRODUCCIÓN | Scrum | |
| DIFICULTADES DE PRODUCCIÓN | <ul style="list-style-type: none"> • Priorizar funcionalidades por sobre el tiempo para presentar tangiblemente los prototipos a los usuarios. • Pretender un diseño que guste a todo el mundo. • Surgen nuevas necesidades que hacen crecer los costos. | |

| | |
|-------------------------|---|
| | <ul style="list-style-type: none"> • Incompatibilidades entre las prestaciones del hosting y lo necesario. • Fallas de seguridad. • El sitio no se ve en dispositivos móviles. |
| ASPECTOS LEGALES | <ul style="list-style-type: none"> • Establecer los términos y condiciones del sitio web: • Titular • Material que se ofrece • Información relativa a la aceptación de cookies • Informar de las políticas de la página web • Informar sobre la responsabilidad de los autores de la nueva metodología CSKT • Agregar un formulario de quejas en caso de plagio. |

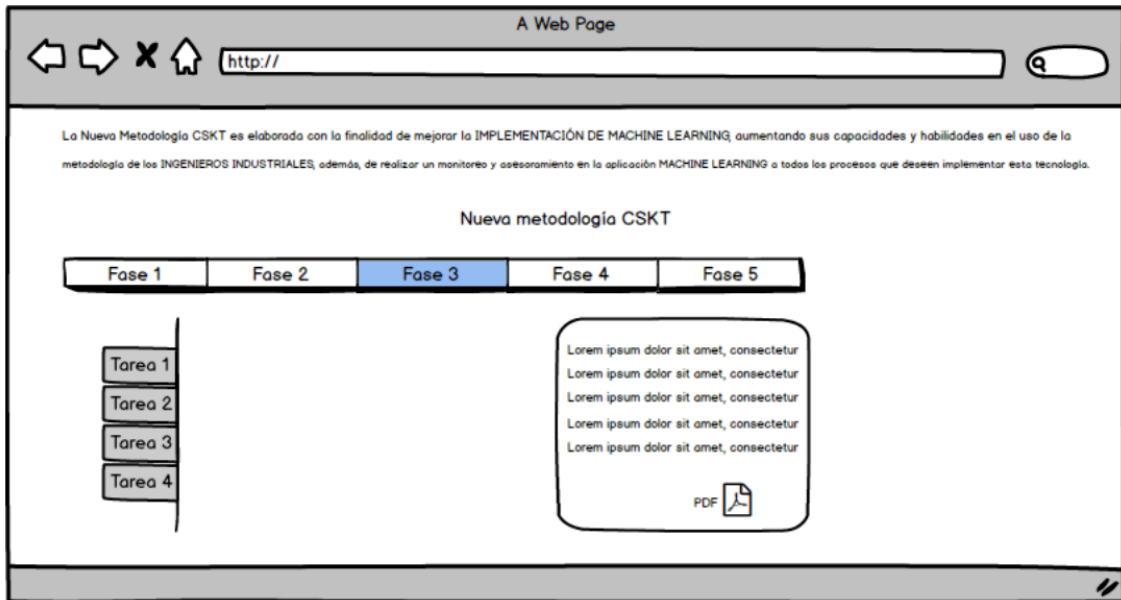
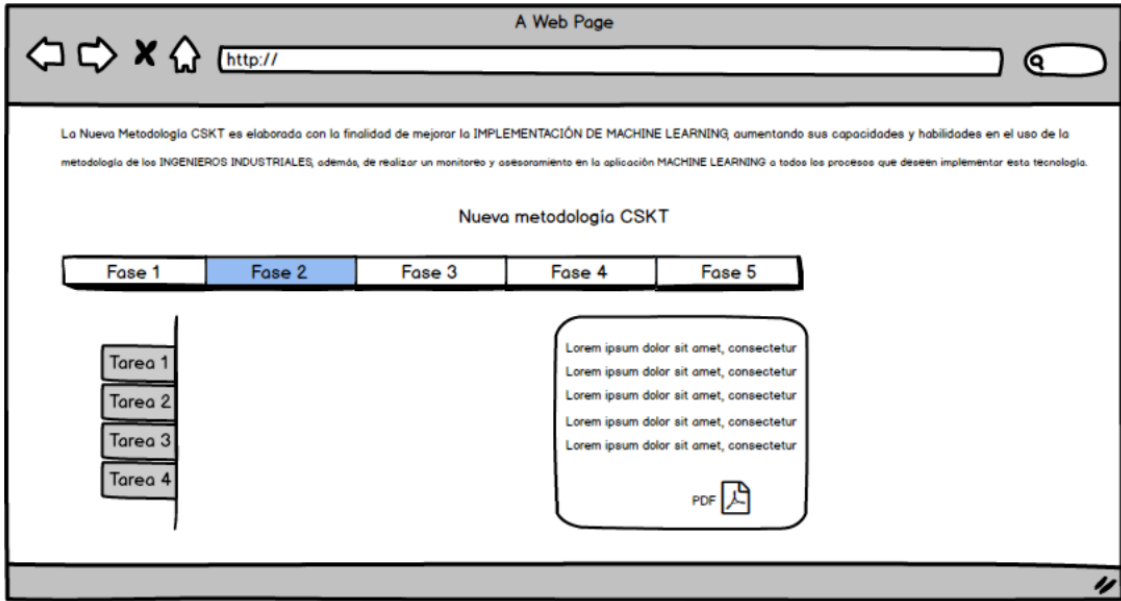




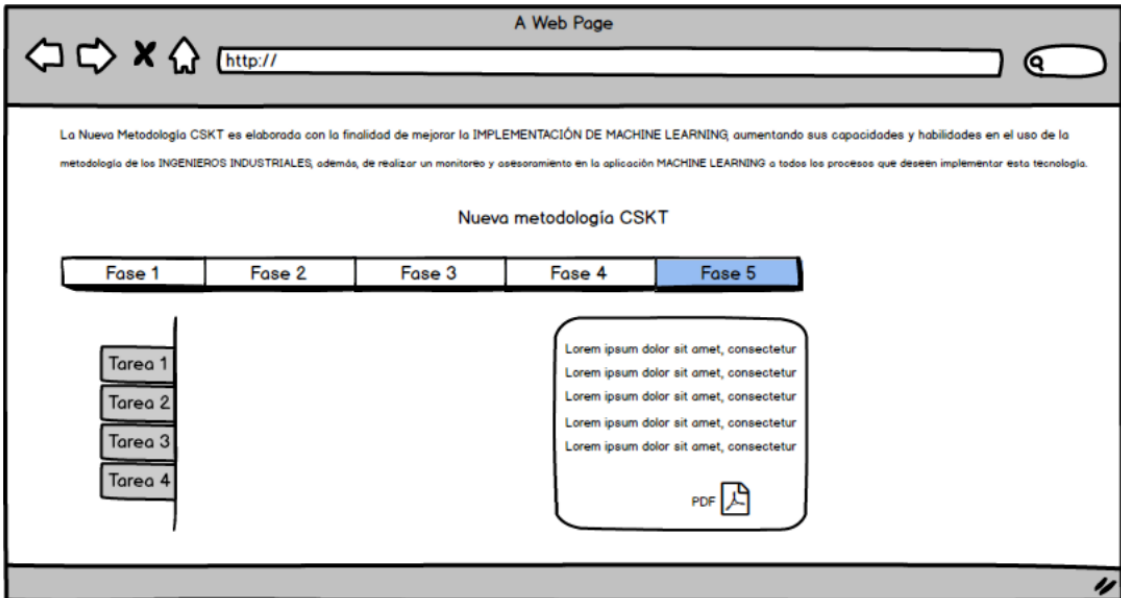
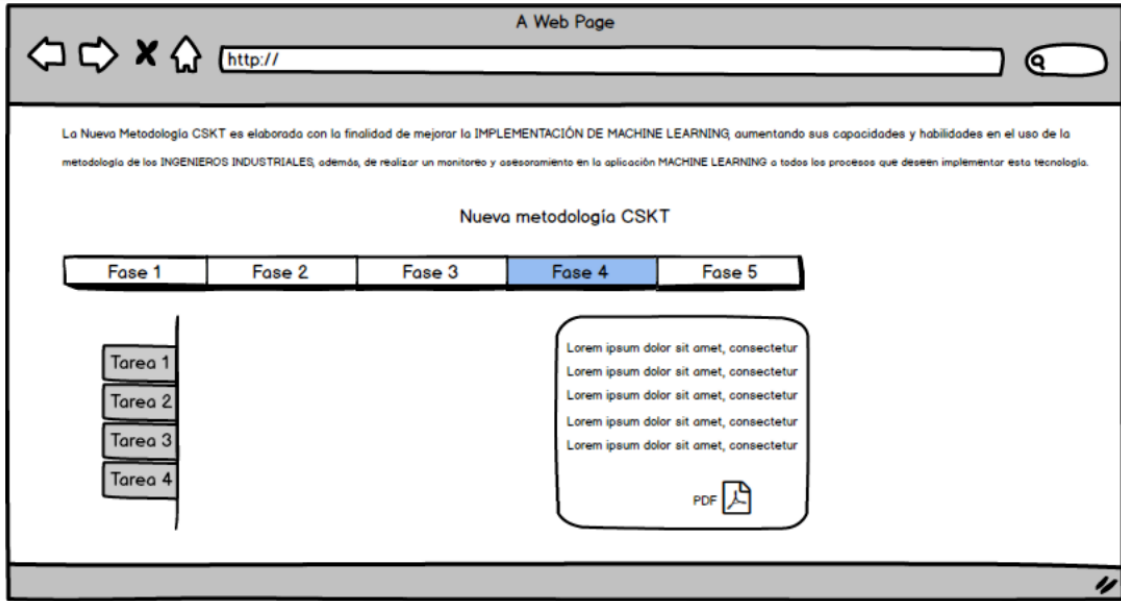
Nueva metodología CSKT para mejorar los proyectos de implementación de Machine Learning en Ingeniería Industrial



Nueva metodología CSKT para mejorar los proyectos de implementación de Machine Learning en Ingeniería Industrial



Nueva metodología CSKT para mejorar los proyectos de implementación de Machine Learning en Ingeniería Industrial



3..6. Selección de prototipos y criterio para hacerlo (evaluar)

| MALLA RECEPTORA DE INFORMACIÓN | |
|--|---|
| <p>IDEAS INTERESANTES</p> <ul style="list-style-type: none"> • Utilizar la estrategia en numerosas ocasiones para ayudar a los estudiantes de Ingeniería industrial a generar nuevas ideas para sus proyectos de tesis. • Cuando hablamos de metodología podemos aplicarlo al área de desarrollo y enseñanza para los proyectos considerados adecuados. • Todas las diferencias con respecto a otras metodologías de Machine Learning deben usar para beneficiar a todos los estudiantes de Ingeniería Industrial para los proyectos que consideran adecuados. | <p>CRÍTICAS CONSTRUCTIVAS</p> <ul style="list-style-type: none"> • Una metodología cualquier estudiante lo conoce, pero que normalmente no se aplican porque el esfuerzo que requiere es muy alto. • Este material aplica al enfoque metodológico basado en otras metodologías en Machine Learning que han probado tener éxito en la optimización de procesos. • Los asesores de tesis ya han echado un vistazo a la metodología y la quieren probar para los proyectos que consideren adecuados. |
| <p>PREGUNTAS NUEVAS</p> <ul style="list-style-type: none"> • ¿Encontrar metodologías nuevas más significativas disminuirá la deserción o fracaso en el proceso de elaboración de tesis en Ingeniería Industrial? | <p>IDEAS NUEVAS</p> <ul style="list-style-type: none"> • Aplicar la enseñanza aprendizaje por investigación ligada a la metodología. • Las metodologías educativas suelen girar en torno a las teorías de aprendizaje. |

| | |
|--|--|
| <ul style="list-style-type: none">• ¿Cuándo algún estudiante me pregunte qué metodología hacer para el proyecto de investigación, la primera pregunta que le hago es: usarás Machine Learning?• La nueva metodología CSKT se presenta como un modelo educativo para Ingeniería Industrial porque la propuesta tiene en cuenta que la carrera se dedica a la optimización de procesos y las tesis no presentan de manera explícita un proyecto con Machine Learning?• ¿Los estudiantes que cursan metodologías de investigación poseen los requisitos necesarios para abordar la propuesta? | <ul style="list-style-type: none">• Cuando el estudiante pone en juego sus conocimientos es cuando verdaderamente descubre el significado. |
|--|--|

4. Metodología

4.1. Scrum

Para la elaboración del trabajo se emprende la metodología Scrum que se pasa a detallar.

Tabla 1

Caso del Negocio

| Descripción del Caso del negocio |
|---|
| <p>Se realizó un estudio de las carreras universitarias que se dedican a la optimización de profesor en donde identificó que Ingeniería Industrial en el país de Perú es una carrera con relativamente un plan de estudio que no esencialmente desemboca en tesis de investigación de optimización de procesos, por ello como interés por el enfoque que tiene la carrera, se espera que esta investigación les permita avanzar entre optimización de procesos utilizando las herramientas que se utilizan para tal propósito; para ello debe existir una metodología con al menos tres fases y que, a diferencia de otras metodologías como SEMMA y CRISP-DM, los beneficiados sean los estudiantes que están elaborando una tesis en Ingeniería Industrial para que les proporcione beneficios como interacción más rápida con el negocio y que literalmente logren automatizar sus proyectos que consideren adecuado utilizar Machine Learning.</p> <p>La universidad quiere entender cómo funcionará la metodología antes de iniciar cualquier desarrollo, han seleccionado los mejores estudiantes para crear un prototipo, en general lo que esperan es que cada investigación contenga la nueva metodología que se les proporcionará para implementar proyectos de Machine Learning por orden importancia lo siguiente:</p> <p>1) Tres fases con al menos 3 actividades.</p> |

- 2) Cada fase de la nueva metodología debe tener su desarrollo de ejemplo sobre un proyecto relacionado a la carrera de Ingeniería Industrial, pues será de gran valor al momento de aplicarla a las investigaciones de los estudiantes, aunque no es el único factor determinante.
- 3) Los estudiantes deben avanzar a través de las fases planteadas de la nueva metodología y tener que pasar al menos todos los recursos que se consideren convenientes compartir para el estudiante que la utilizará.
- 4) Es importante contar con un plan que permita al usuario entender la secuencia de la nueva metodología sin necesidad que le expliquen de forma verbal.

Tabla 2

Visión

| Project Visión | |
|----------------------------------|--|
| PARA | Estudiantes de Ingeniería Industrial |
| OPORTUNIDAD NECESIDAD | O Interés por optimización de procesos |
| LA | "Tesis" |
| BENEFICIOS | Mejorar los proyectos de implementación de Machine Learning en Ingeniería Industrial |
| A DIFERENCIA | CRISP-DM, TDSP, SEMMA, KDD |

NUESTRO PRODUCTO

Nueva metodología CSKT implementada en una página web que es fácil de ser consultada y de ser entendida por el usuario de modo que no necesite una explicación verbal para implementar sus proyectos de Tesis con Machine Learning para la optimización de procesos, que es el enfoque de su carrera profesional.

Tabla 3

Involucrados del proyecto

| INVOLUCRADOS EN EL PROYECTO | | | | |
|-----------------------------|-------------|----------------------|---------|--|
| | Involucrado | Rol | Tipo | Descripción del rol |
| SCRUM TEAM | Ogosi | SCRUM MASTER | CORE | Pensamiento sistémico, conocimientos en ingeniería industrial y desarrollo software |
| | Ogosi | PRODUCT OWNER | CORE | Negociador, alto conocimiento de mejora de procesos, comunicación efectiva, Networking, pensamiento crítico y constructivo |
| | Karla | SCRUM DEVELOPER TEAM | CORE | Diseñadora de páginas web |
| | Angella | Cientes1 | NO CORE | Tesista de Ingeniería Industrial |
| | Lina | Cientes2 | NO CORE | Tesista de Ingeniería Industrial |
| | Almudena | Cientes3 | NO CORE | Tesista de Ingeniería Industrial |

Tabla 4

Product Backlog Item

| ESTRUCTURA DE LA HISTORIA DE USUARIO | | | | | | | | | |
|--------------------------------------|----------|--------------|--------------|--|---|---------------------|------------|--------------|---|
| Product Backlog de Proyecto | | | | | | | | | |
| ID | TIPO | Priorización | Rol/Personal | Historia de Usuario | DC/UAT | Valor/ Beneficio | Estimación | Esfuerzo Min | Product Backlog |
| 101 | Historia | 1 | Almudena | necesita fijar un precio por cada predicción realizada porque quiere crear un modelo de mercado para realizar un análisis de las predicciones de demanda | Me den un plan de acción | Mejor exploración | 30 | 40 | Almudena necesita fijar un precio por cada predicción realizada porque quiere crear un modelo de mercado para realizar un análisis de las predicciones de demanda Me den un plan de acción Mejor exploración |
| 102 | Historia | 2 | Almudena | necesita validar el modelo de predicción de precio del cobre porque quiere verificar si el modelo hace buenas predicciones | Me den una visualización de planes de acción | Mejor visualización | 40 | 38 | Almudena necesita validar el modelo de predicción de precio del cobre porque quiere verificar si el modelo hace buenas predicciones Me den una visualización de planes de acción Mejor visualización |
| 103 | Historia | 3 | Angella | necesita realizar el proceso de auditoria porque quiere promover la mejora continua de los procesos de la organización | Me den una visualización de los procesos | Mejor visualización | 40 | 35 | Angella necesita realizar el proceso de auditoria porque quiere promover la mejora continua de los procesos de la organización Me den una visualización de los procesos Mejor visualización |
| 104 | Historia | 4 | Almudena | necesita digitalizar el proceso de predicción del precio de cobre porque quiere ampliar la información sobre el precio del cobre | Me den una visualización de datos específicos del cobre | Mejor visualización | 40 | 33 | Almudena necesita digitalizar el proceso de predicción del precio de cobre porque quiere ampliar la información sobre el precio del cobre Me den una visualización de datos específicos del cobre Mejor visualización |
| 105 | Historia | 5 | Angella | necesita elaborar una lista de verificación para sus procesos porque quiere detallar la comprobación de sus resultados | Me den un reporte de acciones realizadas | Mejor exploración | 30 | 30 | Angella necesita elaborar una lista de verificación para sus procesos porque quiere detallar la comprobación de sus resultados Me den un reporte de acciones realizadas Mejor exploración |

Nueva metodología CSKT para mejorar los proyectos de implementación de Machine Learning en Ingeniería Industrial

| | | | | | | | | | |
|-----|----------|----|----------|--|---|---------------------|----|----|---|
| 106 | Historia | 6 | Lina | necesita mencionar los costos de operación porque quiere analizar la importancia de los costos de producción | Me den una visualización de los costos | Mejor visualización | 40 | 28 | Lina necesita mencionar los costos de operación porque quiere analizar la importancia de los costos de producción Me den una visualización de los costos Mejor visualización |
| 107 | Historia | 7 | Lina | necesita medir la satisfacción del cliente porque quiere cumplir las expectativas por la falta de personalización | Me den una visualización de los clientes | Mejor visualización | 40 | 27 | Lina necesita medir la satisfacción del cliente porque quiere cumplir las expectativas por la falta de personalización Me den una visualización de los clientes Mejor visualización |
| 108 | Historia | 8 | Almudena | necesita construir el modelo de predicción del precio de cobre porque quiere brindar información sobre los futuros precios del cobre | Me den una visualización del precio de cobre | Mejor visualización | 40 | 26 | Almudena necesita construir el modelo de predicción del precio de cobre porque quiere brindar información sobre los futuros precios del cobre Me den una visualización del precio de cobre Mejor visualización |
| 109 | Historia | 9 | Almudena | necesita formar un equipo de auditoria porque quiere ejecutar instrumentos de capacitación en la empresa | Me den un plan de acción | Mejor preparación | 20 | 25 | Almudena necesita formar un equipo de auditoria porque quiere ejecutar instrumentos de capacitación en la empresa Me den un plan de acción Mejor preparación |
| 110 | Historia | 10 | Angella | necesita corroborar la información de satisfacción del cliente porque quiere obtener información sobre la eficacia de su aplicación | Me den información de la aceptación del cliente | Mejor visualización | 40 | 23 | Angella necesita corroborar la información de satisfacción del cliente porque quiere obtener información sobre la eficacia de su aplicación Me den información de la aceptación del cliente Mejor visualización |
| 111 | Historia | 11 | Angella | necesita prestar servicio a los clientes porque quiere conocer el comportamiento de los clientes | Me den una visualización de los clientes | Mejor visualización | 40 | 22 | Angella necesita prestar servicio a los clientes porque quiere conocer el comportamiento de los clientes Me den una visualización de los clientes Mejor visualización |
| 112 | Historia | 12 | Angella | necesita la aplicación del pronóstico porque quiere el pronóstico de la demanda de efectivo | Me den un plan de acción | Mejor visualización | 40 | 20 | Angella necesita la aplicación del pronóstico porque quiere el pronóstico de la demanda de efectivo Me den un plan de acción Mejor visualización |

Nueva metodología CSKT para mejorar los proyectos de implementación de Machine Learning en Ingeniería Industrial

| | | | | | | | | | |
|-----|----------|----|----------|--|---|---------------------|----|----|---|
| 113 | Historia | 13 | Lina | necesita realizar una auditoria porque quiere poner a disposición de la empresa los balances de la situación | Me den una visualización de los datos obtenidos | Mejor visualización | 40 | 19 | Lina necesita realizar una auditoria porque quiere poner a disposición de la empresa los balances de la situación Me den una visualización de los datos obtenidos Mejor visualización |
| 114 | Historia | 14 | Almudena | necesita desarrollar técnicas de análisis de costos porque quiere identificar los elementos que intervendrán en el modelo predictivo | Me den un reporte inicial de los datos facilitados por la empresa | Mejor preparación | 20 | 18 | Almudena necesita desarrollar técnicas de análisis de costos porque quiere identificar los elementos que intervendrán en el modelo predictivo Me den un reporte inicial de los datos facilitados por la empresa Mejor preparación |
| 115 | Historia | 15 | Lina | necesita comprobar de forma exhaustiva los procesos porque quiere generar evidencia que proporcione mayor fiabilidad a la empresa | Me den una visualización de los datos obtenidos | Mejor visualización | 40 | 16 | Lina necesita comprobar de forma exhaustiva los procesos porque quiere generar evidencia que proporcione mayor fiabilidad a la empresa Me den una visualización de los datos obtenidos Mejor visualización |
| 116 | Historia | 16 | Almudena | necesita cuantificar la satisfacción del cliente porque quiere diseñar un servicio que incluya parámetros de calidad | Me den una visualización de los clientes | Mejor visualización | 40 | 13 | Almudena necesita cuantificar la satisfacción del cliente porque quiere diseñar un servicio que incluya parámetros de calidad Me den una visualización de los clientes Mejor visualización |
| 117 | Historia | 17 | Almudena | necesita establecer un número de indicadores clave porque quiere explicar los precios del cobre | Me den una visualización de algunos datos específicos del cobre | Mejor visualización | 40 | 12 | Almudena necesita establecer un número de indicadores clave porque quiere explicar los precios del cobre Me den una visualización de algunos datos específicos del cobre Mejor visualización |
| 118 | Historia | 18 | Lina | necesita establecer un número de indicadores clave porque quiere brindar una imagen clara de sus logros esperados | Me den los datos facilitados por la empresa | Mejor preparación | 20 | 11 | Lina necesita establecer un número de indicadores clave porque quiere brindar una imagen clara de sus logros esperados Me den los datos facilitados por la empresa Mejor preparación |

Nueva metodología CSKT para mejorar los proyectos de implementación de Machine Learning en Ingeniería Industrial

| | | | | | | | | | |
|-----|----------|-----------|---------|--|---|-------------------|----|----|--|
| 119 | Historia | 19 | Angella | necesita automatizar sus tareas porque quiere ayudar en la optimización laboral | Me den un plan de acción | Mejor preparación | 20 | 10 | Angella necesita automatizar sus tareas porque quiere ayudar en la optimización laboral Me den un plan de acción Mejor preparación |
| 120 | Historia | 20 | Angella | necesita digitalizar sus KPIs porque quiere buscarlos a partir de la información que tiene disponible | Me den los datos facilitados por la empresa | Mejor preparación | 20 | 9 | Angella necesita digitalizar sus KPIs porque quiere buscarlos a partir de la información que tiene disponible Me den los datos facilitados por la empresa Mejor preparación |
| 121 | Historia | 21 | Lina | necesita automatizar tareas sencillas de manufactura porque quiere utilizar mecanismos simples para la planificación de tareas | Me den un plan de acción | Mejor preparación | 20 | 8 | Lina necesita automatizar tareas sencillas de manufactura porque quiere utilizar mecanismos simples para la planificación de tareas Me den un plan de acción Mejor preparación |

Tabla 5

Puntos

| ESTIMACIÓN ENTREGABLES | | | ESTIMACIÓN SPRINTS | | | Consideraciones generales para la estimación |
|--------------------------|-------------------------------|---------|--------------------------|-------------------------------|---------|--|
| Puntos (Planning Pocker) | Equivalencia Release Esfuerzo | Unidad | Puntos (Planning Pocker) | Equivalencia Release Esfuerzo | Unidad | |
| 1 | 30 | Minutos | 1 | 40 | Minutos | Con un sprint de 3 semanas |
| 2 | 40 | Minutos | 2 | 38 | Minutos | |
| 3 | 40 | Minutos | 3 | 35 | Minutos | |
| 4 | 40 | Minutos | 4 | 33 | Minutos | |
| 5 | 30 | Minutos | 5 | 30 | Minutos | |
| 6 | 40 | Minutos | 6 | 28 | Minutos | |
| 7 | 40 | Minutos | 7 | 27 | Minutos | |
| 8 | 40 | Minutos | 8 | 26 | Minutos | |
| 9 | 20 | Minutos | 9 | 25 | Minutos | |
| 10 | 40 | Minutos | 10 | 23 | Minutos | |
| 11 | 40 | Minutos | 11 | 22 | Minutos | |
| 12 | 40 | Minutos | 12 | 20 | Minutos | |
| 13 | 40 | Minutos | 13 | 19 | Minutos | |
| 14 | 20 | Minutos | 14 | 18 | Minutos | |
| 15 | 40 | Minutos | 15 | 16 | Minutos | |
| 16 | 40 | Minutos | 16 | 13 | Minutos | |
| 17 | 40 | Minutos | 17 | 12 | Minutos | |
| 18 | 20 | Minutos | 18 | 11 | Minutos | |
| 19 | 20 | Minutos | 19 | 10 | Minutos | |
| 20 | 20 | Minutos | 20 | 9 | Minutos | |
| 21 | 20 | Minutos | 21 | 8 | Minutos | |

Tabla 6

Tareas

| SPRINT | HISTORIA DE USUARIO | Responsable | Actividad | Story Points | Estatus |
|--------|---|-------------|-------------------------------------|--------------|-----------------|
| 1 | Almudena necesita fijar un precio por cada predicción realizada porque quiere crear un modelo de mercado para realizar un análisis de las predicciones de demanda Me den un plan de acción Mejor exploración | Karla | Mejorar exploración en página web | 3 | WORKING PROCESS |
| 2 | Almudena necesita validar el modelo de predicción de precio del cobre porque quiere verificar si el modelo hace buenas predicciones Me den una visualización de planes de acción Mejor visualización | Karla | Mejorar visualización en página web | 3 | WORKING PROCESS |
| 3 | Angella necesita realizar el proceso de auditoria porque quiere promover la mejora continua de los procesos de la organización Me den una visualización de los procesos Mejor visualización | Karla | Mejorar visualización en página web | 3 | WORKING PROCESS |
| 4 | Almudena necesita digitalizar el proceso de predicción del precio de cobre porque quiere ampliar la información sobre el precio del cobre Me den una visualización de datos específicos del cobre Mejor visualización | Karla | Mejorar visualización en página web | 3 | WORKING PROCESS |
| 5 | Angella necesita elaborar una lista de verificación para sus procesos porque quiere detallar la comprobación de sus resultados Me den un reporte de acciones realizadas Mejor exploración | Karla | Mejorar exploración en página web | 3 | WORKING PROCESS |
| 6 | Lina necesita mencionar los costos de operación porque quiere analizar la importancia de los costos de producción Me den una visualización de los costos Mejor visualización | Karla | Mejorar visualización en página web | 3 | WORKING PROCESS |
| 7 | Lina necesita medir la satisfacción del cliente porque quiere cumplir las expectativas por la falta de personalización Me den una visualización de los clientes Mejor visualización | Karla | Mejorar visualización en página web | 3 | WORKING PROCESS |

| | | | | | |
|----|---|-------|-------------------------------------|---|-----------------|
| 8 | Almudena necesita construir el modelo de predicción del precio de cobre porque quiere brindar información sobre los futuros precios del cobre Me den una visualización del precio de cobre Mejor visualización | Karla | Mejorar visualización en página web | 3 | WORKING PROCESS |
| 9 | Almudena necesita formar un equipo de auditoria porque quiere ejecutar instrumentos de capacitación en la empresa Me den un plan de acción Mejor preparación | Karla | Mejorar preparación en página web | 3 | WORKING PROCESS |
| 10 | Angella necesita corroborar la información de satisfacción del cliente porque quiere obtener información sobre la eficacia de su aplicación Me den información de la aceptación del cliente Mejor visualización | Karla | Mejorar visualización en página web | 3 | WORKING PROCESS |
| 11 | Angella necesita prestar servicio a los clientes porque quiere conocer el comportamiento de los clientes Me den una visualización de los clientes Mejor visualización | Karla | Mejorar visualización en página web | 3 | WORKING PROCESS |
| 12 | Angella necesita la aplicación del pronóstico porque quiere el pronóstico de la demanda de efectivo Me den un plan de acción Mejor visualización | Karla | Mejorar visualización en página web | 3 | WORKING PROCESS |
| 13 | Lina necesita realizar una auditoria porque quiere poner a disposición de la empresa los balances de la situación Me den una visualización de los datos obtenidos Mejor visualización | Karla | Mejorar visualización en página web | 2 | WORKING PROCESS |
| 14 | Almudena necesita desarrollar técnicas de análisis de costos porque quiere identificar los elementos que intervendrán en el modelo predictivo Me den un reporte inicial de los datos facilitados por la empresa Mejor preparación | Karla | Mejorar preparación en página web | 2 | WORKING PROCESS |
| 15 | Lina necesita comprobar de forma exhaustiva los procesos porque quiere generar evidencia que proporcione mayor fiabilidad a la empresa Me den una visualización de los datos obtenidos Mejor visualización | Karla | Mejorar visualización en página web | 2 | WORKING PROCESS |
| 16 | Almudena necesita cuantificar la satisfacción del cliente porque quiere diseñar un servicio que incluya parámetros de calidad Me den una visualización de los clientes Mejor visualización | Karla | Mejorar visualización en página web | 2 | WORKING PROCESS |

Nueva metodología CSKT para mejorar los proyectos de implementación de Machine Learning en Ingeniería Industrial

| | | | | | |
|---------|--|-------|-------------------------------------|----|-----------------|
| 17 | Almudena necesita establecer un número de indicadores clave porque quiere explicar los precios del cobre Me den una visualización de algunos datos específicos del cobre Mejor visualización | Karla | Mejorar visualización en página web | 2 | WORKING PROCESS |
| 18 | Lina necesita establecer un número de indicadores clave porque quiere brindar una imagen clara de sus logros esperados Me den los datos facilitados por la empresa Mejor preparación | Karla | Mejorar preparación en página web | 2 | WORKING PROCESS |
| 19 | Angella necesita automatizar sus tareas porque quiere ayudar en la optimización laboral Me den un plan de acción Mejor preparación | Karla | Mejorar preparación en página web | 2 | WORKING PROCESS |
| 20 | Angella necesita digitalizar sus KPIs porque quiere buscarlos a partir de la información que tiene disponible Me den los datos facilitados por la empresa Mejor preparación | Karla | Mejorar preparación en página web | 1 | WORKING PROCESS |
| 21 | Lina necesita automatizar tareas sencillas de manufactura porque quiere utilizar mecanismos simples para la planificación de tareas Me den un plan de acción Mejor preparación | Karla | Mejorar preparación en página web | 1 | WORKING PROCESS |
| TOTALES | | | | 57 | |

Tabla 7

Burndown Chart

| BRUDOWN CHART | |
|---|------------|
| Fecha de Inicio | 1/09/2022 |
| Fecha de Fin | 22/09/2022 |
| Total de días efectivos de trabajo (ideal time) | 13 |
| Total de puntos | 52 |
| Ideal de puntos por día | 4 |

| | |
|---------------------------------|-----------|
| Sprint 3 semanas | 15 |
| Planning Meeting | 0.750 |
| Reunion de Revision | 0.375 |
| Reunion Retrospectiva | 0.375 |
| Dias festivos | 0.5 |
| Días realmente efectivos | 13 |

13

Tabla 8

Daily Scrum

| GESTIÓN DIARIA | | | |
|----------------|------------------|--------------|--------------|
| Dia | Trabajo Planeado | Trabajo Real | Valor Actual |
| 1 | 52.0 | 52 | 0 |
| 2 | 48.0 | 52 | 0 |
| 3 | 44.0 | 52 | 0 |
| 4 | 40.0 | 52 | 0 |
| 5 | 36.0 | 52 | 0 |
| 6 | 32.0 | 52 | 0 |
| 7 | 28.0 | 52 | 0 |

Nueva metodología CSKT para mejorar los proyectos de implementación de Machine Learning en Ingeniería

| | | | |
|----|-------|----|---|
| 8 | 24.0 | 52 | 0 |
| 9 | 20.0 | 52 | 0 |
| 10 | 16.0 | 52 | 0 |
| 11 | 12.0 | 52 | 0 |
| 12 | 8.0 | 52 | 0 |
| 13 | 4.0 | 52 | 0 |
| 14 | 0.0 | 52 | 0 |
| 15 | -4.0 | 52 | 0 |
| 16 | -8.0 | 52 | 0 |
| 17 | -12.0 | 52 | 0 |
| 18 | -16.0 | 52 | 0 |
| 19 | -20.0 | 52 | 0 |
| 20 | -24.0 | 52 | 0 |
| 21 | -28.0 | 52 | 0 |

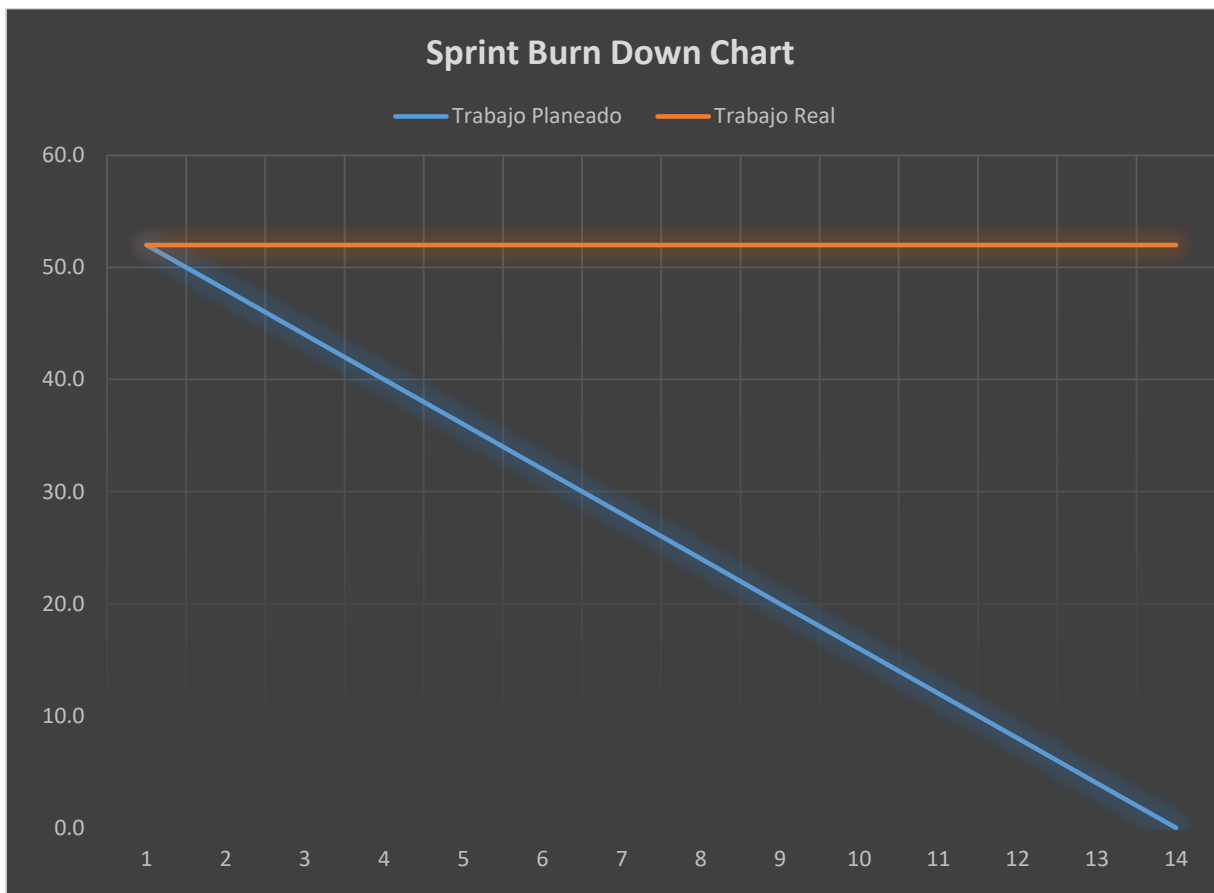


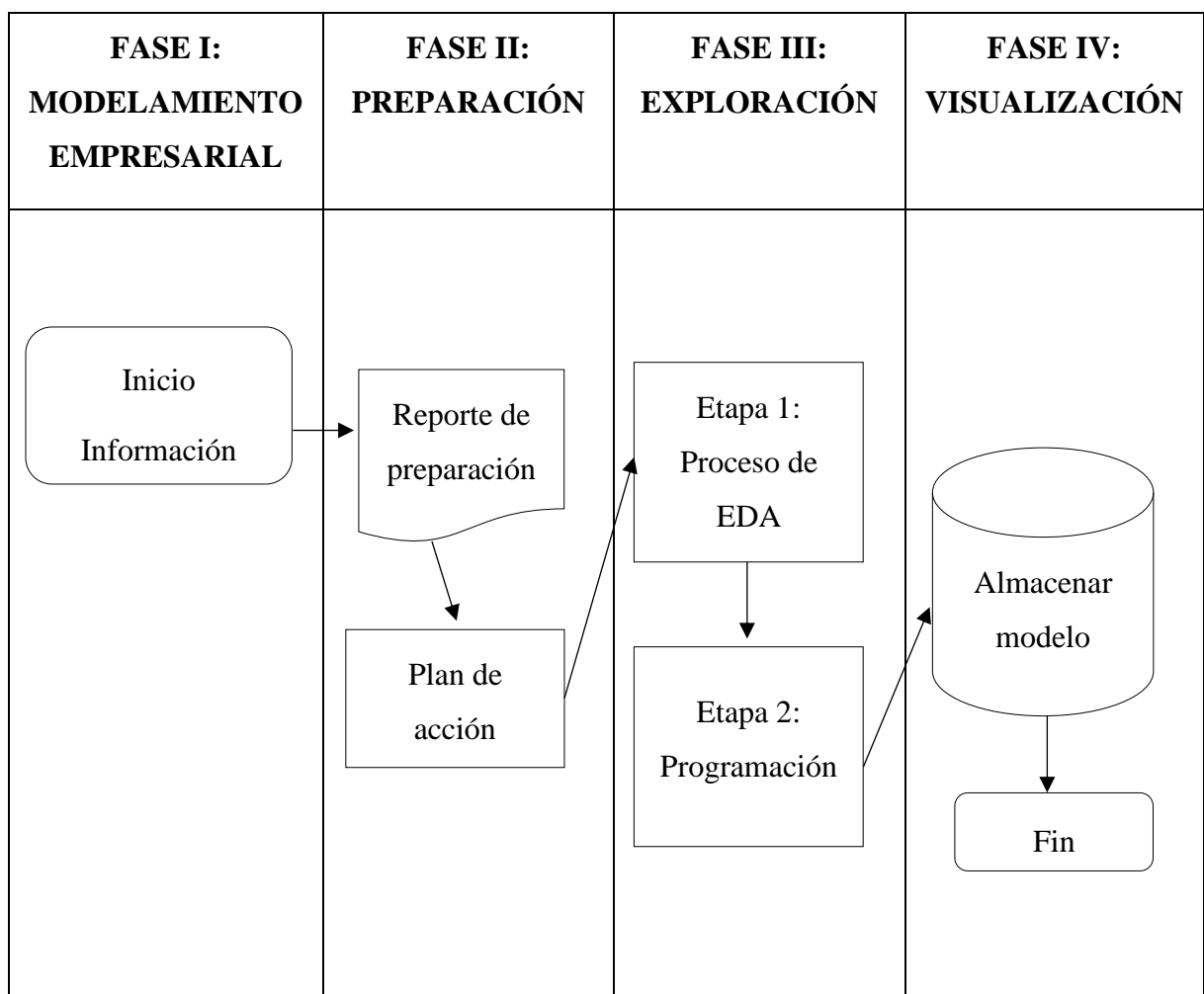
Fig. 9. Burn Down Chart

4..2. Producto Mínimo Viable (MVP)

Para implementar la nueva metodología para optimización de procesos con Machine Learning en especialidades de Ingeniería Industrial es elaborada con la finalidad de mejorar el desarrollo de proyectos relacionados con optimización de procesos, aumentando sus capacidades y habilidades para usar Machine Learning, además, de realizar un monitoreo en la aplicación de esta metodología.

Tabla 9

Cuadro de flujo para la implementación de la nueva metodología CSKT



Fase I: MODELAMIENTO EMPRESARIAL

Es la fase emprendida a preparar el conjunto de datos

1.1. Exploración del negocio: Provee un reporte inicial de los datos facilitados por la empresa que será usada para futuros análisis y desarrollo de planes de acción.

1.2. Explotación de los datos: Como trans fondo del tipo de problema identificado está transformar los datos en conocimiento que se pueda activar, de modo que se comienza a hablar de lenguajes tecnológicos y de negocio para enlazar soluciones, aportando una visión para encontrar ventajas competitivas e innovadoras y en ese sentido, posibilite la apertura a nuevas oportunidades que amplíen el perfil de Ingeniero Industrial, desde su entorno y posibilidad que tengan en resolver los problemas de negocio con Data Analytis.

Fase II: PREPARACIÓN

Es la fase emprendida a introducir el reporte de la preparación de datos.

2.1. Reporte de preparación de datos: Son procesos emprendidos para preparar los datos para su uso. Menciona las bibliotecas para usar, evalúa la estructura de cada uno de los datos que se están usando de los procesos industriales o funcionales. Además, se realiza la preparación de los datos, se evalúa si hay valores únicos y se realizan las distintas limpiezas de los datos, reportando cada uno de los datos, sus especificaciones y las acciones realizadas.

2.2. Plan de acción: Son los procesos que se emprenden para preparar el conjunto de datos, los códigos usados y las soluciones razonables del problema identificado.

Fase III: EXPLORACIÓN

Primera Etapa: Proceso de EDA

Sirve para reportar cada uno de los datos analizados después de la limpieza de datos. Se hace uso de EDA para emprender la exploración de los datos y luego pasar a hacer las visualizaciones.

Segunda Etapa: Programación

Es el conocimiento de los lenguajes de programación, algoritmos y códigos que hace que el programador resuelva problemas con la nueva metodología CSKT basada en las metodologías CRISP-DM, SEMMA, KDD y TDSP.

Fase IV: VISUALIZACIÓN

Es la fase emprendida a obtener visualizaciones de los datos obtenidos usando bibliotecas disponibles para poder encontrar correlaciones en algunos datos específicos de cada uno de los KPIs.

Las fases mencionadas trabajan directamente con una combinación entre CRISP-DM, SEMMA, KDD y TDSP para conformar la nueva metodología CSKT como se ilustra en la siguiente figura, la cual es una representación del Producto Mínimo Viable (MVP) propuesto:

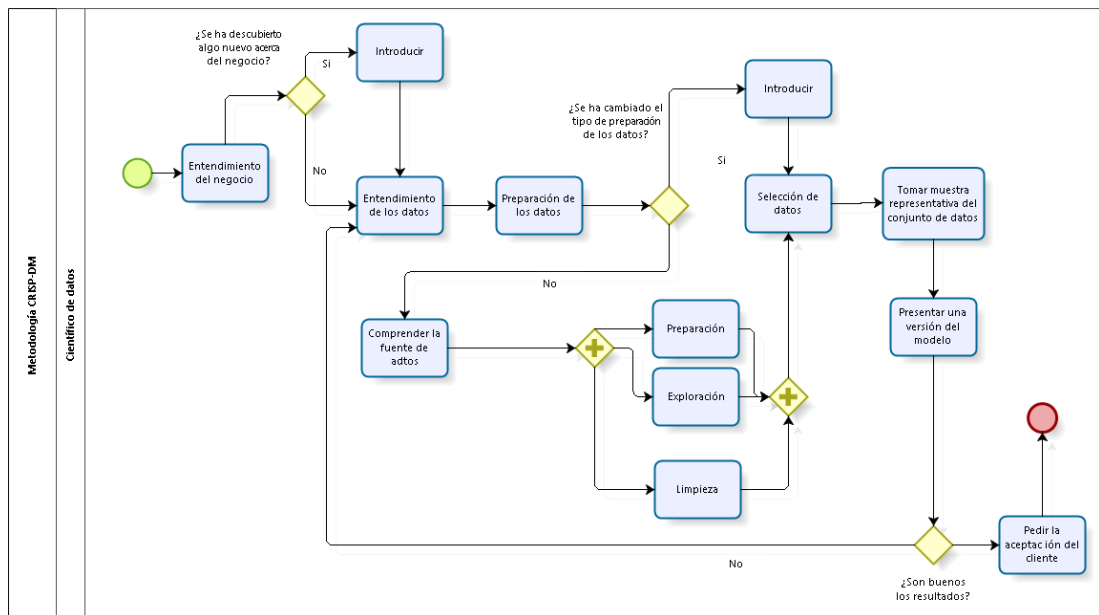


Fig. 10. Diagrama de flujo para la implementación de la metodología CSKT

En definitiva es un cambio cultural, que no es sencillo para el tesista de Ingeniería Industrial que se enfoca en la optimización de procesos, que empiecen a usar la metodología, que está en proceso de desarrollar un análisis del negocio a través de los datos de acuerdo a los escenarios de la carrera Industrial, que tiene el fin de brindar conocimiento para implementar mejores proyectos de optimización de procesos.

5. Implementación de la propuesta

5.1. Planificación y estimación

Como no hay un antecedente sobre esto, la entonces empresa VTT creó una metodología para el desarrollo de aplicaciones móviles por los años 2004, y tiene su página web donde ya está. La página web muestra lo que son las guías de la metodología Mobile-D justamente como parte de su creación, y en esta investigación se presenta esta nueva propuesta de metodología para esa manera para dar facilidades de que esta metodología se encuentre disponible para todo el mundo; por ello la implementación va a consistir en diseñar una nueva metodología CSKT para adecuar los proyectos de implementación de machine learning en ingeniería industrial para que los estudiantes de ingeniería industrial puedan tomar las guías para implementar proyectos de machine learning al ser esta nueva metodología diseñada tomando en cuenta las necesidades, y estas necesidades se realizaron conjuntamente con el desarrollo de la metodología design thinking que se emprendió para tomar las tesis de ingeniería industrial como la muestra para generar una metodología diseñada a sus necesidades o forma de trabajar, para con ello, en otra etapa realizar la metodología Scrum para sobrellevar la gestión de este diseño de la nueva metodología.

Metodología CSKT

- 1) Modelamiento empresarial
 - a. Entendimiento del negocio
- 2) Preparación
 - a. Entendimiento de los datos
 - b. Preparación de los datos
 - c. Tomar muestra del conjunto de datos
- 3) Exploración
 - a. Selección de datos
 - b. Comprender fuente de datos
 - c. Preparación
 - d. Exploración
 - e. Limpieza
- 4) Visualización
 - a. Presentar versión de modelo
 - b. Pedir aceptación del cliente

Fig. 11. Metodología CSKT

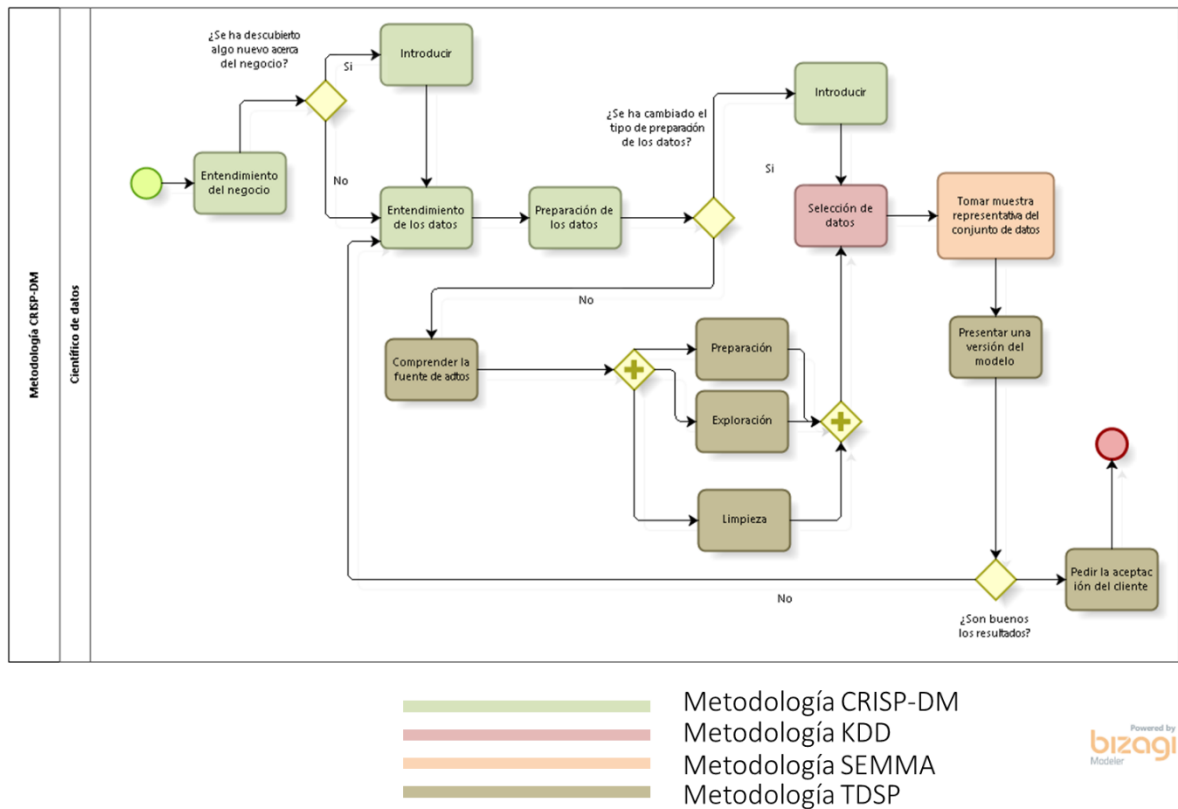


Fig. 12. Planificación del Proyecto - Metodología CSKT

La técnica para estimar la implementación, entre las diferentes técnicas, se ha usado el Software COCOMO II para estimar el costo total del proyecto.

5..2. Implementación

La siguiente información es un recopilado de cada una de las metodologías que se usarán para establecer la nueva metodología CSKT, esto desde un punto de vista más detallado donde se especifica cada paso y actividad de las metodologías base para el desarrollo de esta investigación: CRISP-DM, SEMMA, KDD y TDSP.

Entendimiento del negocio

1.1. Determinar los objetivos empresariales

| | |
|--------------|--|
| Tarea | <p>Determinar los objetivos empresariales</p> <p>El primer objetivo del analista de datos es comprender a fondo, desde una perspectiva empresarial, lo que el cliente realmente quiere conseguir. A menudo, el cliente tiene muchos objetivos y limitaciones que compiten entre sí y que deben equilibrarse adecuadamente. El objetivo del analista es descubrir, desde el principio, los factores importantes que pueden</p> |
|--------------|--|

| | |
|----------------|---|
| | <p>influir en el resultado del proyecto. Una posible consecuencia de descuidar este paso es gastar una gran cantidad de esfuerzo produciendo las respuestas correctas a las preguntas equivocadas.</p> |
| Salidas | <p>Antecedentes</p> <p>Registre la información que se conoce sobre la situación empresarial de la organización situación de la empresa al inicio del proyecto.</p> |
| | <p>Objetivos empresariales</p> <p>Describa el objetivo principal del cliente, desde una perspectiva empresarial. Además del objetivo empresarial principal, suele haber otras cuestiones empresariales relacionadas que el cliente desea abordar. Por ejemplo, el objetivo empresarial principal podría ser mantener a los clientes actuales prediciendo cuándo son propensos a cambiarse a un competidor. Ejemplos de preguntas de negocio relacionadas son: "¿Cómo funciona el canal principal (por ejemplo, el cajero automático)? principal (por ejemplo, cajero automático, visita a la sucursal, Internet) que utiliza un cliente del banco afecta o "¿Reducirán las comisiones de los cajeros automáticos el número de clientes el número de clientes de alto valor que se van?".</p> |
| | <p>Criterios de éxito de la empresa</p> <p>Describa los criterios para un resultado exitoso o útil del proyecto desde el punto de vista empresarial. Esto puede ser bastante específico y capaz de medir objetivamente, como la reducción de la pérdida de clientes a un nivel, o general y subjetivo, como "dar una visión útil de las relaciones" las relaciones". En este último caso debe indicarse quién hace el juicio subjetivo.</p> |

1.2. Evaluar la situación

| | |
|----------------|---|
| Tarea | <p>Evaluar la situación</p> <p>Esta tarea implica una investigación más detallada de todos los recursos, limitaciones, suposiciones y otros factores que deben tenerse en cuenta para determinar el objetivo del análisis de datos y el plan del proyecto. En la tarea anterior, su objetivo es llegar rápidamente al meollo de la situación. Aquí quiere concretar los detalles.</p> |
| Salidas | <p>Inventario de recursos</p> <p>Enumere los recursos disponibles para el proyecto, incluyendo: personal (expertos en negocios, expertos en datos, apoyo técnico, personal de minería de datos), datos (extractos fijos, acceso a datos vivos almacenados u operativos), recursos informáticos (plataformas de hardware) y software (herramientas de datos, otros programas informáticos pertinentes).</p> |

| | |
|--|--|
| | <p>Requisitos, supuestos y limitaciones</p> <p>Enumere todos los requisitos del proyecto, incluido el calendario de ejecución, la comprensibilidad y la calidad de los resultados y la seguridad, así como las cuestiones legales. Como parte de este resultado, asegúrate de que se te permite utilizar los datos.</p> <p>Enumere los supuestos del proyecto. Pueden ser suposiciones sobre los datos que pueden comprobarse durante la extracción de datos, pero también pueden incluir suposiciones no comprobables sobre el negocio en el que se basa el proyecto. Es especialmente importante enumerar estos últimos si constituyen condiciones de validez de los resultados.</p> <p>Enumere las limitaciones del proyecto. Puede tratarse de limitaciones en cuanto a la disponibilidad de recursos, pero también pueden ser limitaciones tecnológicas como el tamaño de los datos que es práctico utilizar para la modelización.</p> |
| | <p>Riesgos y contingencias</p> <p>Enumere los riesgos o acontecimientos que pueden retrasar el proyecto o hacer que el proyecto fracase. Enumere los correspondientes planes de contingencia; qué medidas se tomarán en caso de que se produzcan los riesgos.</p> |
| | <p>Terminología</p> <p>Elabore un glosario con la terminología relevante para el proyecto. Esto puede incluir dos componentes:</p> <ol style="list-style-type: none"> a. Un glosario de terminología empresarial relevante, que forma parte de la comprensión empresarial disponible para el proyecto. La elaboración de este glosario es un ejercicio útil de "obtención de conocimientos" y de formación. b. Un glosario de terminología de minería de datos, ilustrado con ejemplos pertinentes para el problema empresarial en cuestión. |
| | <p>Costes y beneficios</p> <p>Construir un análisis coste-beneficio para el proyecto, que compare los costes del proyecto con el beneficio potencial para la empresa si tiene éxito. éxito. La comparación debe ser lo más específica posible, por ejemplo utilizando medidas monetarias en una situación comercial.</p> |

1.3. Determinar los objetivos de la minería de datos

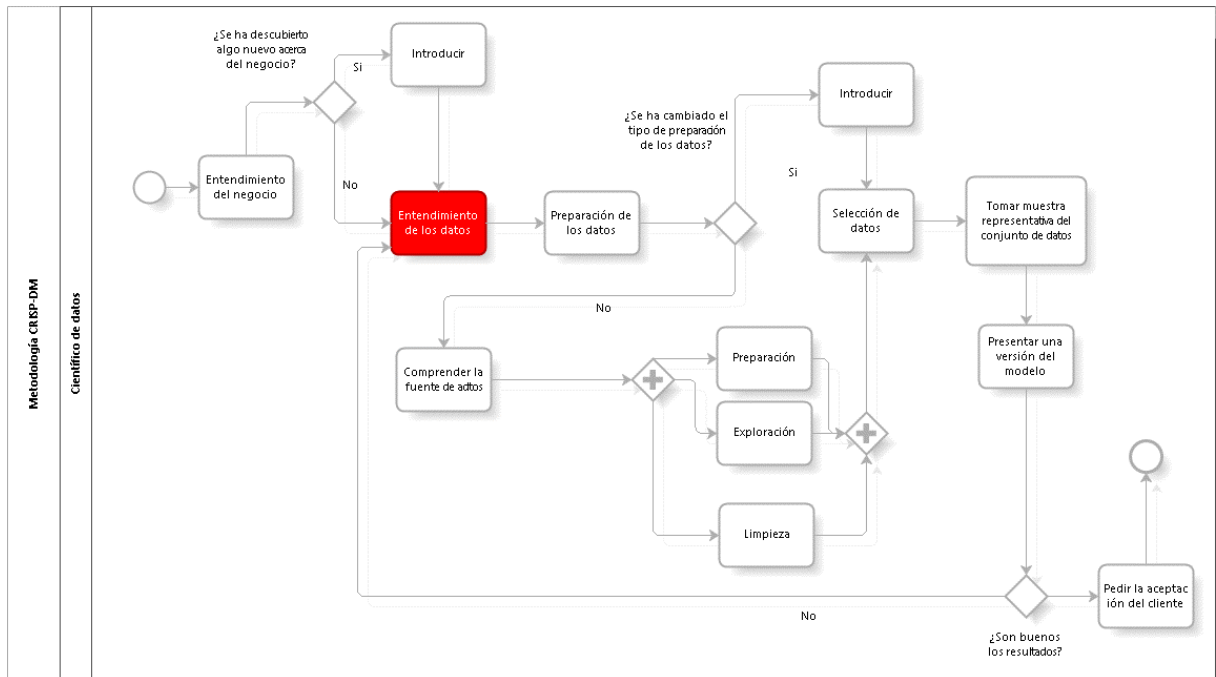
| | |
|---------------------|---|
| <p>Tarea</p> | <p>Determinar los objetivos de la minería de datos</p> <p>Un objetivo empresarial establece los objetivos en terminología empresarial. Un objetivo de minería de datos establece los objetivos del proyecto en términos técnicos. Por ejemplo, el objetivo empresarial podría ser "Aumentar las ventas por catálogo a los clientes existentes". Un objetivo de minería de datos podría ser "Predecir cuántos</p> |
|---------------------|---|

| | |
|---------|--|
| | widgets comparará un cliente comprarán, dadas sus compras de los últimos tres años, la información demográfica información demográfica (edad, salario, ciudad, etc.) y el precio del artículo". |
| Salidas | Objetivos de la minería de datos Describa los resultados previstos del proyecto que permiten la alcanzar los objetivos de la empresa. |
| | Criterios de éxito de la extracción de datos Definir los criterios de éxito del proyecto en términos técnicos términos técnicos, por ejemplo, un determinado nivel de precisión predictiva o una propensión al perfil de compra con un determinado grado de "elevación". Al igual que con los criterios de éxito empresarial puede ser necesario describirlos en términos subjetivos, en cuyo caso En este caso, la persona o personas que emiten el juicio subjetivo deben ser identificados. |

1.4. Elaborar el plan del proyecto

| | |
|---------|--|
| Tarea | Elaborar el plan del proyecto Describa el plan previsto para alcanzar los objetivos de la minería de datos y alcanzar los objetivos de la empresa. El plan debe especificar el conjunto previsto de pasos a realizar durante el resto del proyecto incluyendo una selección inicial de herramientas y técnicas. |
| Salidas | Plan del proyecto Enumerar las etapas que se van a ejecutar en el proyecto, junto con la duración, los recursos necesarios, las entradas, las salidas y las dependencias. Cuando sea posible, explicita las iteraciones a gran escala en el proceso de extracción de datos, por ejemplo, las repeticiones de las fases de modelización y evaluación. Como parte del plan del proyecto, también es importante analizar las dependencias entre el calendario y los riesgos. Marque los resultados de estos análisis explícitamente en el plan del proyecto, idealmente con acciones y recomendaciones si los riesgos aparecen. Nota: el plan del proyecto contiene planes detallados para cada fase. Por ejemplo, decida en este momento qué estrategia de evaluación se utilizará en la fase de evaluación. El plan del proyecto es un documento dinámico en el sentido de que, al final de cada fase de cada fase es necesario revisar los progresos y los logros y se recomienda una actualización del plan del proyecto en consecuencia. Los puntos de revisión específicos de Los puntos de revisión específicos para estas revisiones también forman parte del plan del proyecto. |
| | Evaluación inicial de las herramientas y técnicas |

Al final de la primera fase, el proyecto también realiza una evaluación inicial de herramientas y técnicas. Aquí se selecciona una herramienta de minería de datos que soporte varios métodos para las diferentes etapas del proceso, por ejemplo. Es importante evaluar las herramientas y técnicas al principio del proceso, ya que la selección de herramientas y técnicas posiblemente influya en todo el proyecto.



Entendimiento de los datos

2.1. Recoger los datos iniciales

| | |
|-----------------------|---|
| <p>Tarea</p> | <p>Recoger los datos iniciales</p> <p>Adquirir dentro del proyecto los datos (o el acceso a los datos) que figuran en los recursos del proyecto. Esta recopilación inicial incluye la carga de datos si es necesario para la comprensión de los datos. Por ejemplo, si se aplica una herramienta específica para la comprensión de los datos, tiene mucho sentido cargar los datos en esta herramienta. Este esfuerzo posiblemente conduzca a los pasos iniciales de preparación de los datos.</p> <p>Nota: si adquiere múltiples fuentes de datos, la integración es una cuestión adicional de datos, la integración es una cuestión adicional, ya sea aquí o en la fase posterior de preparación de los datos.</p> |
| <p>Salidas</p> | <p>Informe de recogida de datos iniciales</p> <p>Enumere el conjunto o conjuntos de datos adquiridos, junto con su ubicación dentro del proyecto, los métodos utilizados para adquirirlos y los problemas encontrados. Registre</p> |

| | |
|--|--|
| | los problemas encontrados y las soluciones logradas para facilitar la repetición de este proyecto en el futuro o la ejecución de proyectos similares. proyectos futuros similares. |
|--|--|

2.2. Describir los datos

| | |
|----------------|--|
| Tarea | Describir los datos Examinar las propiedades "brutas" o "superficiales" de los datos adquiridos e informar de los resultados. |
| Salidas | Informe de descripción de datos Describa los datos que se han adquirido, incluyendo: el formato de la cantidad de datos, por ejemplo, el número de registros y campos en cada tabla, las identidades de los campos y cualquier otra característica superficial de los datos que se han descubierto. ¿Los datos adquiridos satisfacen los requisitos pertinentes? |

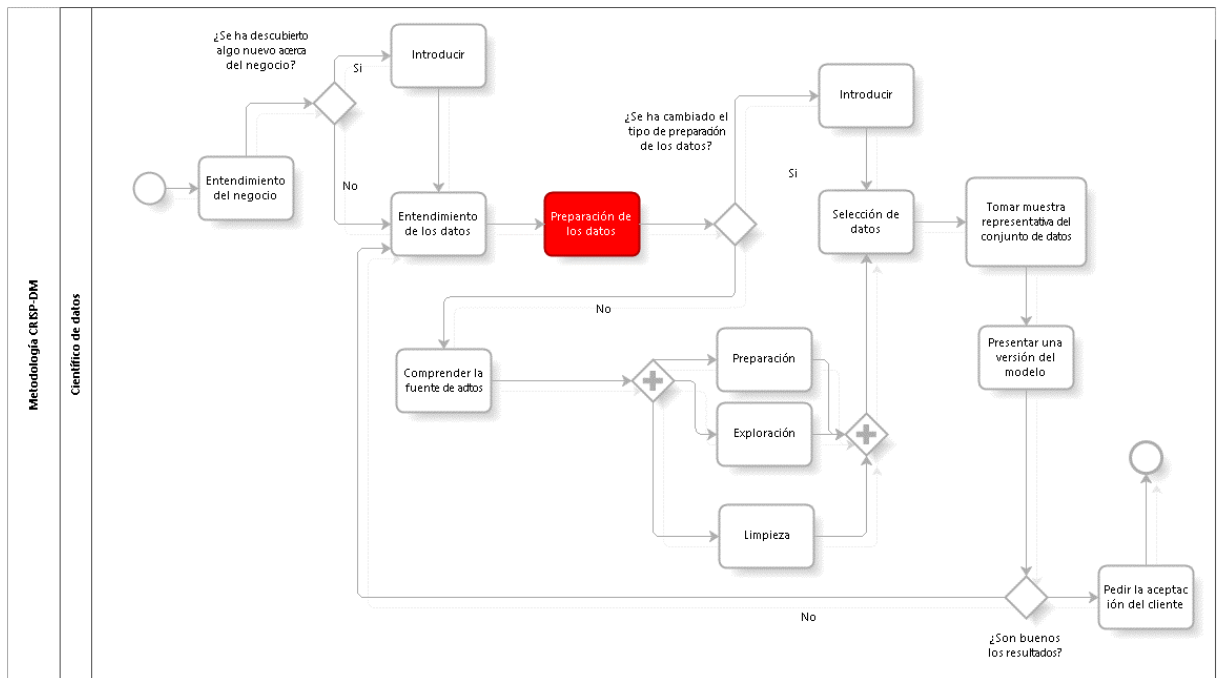
2.3. Explorar los datos

| | |
|----------------|--|
| Tarea | Explorar los datos Esta tarea aborda las cuestiones de la minería de datos, que se pueden abordar mediante la consulta, la visualización y la elaboración de informes. Entre ellas se encuentran: la distribución de los atributos clave, por ejemplo, el atributo objetivo de una tarea de predicción; las relaciones entre pares o pequeños números de atributos; los resultados de agregaciones sencillas; las propiedades de subpoblaciones significativas; los análisis estadísticos sencillos. Estos análisis pueden abordar directamente los objetivos de la minería de datos; también pueden contribuir a la descripción de los datos y a los informes de calidad, o perfeccionarlos, y alimentar la transformación. y los informes de calidad y alimentar la transformación y otros preparativos de los datos de datos necesarios para el análisis posterior. |
| Salidas | Informe de exploración de datos Describa los resultados de esta tarea, incluyendo las primeras conclusiones o hipótesis iniciales y su impacto en el resto del proyecto. Si procede, incluya gráficos y diagramas que indiquen las características de los datos o conduzcan a subconjuntos de datos interesantes para un examen más profundo. |

2.4. Verificar la calidad de los datos

| | |
|--------------|--|
| Tarea | Verificar la calidad de los datos |
|--------------|--|

| | |
|----------------|--|
| | Examinar la calidad de los datos, abordando cuestiones como: ¿son los datos completos (cubren todos los casos necesarios)? ¿Son correctos o ¿contiene errores y, si los hay, cuán comunes son? ¿Faltan valores en los datos? En caso afirmativo, ¿cómo se representan? ¿dónde se producen y qué frecuencia tienen? |
| Salidas | Informe sobre la calidad de los datos Enumere los resultados de la verificación de la calidad de los datos; si existen problemas de calidad Si existen problemas de calidad, enumere las posibles soluciones. Las soluciones a los problemas de calidad de los datos generalmente dependen en gran medida del conocimiento de los datos y del negocio. |



Preparación de los datos

3.1. Seleccione los datos

| | |
|----------------|---|
| Tarea | Seleccione los datos Decidir los datos que se utilizarán para el análisis. Los criterios incluyen la relevancia para los objetivos de la minería de datos, la calidad y las restricciones técnicas, como límites en el volumen o los tipos de datos. Tenga en cuenta que la selección de datos abarca selección de atributos (columnas), así como la selección de registros (filas) en una tabla. |
| Salidas | Justificación de la inclusión/exclusión Enumere los datos que se van a incluir/excluir y los motivos de estas decisiones. |

3.2. Limpiar los datos

| | |
|----------------|--|
| Tarea | <p>Limpiar los datos</p> <p>Elevar la calidad de los datos al nivel requerido por las técnicas de análisis seleccionadas. Esto puede implicar la selección de subconjuntos limpios de los datos, la inserción de valores predeterminados adecuados o técnicas más ambiciosas como la estimación de los datos que faltan mediante modelización.</p> |
| Salidas | <p>Informe de limpieza de datos</p> <p>Describe qué decisiones y acciones se tomaron para solucionar los problemas de calidad de los datos notificados durante la tarea de verificación de la fase de comprensión de los datos. Las transformaciones de los datos para su limpieza y el posible impacto en los resultados del análisis. en los resultados del análisis.</p> |

3.3. Construir los datos

| | |
|----------------|--|
| Tarea | <p>Construir los datos</p> <p>Esta tarea incluye operaciones constructivas de preparación de datos como la producción de atributos derivados, nuevos registros completos o valores transformados para los atributos existentes.</p> |
| Salidas | <p>Atributos derivados</p> <p>Los atributos derivados son nuevos atributos que se construyen a partir de uno o varios atributos existentes en el mismo registro. Ejemplos: $\text{área} = \text{longitud} * \text{anchura}$.</p> |
| | <p>Registros generados</p> <p>Describe la creación de registros completamente nuevos. Ejemplo: crear registros para los clientes que no han realizado ninguna compra durante el último año. No había ninguna razón para tener estos registros en los datos brutos, pero para la modelización, podría tener sentido representar explícitamente el hecho que ciertos clientes no hicieron ninguna compra.</p> |

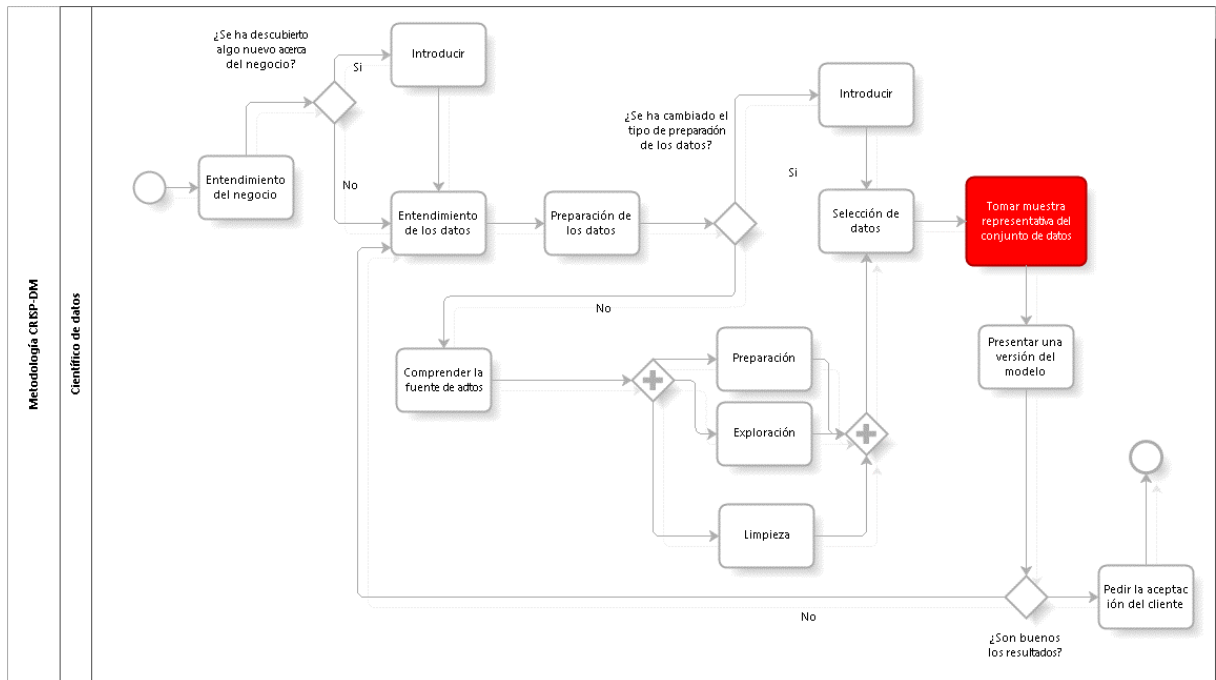
3.4. Integrar datos

| | |
|----------------|---|
| Tarea | <p>Integrar datos</p> <p>Son métodos por los que se combina la información de varias tablas o registros para crear nuevos registros o valores.</p> |
| Salidas | <p>Fusionar datos</p> |

| | |
|--|--|
| | <p>La fusión de tablas se refiere a la unión de dos o más tablas que tienen información diferente sobre los mismos objetos. Ejemplo: una cadena de tiendas tiene una tabla con información sobre las características generales de cada tienda (por ejemplo, superficie, tipo de centro comercial), otra tabla con datos resumidos sobre las ventas de ventas resumidas (por ejemplo, beneficios, porcentaje de variación de las ventas respecto al año anterior) y otra con información sobre la demografía de la zona circundante. alrededores. Cada una de estas tablas contiene un registro por cada tienda. Estas tablas pueden fusionarse en una nueva tabla con un registro para cada un registro para cada tienda, combinando los campos de las tablas de origen.</p> <p>Los datos fusionados también abarcan las agregaciones. La agregación se refiere a las operaciones en las que los nuevos valores se calculan resumiendo la información de múltiples registros y/o tablas. Por ejemplo, convertir una tabla de compras de clientes en la que hay un registro por cada compra en una nueva tabla en la que hay un registro para cada cliente, con campos como como el número de compras, el importe medio de las compras, el porcentaje de pedidos cargados a la tarjeta de crédito, porcentaje de artículos en promoción, etc.</p> |
|--|--|

3.5. Formatear datos

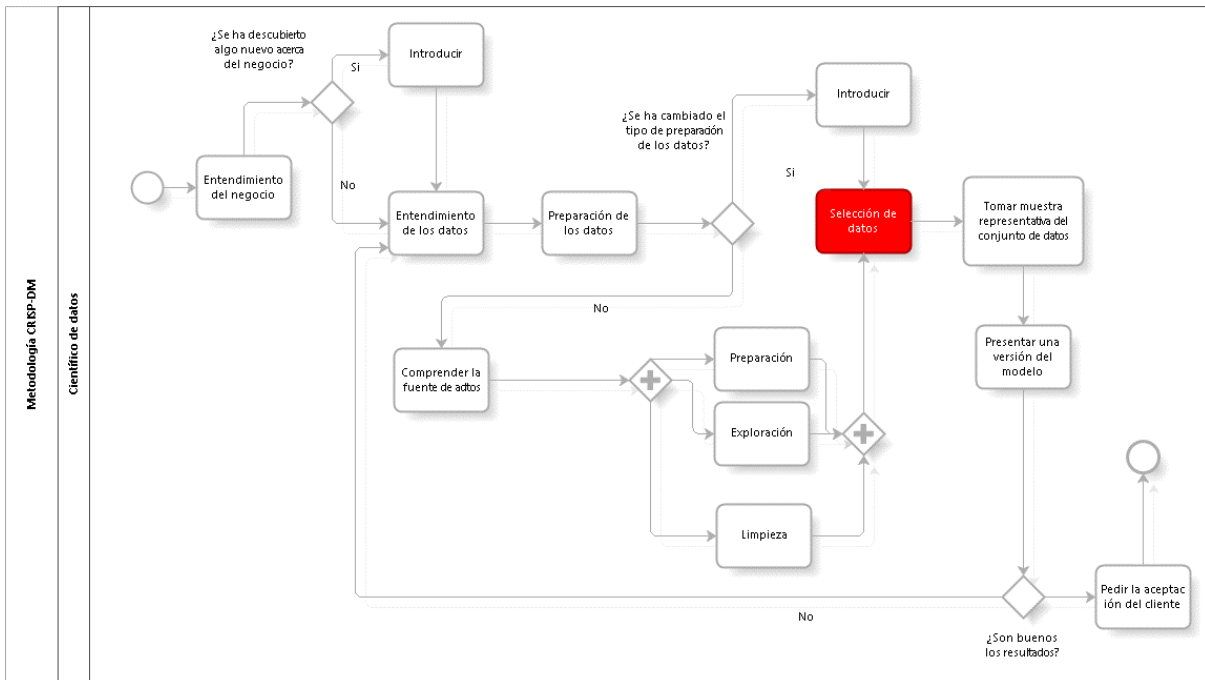
| | |
|----------------|--|
| Tarea | Formatear datos Las transformaciones de formato se refieren principalmente a las modificaciones sintácticas que se hacen a los datos y que no cambian su significado, pero que pueden ser requeridas por la herramienta de modelado. |
| Salidas | Datos reformateados Algunas herramientas tienen requisitos sobre el orden de los atributos, como por ejemplo que el primer campo sea un identificador único para cada registro o que el último campo es el campo de resultado que el modelo debe predecir. Puede ser importante cambiar el orden de los registros en el conjunto de datos. Tal vez la herramienta de modelización requiera que los registros se ordenen según el valor del atributo de resultado. según el valor del atributo de resultado. Una situación común es que los registros del conjunto de datos estén inicialmente ordenados de alguna manera, pero el algoritmo de modelización necesita que estén en un orden bastante aleatorio. Por ejemplo, cuando se utilizan Cuando se utilizan redes neuronales, suele ser mejor que los registros se presenten en un orden aleatorio, aunque algunas herramientas lo gestionan automáticamente sin la intervención explícita del usuario. |



Tomar muestra representativa del conjunto de datos

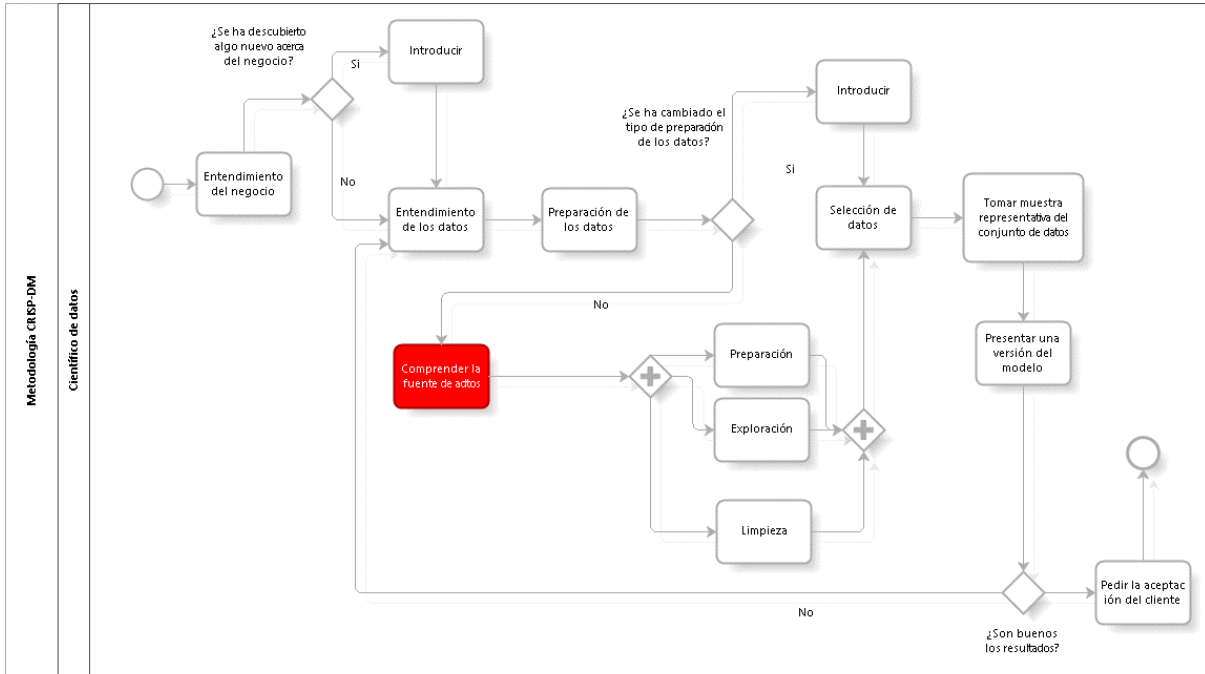
| | |
|-----------------------|---|
| <p>Tarea</p> | <p>Tomar muestra del conjunto de datos</p> <p>Estos identifican, fusionan, dividen y muestrean conjuntos de datos de entrada, entre otras tareas.</p> |
| <p>Salidas</p> | <p>Agregar</p> <p>Los datos que se importan deben estar en un diagrama de flujo de proceso. De acuerdo con la importancia de los datos, se deben unir las fuentes de datos en este diagrama de flujo para ser unidos en un entrenamiento con datos de las transacciones, de puntaje, etc. Estos datos se deben particionar previamente en roles de prueba y validación en un conjunto de datos de entrenamiento.</p> <p>Partición de datos</p> <p>Los datos particionados deben tener datos de prueba, de validación y de entrenamiento, los cuales se utilizarán para el ajuste preliminar del modelo de machine learning. El conjunto de datos de validación se usa para monitorear y ajustar los pesos del modelo durante la estimación y también se usa para evaluar el modelo. El conjunto de datos de prueba es un conjunto de datos reservado para la evaluación del modelo. La partición de datos utiliza un muestreo aleatorio simple, muestreo aleatorio estratificado definidas por el usuario (experto cuantitativo) para crear el conjunto de datos particionados.</p> <p>Importación de archivos</p> <p>La importación de los archivos con la debida consideración del diagrama de flujo permite almacenar los datos de externos a una fuente de datos que se puede interpretar para el ajuste preliminar del modelo de machine learning. La importación de archivos puede</p> |

| | |
|--|---|
| | soportar archivos planos CSV, tablas de hojas de cálculo de Microsoft Excel y tablas de bases de datos. |
| | <p>Filtro</p> <p>Se les aplica un filtro a estos archivos importados para obtener un conjunto de datos de entrenamiento filtrados para excluir valores atípicos u otras observaciones de las que no se desea incluir al análisis de minería de datos. Los valores atípicos pueden afectar en gran medida los resultados del modelado y posteriormente, la precisión y confiabilidad de los modelos entrenados.</p> |
| | <p>Datos de entrada</p> <p>Los datos de entrada le permiten que los conjuntos de datos y otros tipos de datos sean accedidos. Los datos de entrada representan la introducción a los metadatos predefinidos en un espacio de trabajo para su procesamiento. Puede ver información de los metadatos sobre sus datos de origen por los datos de entrada, como valores iniciales de medición y función de modelo de cada variable para el ajuste preliminar del modelo de machine learning.</p> |
| | <p>Fusionar</p> <p>La fusión le permite fusionar observaciones de dos o más conjuntos de datos en una sola emprendiendo un nuevo conjunto de datos. La fusión combina uno a uno, además, en ello se cambian el nombre de ciertas variables (por ejemplo, valores pronosticados y probabilidades posteriores) según la configuración para el ajuste preliminar del modelo de machine learning.</p> |
| | <p>Muestra</p> <p>La muestra le permite tomar muestras aleatorias simples, estratificadas y de conglomerados del conjunto de datos. Se recomienda el muestro para bases de datos extremadamente grandes porque puede reducir significativamente el tiempo de entrenamiento del modelo. Si la muestra es suficientemente representativa (1000 registros como mínimo), se puede esperar que las relaciones encontradas en la muestra se generalicen hacia un conjunto de datos completo y óptimo para el ajuste preliminar del modelo de machine learning.</p> |



Selección de datos

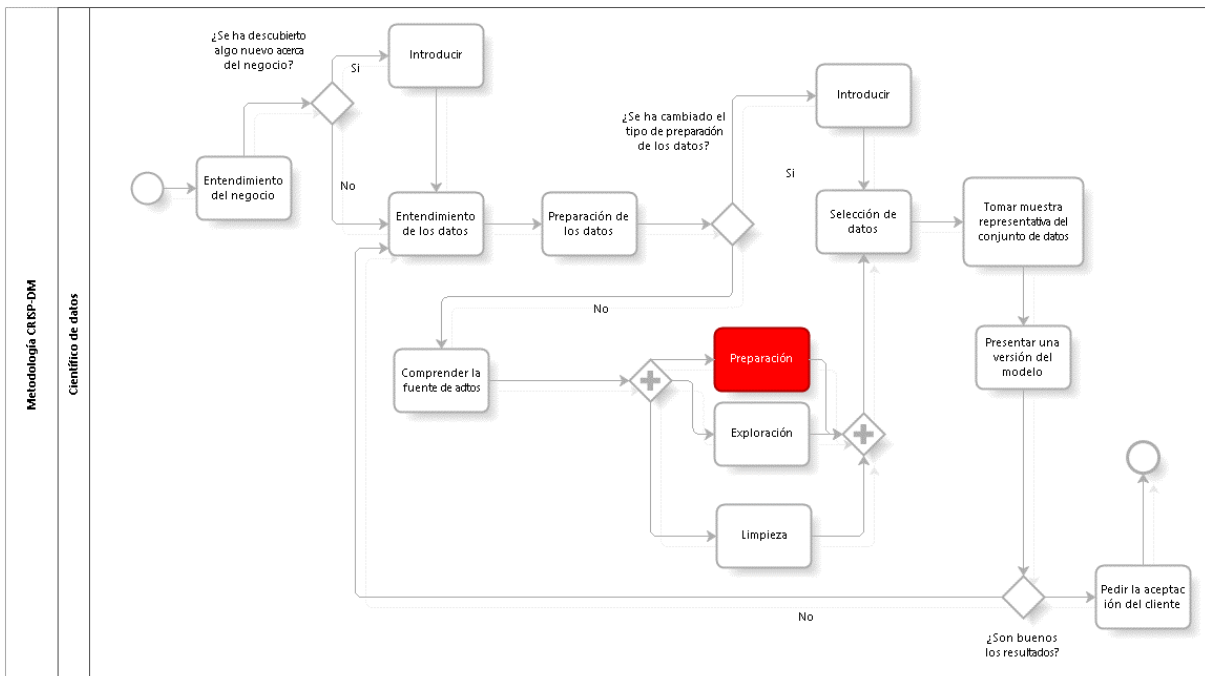
| | |
|-----------------------|--|
| <p>Tarea</p> | <p>Selección de datos</p> <p>Se realiza una selección de aquellos datos que son relevantes para el proceso de extracción de conocimiento que se desea realizar.</p> |
| <p>Salidas</p> | <p>Reducción y proyección de los datos</p> <p>En este paso se detectan características útiles de representación de los datos dependiendo del objetivo de la tarea de minería (descripción o predicción). Se incluye la utilización de técnicas de reducción de la dimensionalidad y métodos de transformación de los datos para reducir la cantidad de variables en discusión o para encontrar representaciones invariantes de los datos. En esta etapa es frecuente la transformación de los datos, calculando nuevos atributos o bien redefiniendo los existentes con otro formato.</p> <p>Selección del algoritmo para extracción de datos</p> <p>La selección de datos se define como el proceso en el que los datos relevantes para el análisis se deciden y recuperan de la recopilación de datos.</p> <p>Selección de datos mediante red neuronal.</p> <p>Selección de datos mediante árboles de decisión.</p> <p>Selección de datos mediante Naive bayes.</p> <p>Selección de datos mediante Clustering, Regresión, etc.</p> |



Comprender la fuente de datos

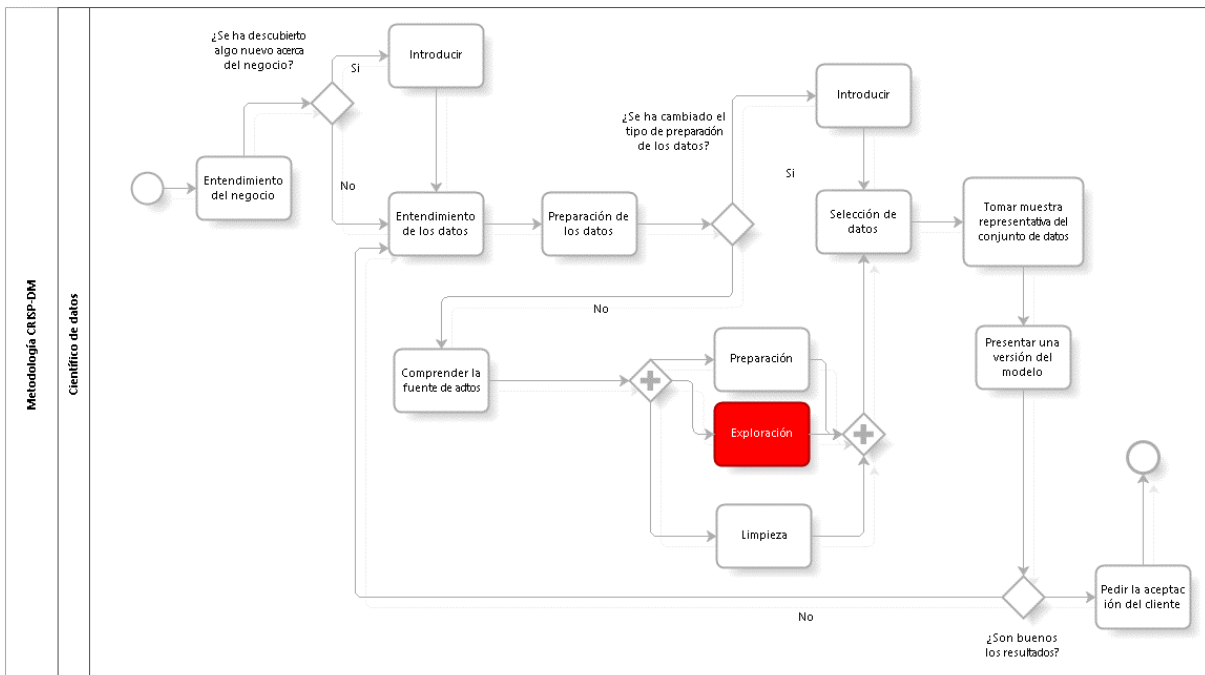
| | |
|-----------------------|---|
| <p>Tarea</p> | <p>Comprender la fuente de datos</p> <p>Se realiza la plantilla para la definición de estructura del proyecto de CSKT, dicha plantilla se actualiza a lo largo del proyecto.</p> |
| <p>Salidas</p> | <p>Identificar el origen de datos</p> <p>Identificar el origen de los datos que contienen ejemplos conocidos de respuestas a las preguntas certeras. Por ejemplo:</p> <ul style="list-style-type: none"> ▪ Datos relacionados para la pregunta ¿Tenemos indicadores para medir el objetivo y las características que están relacionadas a él? ▪ Datos que representen una medida precisa del objetivo del modelo y las características de interés. <p>Se especifica la ubicación original y de destino de los datos sin procesar.</p> <p>Se provee el diccionario de datos: Documento que provee de las descripciones de los datos facilitados por el cliente. Por ejemplo: tipos de datos, información sobre las reglas de validación, etc.</p> |
| | <p>Selección del algoritmo para extracción de datos</p> |

| | |
|--|--|
| | <p>La selección de datos se define como el proceso en el que los datos relevantes para el análisis se deciden y recuperan de la recopilación de datos.</p> <p>Selección de datos mediante red neuronal.</p> <p>Selección de datos mediante árboles de decisión.</p> <p>Selección de datos mediante Naive bayes.</p> <p>Selección de datos mediante Clustering, Regresión, etc.</p> |
| | <p>Aprovisionar infraestructura de datos</p> <p>La infraestructura de análisis y almacenamiento, donde se almacenan los conjuntos de datos sin procesar y procesados, puede estar en la nube o en las instalaciones. Esta infraestructura permite un análisis reproducible. También evita la duplicación, que puede dar lugar a incoherencias y costes de infraestructura innecesarios. Se proporcionan herramientas para aprovisionar los recursos compartidos, rastrearlos y permitir que cada miembro del equipo se conecte a esos recursos de forma segura. También es una buena práctica que los miembros del proyecto creen un entorno informático coherente. Luego, diferentes miembros del equipo pueden replicar y validar los experimentos.</p> |



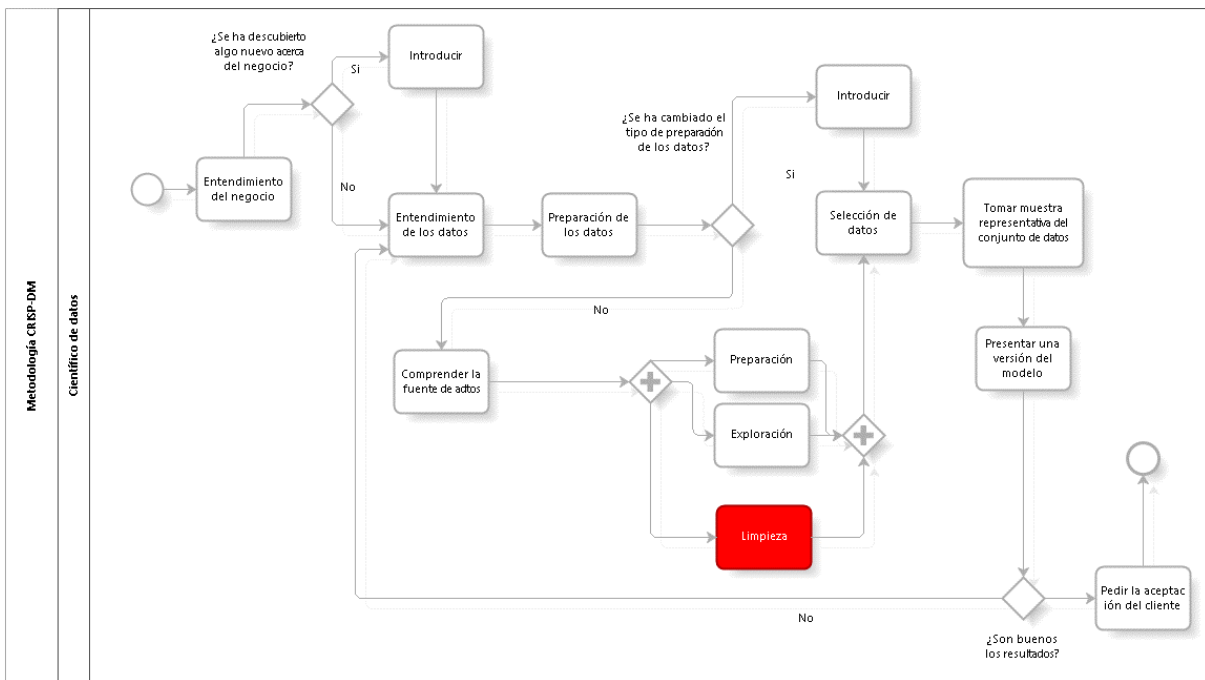
Preparación

| | |
|-----------------------|---|
| <p>Tarea</p> | <p>Preparación</p> <p>Configuración del proceso para mover los datos desde las ubicaciones de origen a las ubicaciones de destino donde se ejecutan las operaciones de análisis.</p> |
| <p>Salidas</p> | <p>Instalar Python</p> <p>Antes de llegar a la exploración de datos, primero debemos asegurarnos de que tiene Python instalado correctamente y que lo está usando en el entorno que terminará usándolo para varias tareas y proyectos.</p> <p>El primer paso es instalar un paquete de software llamado Anaconda o Anaconda Navigator. Puede ir al sitio web de anaconda, descargar la edición individual y asegurarse de instalar la versión python tres de la herramienta. Una vez que lo haya instalado, puede iniciar Anaconda Navigator y ejecutar Jupyter Notebook, lo que le permitirá ejecutar un intérprete de python en tiempo real dentro de su navegador web. Python proporciona una serie de bibliotecas de datos muy importantes para ingerir y explorar datos. Es posible que no termine usándolos tanto directamente, aunque definitivamente usará bibliotecas construidas sobre esas bibliotecas, incluidos pandas, que le permiten manipular estructuras de datos y proporciona herramientas para analizar y disputar la línea de la tabla de datos por ello puede usar Colab que es especialmente adecuado para tareas de aprendizaje automático, análisis de datos y educación que proporciona cuadernos que en su mayoría están construido de la inspiración de Python en particular.</p> |



Exploración

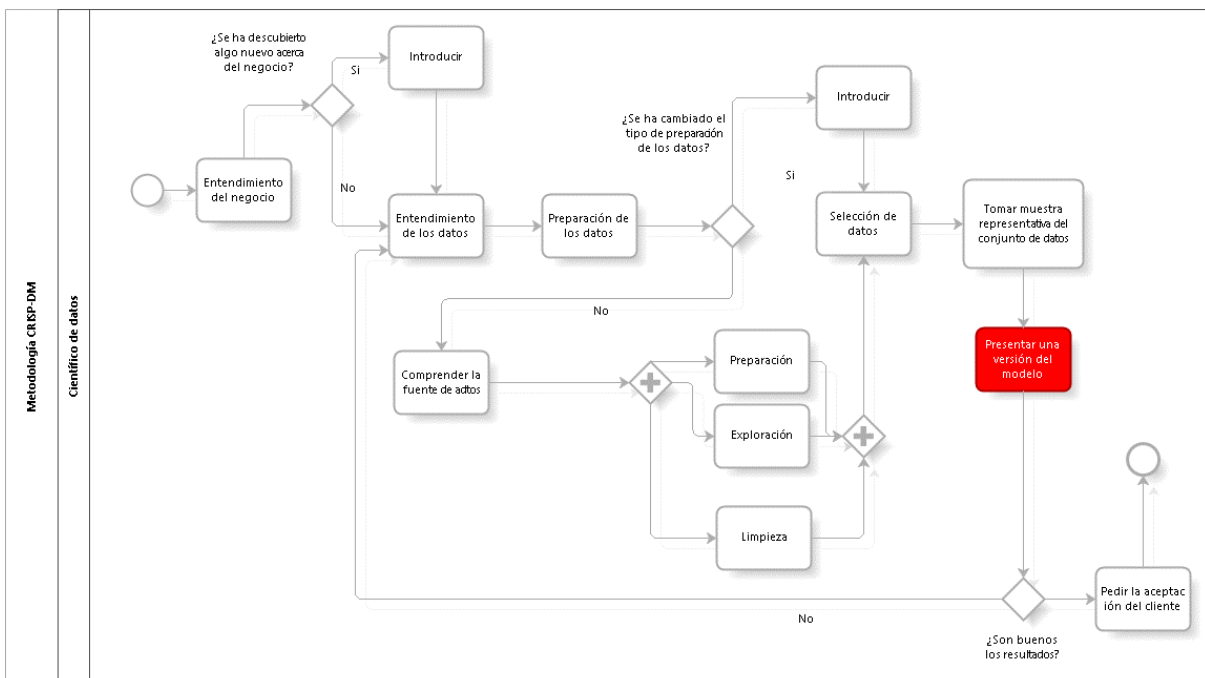
| | |
|----------------|--|
| Tarea | Exploración Debido a que comúnmente los conjuntos de datos reales tienen ruido, les faltan datos o tienen diversas discrepancias es necesario utilizar funciones de resumen y visualización de los datos para auditar su calidad y dar la información necesaria para procesarlos y dejarlos listos para el modelado, posteriormente se debe elegir un modelo de predicción adecuado para el destino y poder avanzar con los siguientes pasos del modelado. |
| Salidas | Comprender valores atípicos Una de las cosas más importantes que se deben hacer con el conjunto de datos antes de colocarlo ciegamente en un modelo es explorarlo para comprender qué hay en los datos. Comprender si hay varios valores atípicos y buscar varios tipos de anomalías. Las anomalías pueden incluir datos faltantes, datos con valores cero, datos con números negativos, etc. A veces esos valores atípicos representan errores en los datos y otras veces representan eventos interesantes. Pero independientemente, cualquier tipo de valor atípico es algo que, una vez que lo introduce en el modelo, puede afectar en última instancia lo que el modelo está aprendiendo. |



Limpeza

| | |
|----------------|--|
| Tarea | Limpeza Proceso para puntuar los datos nuevos o actualizarlos con regularidad durante el proceso del aprendizaje continuo. |
| Salidas | Informe de la calidad de los datos |

| | |
|--|---|
| | <p>Proporcionamos plantillas para la estructura de carpetas y los documentos requeridos en ubicaciones estándar. Esta estructura de carpetas organiza los archivos que contienen código para la exploración de datos y la extracción de características, y que registran las iteraciones del modelo. Estas plantillas facilitan que los miembros del equipo comprendan el trabajo realizado por otros y agreguen nuevos miembros a los equipos.</p> |
| | <p>Arquitectura de la solución</p> <p>Diagrama o descripción de la canalización de datos que utiliza para llevar a cabo la tarea de puntuación o las predicciones con los nuevos datos después de que se haya compilado el modelo. Según las necesidades de su negocio y las limitaciones de sus sistemas existentes en los que se integra esta solución, la canalización puede ser una de las siguientes opciones:</p> <ul style="list-style-type: none"> ▪ basado en grandes conjuntos de datos ▪ Streaming o en tiempo real ▪ Un híbrido |
| | <p>Decisión de punto de control</p> <p>Reevaluación del proyecto para determinar si el valor que está previsto que aporte es suficiente para seguir adelante con él.</p> |

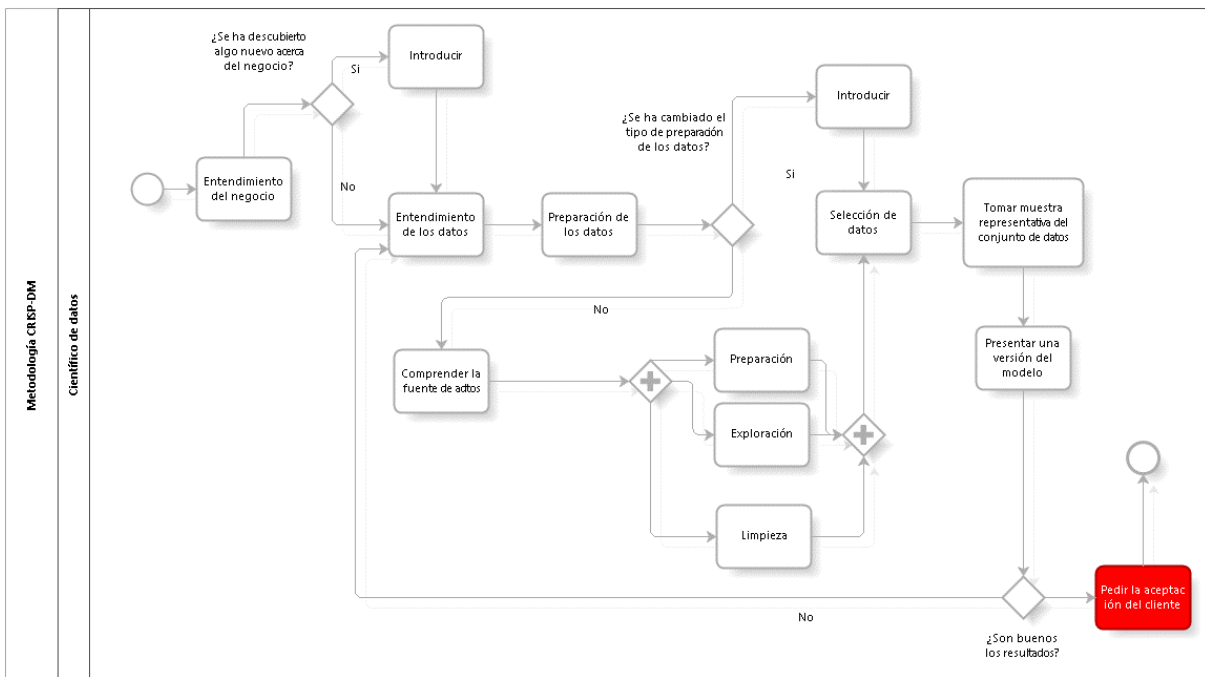


Presentar una versión del modelo

| | |
|-------|----------------------------------|
| Tarea | Presentar una versión del modelo |
|-------|----------------------------------|

| | |
|-----------------------|--|
| | <p>Determine las características de datos óptimas para el modelo de aprendizaje automático.</p> <p>Cree un modelo informativo de aprendizaje automático que prediga el objetivo con mayor precisión.</p> <p>Cree un modelo de aprendizaje automático que sea adecuado para la producción.</p> |
| <p>Salidas</p> | <p>Ingeniería de características</p> <p>Este paso requiere una combinación creativa de experiencia en el dominio y los conocimientos obtenidos del paso de exploración de datos.</p> <p>La ingeniería de características es un acto de equilibrio para encontrar e incluir variables informativas, pero al mismo tiempo tratar de evitar demasiadas variables no relacionadas. Las variables informativas mejoran su resultado; las variables no relacionadas introducen ruido innecesario en el modelo.</p> <p>También necesita generar estas características para cualquier dato nuevo obtenido durante la puntuación. Como resultado, la generación de estas características solo puede depender de los datos disponibles en el momento de la puntuación obteniendo la importancia de las características del modelo de machine learning con los ajustes preliminares.</p> |
| | <p>Desarrollo del modelo</p> <p>El proceso de formación de modelos incluye los siguientes pasos:</p> <ul style="list-style-type: none"> ▪ Dividir los datos de entrada aleatoriamente para el modelado en un conjunto de datos de aprendizaje y un conjunto de datos de prueba. ▪ Compilar los modelos mediante el conjunto de datos de aprendizaje. ▪ Evaluar el entrenamiento y el conjunto de datos de prueba. Use una serie de algoritmos de aprendizaje automático paralelos junto con los diversos parámetros de ajuste asociados (lo que se denomina barrido de parámetros) que están orientados a responder a la pregunta de interés con los datos actuales. ▪ Determinar la "mejor" solución para responder a la pregunta mediante la comparación de las métricas de éxito entre los métodos alternativos. |
| | <p>Desarrollar canalización de datos</p> <p>Evalúe si el modelo funciona lo suficiente para la producción. Algunas preguntas clave para hacer son:</p> <ul style="list-style-type: none"> ▪ ¿Responde el modelo a la pregunta con suficiente confianza dados los datos de prueba? ▪ ¿Deberías probar algún enfoque alternativo? ▪ ¿Debe recopilar datos adicionales, hacer más ingeniería de funciones o experimentar con otros algoritmos? |
| | <p>Diseñar la solución de la arquitectura</p> |

| | |
|--|--|
| | <p>Consiste en prestar atención a cómo poner en marcha la transición sistemática de una solución, así como alinearla al proceso comercial teniendo en cuenta:</p> <ul style="list-style-type: none"> ▪ Cómo se admite el sistema existente y la implementación de la tecnología durante la migración. ▪ Ilustrar cuándo y en qué momento se requieren recursos clave. ▪ Describir tanto las soluciones de prueba de migración como las de prueba de implementación. ▪ Cómo se realizará la migración de datos. ▪ Quién tiene la responsabilidad de la entrega del hardware. |
| | <p>Evaluación de la equidad</p> <p>Evalúe la imparcialidad de su modelo con métricas de matriz de confusión por ejemplo, para la evaluar la realidad de lo que predice o clasifica el modelo de machine learning ajustado previamente; este proceso le ayudará a obtener más información sobre la equidad en el modelo para producción.</p> |

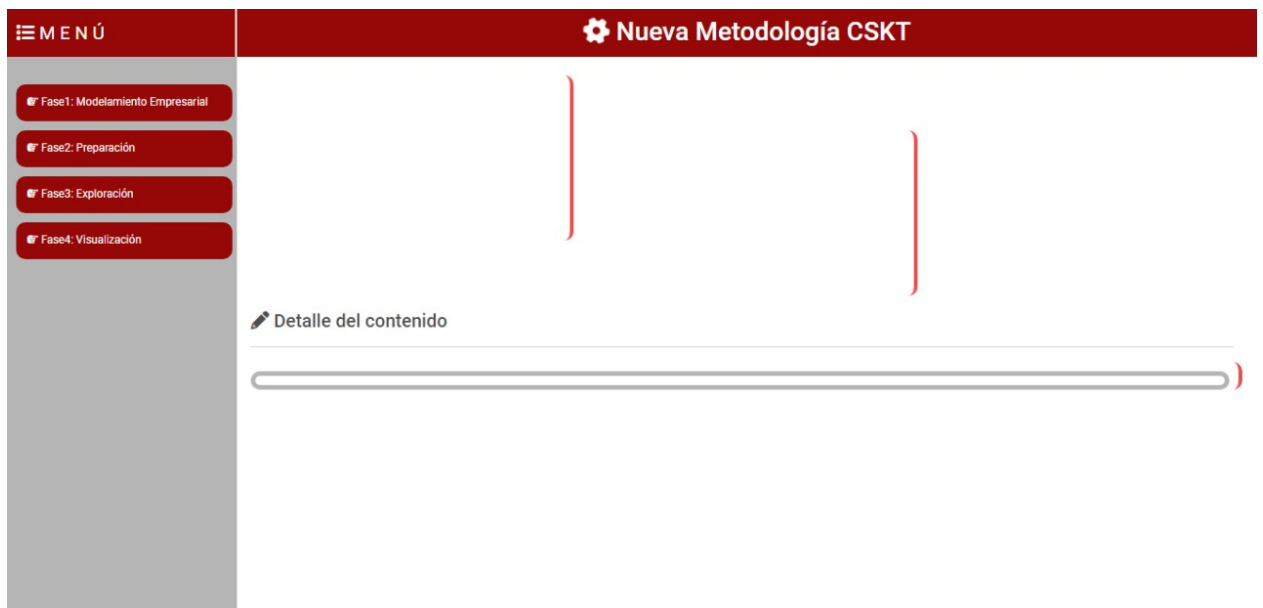


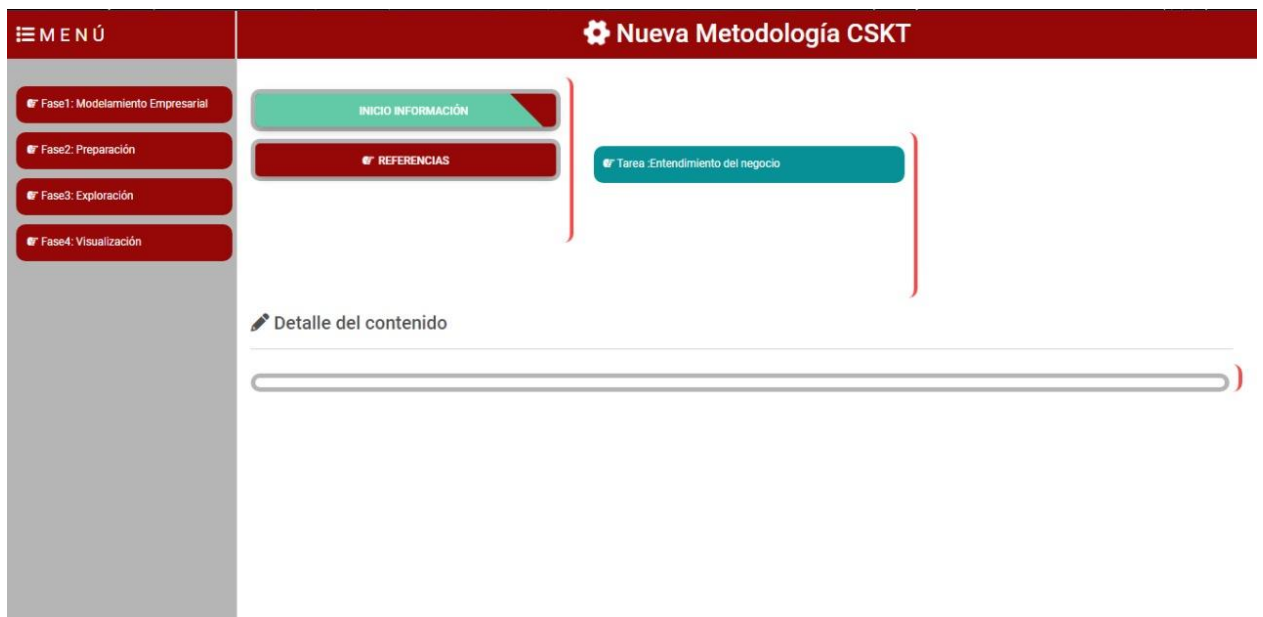
Pedir la aceptación del cliente

| | |
|----------------|--|
| <p>Tarea</p> | <p>Pedir la aceptación del cliente Confirmación que el proyecto desarrollado cumple con los objetivos del cliente.</p> |
| <p>Salidas</p> | <p>Validación del sistema</p> |

| | |
|--|---|
| | <p>El cliente debe validar que el sistema satisface sus necesidades empresariales y responde a las preguntas con una precisión aceptable para implementarlo en el entorno de producción para ser usado. La revisión incluye la documentación.</p> |
| | <p>Entrega del proyecto</p> <p>Se entrega a la entidad responsable de las operaciones.</p> |
| | <p>Informe de resultados del proyecto</p> <p>Informa de salida del proyecto para el cliente. Contiene los detalles técnicos (hardware y software) del proyecto para el funcionamiento del sistema.</p> |

5..3. Despliegue





MENÚ Nueva Metodología CSKT

- Fase1: Modelamiento Empresarial
- Fase2: Preparación
- Fase3: Exploración
- Fase4: Visualización

INICIO INFORMACIÓN

REFERENCIAS

Tarea :Entendimiento del negocio

Detalle del contenido

Entendimiento del negocio

Determinar los objetivos empresariales: El primer objetivo del analista de datos es comprender a fondo, desde una perspectiva empresarial, lo que el cliente realmente quiere conseguir. A menudo, el cliente tiene muchos objetivos y limitaciones que compiten entre sí y que deben equilibrarse adecuadamente. El objetivo del analista es descubrir, desde el principio, los factores importantes que pueden influir en el resultado del proyecto. Una posible consecuencia de descuidar este paso es gastar una gran cantidad de esfuerzo produciendo las respuestas correctas a las preguntas equivocadas.



MENÚ Nueva Metodología CSKT


- Fase1: Modelamiento Empresarial
- Fase2: Preparación
- Fase3: Exploración
- Fase4: Visualización

INICIO INFORMACIÓN

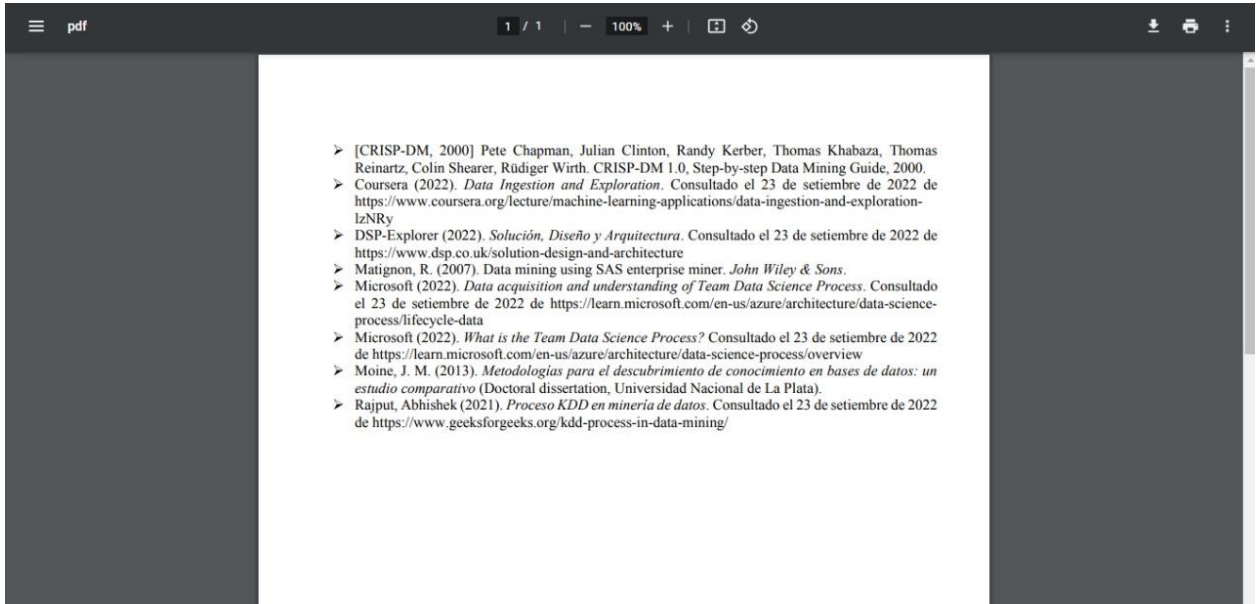
REFERENCIAS

Referencia

Detalle del contenido

Referencia Fase 1: 

Nueva metodología CSKT para mejorar los proyectos de implementación de Machine Learning en Ingeniería Industrial



☰ MENÚ

⚙ Nueva Metodología CSKT

- ☑ Fase1: Modelamiento Empresarial
- ☑ Fase2: Preparación
- ☑ Fase3: Exploración
- ☑ Fase4: Visualización

☑ REPORTE DE PREPARACION

☑ PLAN DE ACCION

☑ REFERENCIAS

☑ Tarea1 :Entendimiento de los Datos

☑ Tarea2 :Preparacion de los Datos

✎ Detalle del contenido

☰ MENÚ

⚙ Nueva Metodología CSKT

- ☑ Fase1: Modelamiento Empresarial
- ☑ Fase2: Preparación
- ☑ Fase3: Exploración
- ☑ Fase4: Visualización

☑ REPORTE DE PREPARACION

☑ PLAN DE ACCION

☑ REFERENCIAS

☑ Tarea1 :Entendimiento de los Datos

☑ Tarea2 :Preparacion de los Datos

✎ Detalle del contenido

Recoger los datos iniciales: Adquirir dentro del proyecto los datos (o el acceso a los datos) que figuran en los recursos del proyecto. Esta recopilación inicial incluye la carga de datos si es necesario para la comprensión de los datos. Por ejemplo, si se aplica una herramienta específica para la comprensión de los datos, tiene mucho sentido cargar los datos en esta herramienta. Este esfuerzo posiblemente conduzca a los pasos iniciales de preparación de los datos. Nota: si adquiere múltiples fuentes de datos, la integración es una cuestión adicional de datos, la integración es una cuestión adicional, ya sea aquí o en la fase posterior de preparación de los datos.

 pdf

MENÚ Nueva Metodología CSKT

- Fase1: Modelamiento Empresarial
- Fase2: Preparación
- Fase3: Exploración
- Fase4: Visualización

REPORTE DE PREPARACION

PLAN DE ACCION

REFERENCIAS

Tarea: Tomar Muestra del Conjunto de Datos

Detalle del contenido

Tomar muestra del conjunto de datos: Estos identifican, fusionan, dividen y muestrean conjuntos de datos de entrada, entre otras tareas



MENÚ Nueva Metodología CSKT

- Fase1: Modelamiento Empresarial
- Fase2: Preparación
- Fase3: Exploración
- Fase4: Visualización


REPORTE DE PREPARACION

PLAN DE ACCION

REFERENCIAS

Referencia

Detalle del contenido

Referencia Fase 2: 

MENÚ Nueva Metodología CSKT

- Fase1: Modelamiento Empresarial
- Fase2: Preparación
- Fase3: Exploración
- Fase4: Visualización

- PROCESO EDA
- PROGRAMACIÓN
- REFERENCIAS

- Tarea1: Preparacion
- Tarea2: Exploracion
- Tarea3: Limpieza

Detalle del contenido

Preparación: Configuración del proceso para mover los datos desde las ubicaciones de origen a las ubicaciones de destino donde se ejecutan las operaciones de análisis.



MENÚ Nueva Metodología CSKT

- Fase1: Modelamiento Empresarial
- Fase2: Preparación
- Fase3: Exploración
- Fase4: Visualización

- PROCESO EDA
- PROGRAMACIÓN
- REFERENCIAS

- Tarea1: Preparacion
- Tarea2: Exploracion
- Tarea3: Limpieza

Detalle del contenido

Preparación: Configuración del proceso para mover los datos desde las ubicaciones de origen a las ubicaciones de destino donde se ejecutan las operaciones de análisis.



MENÚ Nueva Metodología CSKT

- Fase1: Modelamiento Empresarial
- Fase2: Preparación
- Fase3: Exploración
- Fase4: Visualización

PROCESO EDA
PROGRAMACIÓN
REFERENCIAS

Referencia

Detalle del contenido

Referencia Fase 3: 

MENÚ Nueva Metodología CSKT

- Fase1: Modelamiento Empresarial
- Fase2: Preparación
- Fase3: Exploración
- Fase4: Visualización

ALMACENAR MODELO
REFERENCIAS

Tarea1 :Presentar Version de Modelo
Tarea2 :Pedir Aceptacion del Cliente

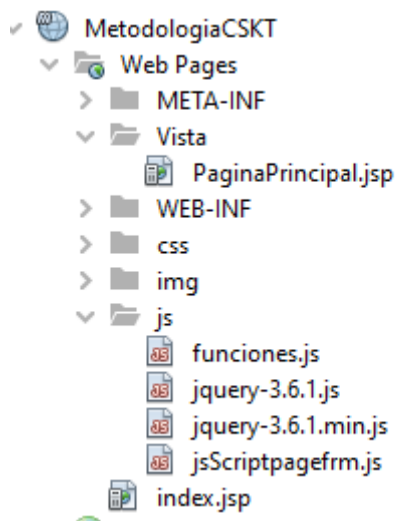
Detalle del contenido

Presentar una versión del modelo. Determine las características de datos óptimas para el modelo de aprendizaje automático. Cree un modelo informativo de aprendizaje automático que prediga el objetivo con mayor precisión. Cree un modelo de aprendizaje automático que sea adecuado para la producción.





5.4. Mantenimiento



PaginaPrincipal.jsp

```
<%@page session="true" %>
<%@page import="java.util.ArrayList"%>
<%@page import="Bean.Pdf"%>
<%!ArrayList<Pdf> lista = null;%>
<% lista = (ArrayList) request.getAttribute("lista");%>

<%@page contentType="text/html" pageEncoding="UTF-8"%>
<!DOCTYPE html>
<html>
```



```

<head>
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">

  <title>Menú De Navegacion Con Listas</title>
  <link rel="preconnect" href="https://fonts.googleapis.com">
  <link rel="preconnect" href="https://fonts.gstatic.com" crossorigin>
  <link href="https://fonts.googleapis.com/css2?family=Roboto:wght@400;500&display=swap" rel="stylesheet">
  <link href="https://cdn.jsdelivr.net/npm/bootstrap@5.0.2/dist/css/bootstrap.min.css" rel="stylesheet" integrity="sha384-
EVSTQN3/azprG1Anm3QDgpJLIm9Nao0Yz1ztcQTWfspd3yD65VohhpuuCOMLASjC" crossorigin="anonymous">
  <script src="<%=request.getContextPath()%>/js/funciones.js" type="text/javascript"></script>
  <link href="<%=request.getContextPath()%>/css/Estilo.css" rel="stylesheet" type="text/css"/>
  <link href="<%=request.getContextPath()%>/css/cssBoton.css" rel="stylesheet" type="text/css"/>
  <link href="<%=request.getContextPath()%>/css/stylepagefrm.css" rel="stylesheet" type="text/css" />
  <link href="https://cdn.jsdelivr.net/npm/bootstrap@5.0.2/dist/css/bootstrap.min.css" rel="stylesheet"
    integrity="sha384-EVSTQN3/azprG1Anm3QDgpJLIm9Nao0Yz1ztcQTWfspd3yD65VohhpuuCOMLASjC" crossorigin="anonymous">
  <!-- Font Awesome -->
  <link href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/6.0.0/css/all.min.css" rel="stylesheet" />
  <!-- Google Fonts -->
  <link href="https://fonts.googleapis.com/css?family=Roboto:300,400,500,700&display=swap" rel="stylesheet" />
  <!-- MDB -->
  <link href="https://cdnjs.cloudflare.com/ajax/libs/mdb-ui-kit/5.0.0/mdb.min.css" rel="stylesheet" />
  <script type="text/javascript" src="https://cdnjs.cloudflare.com/ajax/libs/mdb-ui-kit/5.0.0/mdb.min.js"></script>
  <script src="https://cdn.jsdelivr.net/npm/bootstrap@5.0.2/dist/js/bootstrap.bundle.min.js"
    integrity="sha384-MrcW6ZMFYlzcL88NI+NtUVF0sA7MsXsP1UyJoMp4YLEuNSfAP+JcXn/tWtIaxVXM"
    crossorigin="anonymous"></script>
  <script src="<%=request.getContextPath()%>/js/jsScriptpagefrm.js" type="text/javascript"></script>
  <script src="https://cdn.jsdelivr.net/npm/bootstrap@5.0.2/dist/js/bootstrap.bundle.min.js" integrity="sha384-
MrcW6ZMFYlzcL88NI+NtUVF0sA7MsXsP1UyJoMp4YLEuNSfAP+JcXn/tWtIaxVXM" crossorigin="anonymous"></script>
  <script src="https://code.jquery.com/jquery-3.6.1.min.js"></script>
  <script src="<%=request.getContextPath()%>/js/jquery-3.6.1.min.js" type="text/javascript"></script>
  <script src="<%=request.getContextPath()%>/js/jquery-3.6.1.js" type="text/javascript"></script>
  </script>

  $(document).ready(function () {
    $("#Fase1").click(function () {
      mostrarmenu1();
      limpiar();
      limpiar2();
    });
  });

  function mostrarmenu1() {
    $.ajax({
      url: 'ControladorFases',
      type: 'POST',
      data: 'op=1',
      success: function (result) {
        $('#submenu').html(result);
      }
    });
  }

  $(document).ready(function () {
    $("#Fase2").click(function () {
      mostrarmenu2();
      limpiar();
      limpiar2();
    });
  });

  function mostrarmenu2() {
    $.ajax({
      url: 'ControladorFases',
      type: 'POST',
      data: 'op=2',
      success: function (result) {
        $('#submenu').html(result);
      }
    });
  }

```

```
    });  
  }  
  $(document).ready(function () {  
    $("#Fase3").click(function () {  
      mostrarmenu3();  
      limpiar();  
      limpiar2();  
    });  
  });  
  
function mostrarmenu3() {  
  $.ajax({  
    url: 'ControladorFases',  
    type: 'POST',  
    data: 'op=3',  
    success: function (result) {  
      $('#submenu').html(result);  
    }  
  });  
}  
$(document).ready(function () {  
  $("#Fase4").click(function () {  
    mostrarmenu4();  
    limpiar();  
    limpiar2();  
  });  
});  
  
function mostrarmenu4() {  
  $.ajax({  
    url: 'ControladorFases',  
    type: 'POST',  
    data: 'op=4',  
    success: function (result) {  
      $('#submenu').html(result);  
    }  
  });  
}  
  
function limpiar() {  
  $.ajax({  
    url: 'ControladorLimpiarDiv',  
    type: 'POST',  
    data: 'op2=1',  
    success: function (result) {  
      $('#mostrartareas').html(result);  
    }  
  });  
}  
  
function limpiar2() {  
  $.ajax({  
    url: 'ControladorLimpiarDiv',  
    type: 'POST',  
    data: 'op2=2',  
    success: function (result) {  
      $('#mostrarpdf').html(result);  
    }  
  });  
}  
  
</script>  
  
</head>  
<body>  
  <div id="viewport">  
    <!-- Sidebar -->
```

Nueva metodología CSKT para mejorar los proyectos de implementación de Machine Learning en Ingeniería Industrial

```

<div id="sidebar">
  <header>
    <a href="#"><i class="fas fa-list"></i> M E N Ú</a>
  </header>
  <form name="form">
    <div class="container">
      <div class="row">

        <input type="hidden" name="op">
        <input type="hidden" name="op2">
        <div class="col-12 menu">
          <a id="Fase1" href="#"><i class="fas fa-hand-point-right"></i> Fase1: Modelamiento Empresarial</a>
        </div>
        <div class="col-12 menu">
          <a id="Fase2" href="#"><i class="fas fa-hand-point-right"></i> Fase2: Preparación</a>
        </div>
        <div class="col-12 menu">
          <a id="Fase3" href="#"><i class="fas fa-hand-point-right"></i> Fase3: Exploración</a>
        </div>
        <div class="col-12 menu">
          <a id="Fase4" href="#"><i class="fas fa-hand-point-right"></i> Fase4: Visualización</a>
        </div>

      </div>
    </div>
  </form>
</div>
<!-- Content -->
<div id="content">
  <div class="container-fluid">
    <div class="row">
      <div class="col-12" id="cabecera">
        <h3><i class="fas fa-cog fa-spin"></i> Nueva Metodología CSKT</h3>
      </div>
      <div class="row col-12" id="contenido">
        <div class="col-4" id="contenidosubmenuuno">
          <div id="submenu">

            </div>
          </div>
          <div class="col-4" id="contenidosubmenudos">
            <div id="mostrartareas" class="col-12 mostrartareas">

              </div>
            </div>
            <div class="col-12" id="contenidopdfitulo">
              <div class="col-12">
                <h5><i class="fas fa-pencil-alt"></i> Detalle del contenido</h5>
                <hr>
              </div>
            </div>
            <div class="col-12" id="contenidopdf">
              <div class="col-12 divcontenidopdf">
                <div class="divnfpdf">
                  <div id="mostrarpdf">

                    </div>
                  </div>
                </div>
              </div>
            </div>
          </div>
        </div>
      </div>
    </div>
  </div>
</body>
</html>

```

funciones.js

```
function iniciar(ruta)
{
    var url=ruta+"/ControladorIndex"
    window.open(url);
    this.close();
}

function Enviar(ruta)
{
    window.document.form.action=ruta+"/ControladorPdf";
    window.document.form.method="POST";
    window.document.form.submit();
}
```

Index.jsp

```
<%@page contentType="text/html" pageEncoding="UTF-8"%>
<!DOCTYPE html>
<html>
    <head>
        <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
        <title>JSP Page</title>
        <script src="<%=request.getContextPath()%>/js/funciones.js" type="text/javascript"></script>
    </head>
    <body onload="iniciar('<%=request.getContextPath()%>')">
    </body>
</html>
```

6. Validación y diseño experimental

Se toma una muestra de estudiantes para su opinión con el objetivo de ayudar a mejorar la manera de presentar esta nueva metodología en el futuro. Esta se tomó en el 2022 de los estudiantes de Ingeniería en Perú.

La escala de los datos es:

- 1 = Totalmente en desacuerdo
- 2 = En desacuerdo
- 3 = Ni de acuerdo ni desacuerdo

4 = De acuerdo

5 = Totalmente de acuerdo

Para comprobar la validación se realizó una prueba piloto:

A. Prueba Piloto

| Ítem | Número de observaciones |
|---|-------------------------|
| Podrá utilizar mecanismos simples para la planificación de tareas cuando necesita automatizar tareas sencillas de manufactura | 9 |
| Podrá buscar a partir de información disponible cuando necesita digitalizar KPI's | 9 |
| Podrá ayudar en la automatización laboral cuando necesita automatizar tareas | 9 |
| Podrá brindar una imagen clara de sus logros esperados cuando necesita establecer un número de indicadores clave | 9 |
| Podrá identificar los elementos que intervendrán en el modelo predictivo cuando necesita desarrollar técnicas de análisis de costos | 9 |
| Podrá ejecutar instrumentos de capacitación en la empresa cuando necesita formar un equipo de auditoría | 9 |
| Podrá detallar la comprobación de sus resultados cuando necesita elaborar una lista de verificación | 9 |

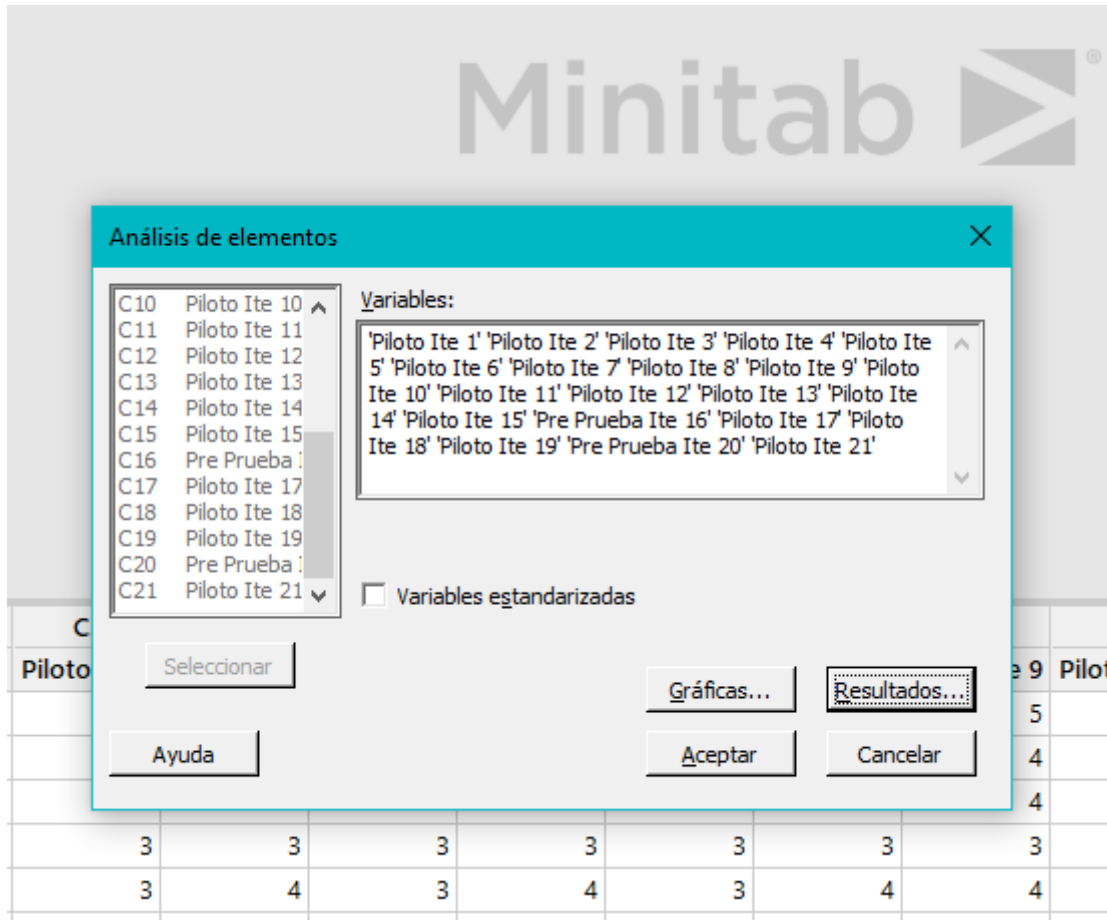
| | |
|---|---|
| Podrá realizar un análisis de las predicciones de demanda cuando necesita fijar un precio por cada predicción realizada | 9 |
| Podrá explicar los precios del cobre cuando necesita establecer un número de indicadores clave | 9 |
| Podrá diseñar un servicio que incluya parámetros de calidad cuando necesita cuantificar la satisfacción del cliente | 9 |
| Podrá generar evidencia que proporcione mayor fiabilidad a la empresa cuando se necesita comprobar de forma exhaustiva los procesos | 9 |
| Podrá poner a disposición de la empresa los balances de la situación cuando necesita realizar una auditoria | 9 |
| Podrá pronosticar la demanda de efectivo cuando necesita la aplicación del pronóstico | 9 |
| Podrá conocer el comportamiento de los clientes cuando necesita prestar servicio a los clientes | 9 |
| Podrá obtener información sobre la eficacia de su aplicación cuando necesita corroborar la información de satisfacción del cliente | 9 |
| Podrá brindar información sobre los futuros precios del cobre cuando necesita construir el modelo de predicción del precio de cobre | 9 |
| Podrá cumplir las expectativas por la falta de personalización cuando necesita medir la satisfacción del cliente | 9 |

| | |
|---|---|
| Podrá analizar la importancia de los costos de producción cuando necesita mencionar los costos de operación | 9 |
| Podrá ampliar la información sobre el precio del cobre cuando necesita digitalizar el proceso de predicción del precio de cobre | 9 |
| Podrá promover la mejora continua de los procesos de la organización cuando necesita realizar el proceso de auditoria | 9 |
| Podrá verificar si el modelo hace buenas predicciones cuando necesita validar el modelo de predicción de precio del cobre | 9 |

a) Análisis estadístico

The screenshot shows the Minitab software interface. The 'Estadísticas' menu is open, and 'Análisis de elementos...' is selected. A tooltip for 'Análisis de elementos' is visible, stating: 'Evaluar la confiabilidad de múltiples elementos de un sondeo para medir la misma característica.' The background data table is as follows:

| | C1 | C2 | | C7 | C8 | C9 |
|----|--------------|--------------|--------------|--------------|--------------|--------------|
| | Piloto Ite 1 | Piloto Ite 2 | Piloto Ite 3 | Piloto Ite 7 | Piloto Ite 8 | Piloto Ite 9 |
| 1 | 5 | 5 | 5 | 5 | 5 | 5 |
| 2 | 4 | 4 | 4 | 5 | 4 | 4 |
| 3 | 5 | 5 | 5 | 5 | 5 | 4 |
| 4 | 3 | 3 | 3 | 3 | 3 | 3 |
| 5 | 3 | 3 | 3 | 4 | 3 | 4 |
| 6 | 4 | 4 | 4 | 4 | 4 | 5 |
| 7 | 5 | 4 | 5 | 4 | 4 | 4 |
| 8 | 5 | 4 | 5 | 5 | 4 | 4 |
| 9 | 5 | 5 | 5 | 4 | 5 | 4 |
| 10 | | | | | | |



b) Interpretación estadística

HOJA DE TRABAJO 1

Análisis de elementos de Piloto Ite 1; Piloto Ite 2; Piloto Ite 3; I

| Pre Prueba Ite 20 | 9 | 4.222 | 0.855 |
|-------------------|---|--------|--------|
| Piloto Ite 21 | 9 | 4.222 | 0.667 |
| Total | 9 | 89.222 | 14.016 |

Alfa de Cronbach

Alfa
0.9683

Estadísticas de elementos omitidas

| Variable omitida | Media total ajustada | Desv.Est. total ajustada | Correlación total ajustada por elemento | Correlación múltiple cuadrada | Alfa de Cronbach |
|------------------|----------------------|--------------------------|---|-------------------------------|------------------|
| Piloto Ite 1 | 84.89 | 13.29 | 0.8291 | * | 0.9660 |
| Piloto Ite 2 | 85.11 | 13.36 | 0.8242 | * | 0.9662 |

En la teoría según Sampieri (2014) se muestra la tabla de rangos para los resultados de confiabilidad

| Rangos | Magnitud |
|-------------|----------|
| 0,81 a 1,00 | Muy Alta |
| 0,61 a 0,80 | Alta |
| 0,41 a 0,60 | Moderada |
| 0,21 a 0,40 | Baja |
| 0,01 a 0,20 | Muy Baja |

Con base en Sampieri (2014), al obtener una alfa de Cronbach de 0.9683, la encuesta es confiable y se puede avanzar con esta etapa tomando las 36 encuestas que se necesitan.

B. Diseño de resultados

a) Dimensionalidad

1.- Preparación

| N | Afirmaciones | Respuestas | | | | |
|---|---|------------|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| 1 | Podrá utilizar mecanismos simples para la planificación de tareas cuando necesita automatizar tareas sencillas de manufactura | | | | | |
| 2 | Podrá buscar a partir de información disponible cuando necesita digitalizar KPI's | | | | | |
| 3 | Podrá ayudar en la automatización laboral cuando necesita automatizar tareas | | | | | |

| | | | | | | |
|---|---|--|--|--|--|--|
| 4 | Podrá brindar una imagen clara de sus logros esperados cuando necesita establecer un número de indicadores clave | | | | | |
| 5 | Podrá identificar los elementos que intervendrán en el modelo predictivo cuando necesita desarrollar técnicas de análisis de costos | | | | | |
| 6 | Podrá ejecutar instrumentos de capacitación en la empresa cuando necesita formar un equipo de auditoria | | | | | |

2.- Exploración

| N | Afirmaciones | Respuestas | | | | |
|---|---|------------|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| 7 | Podrá detallar la comprobación de sus resultados cuando necesita elaborar una lista de verificación | | | | | |
| 8 | Podrá realizar un análisis de las predicciones de demanda cuando necesita fijar un precio por cada predicción realizada | | | | | |

3.- Visualización

| N | Afirmaciones | Respuestas | | | | |
|---|--|------------|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| 9 | Podrá explicar los precios del cobre cuando necesita establecer un número de indicadores clave | | | | | |

| | | | | | | |
|----|---|--|--|--|--|--|
| 10 | Podrá diseñar un servicio que incluya parámetros de calidad cuando necesita cuantificar la satisfacción del cliente | | | | | |
| 11 | Podrá generar evidencia que proporcione mayor fiabilidad a la empresa cuando se necesita comprobar de forma exhaustiva los procesos | | | | | |
| 12 | Podrá poner a disposición de la empresa los balances de la situación cuando necesita realizar una auditoria | | | | | |
| 13 | Podrá pronosticar la demanda de efectivo cuando necesita la aplicación del pronóstico | | | | | |
| 14 | Podrá conocer el comportamiento de los clientes cuando necesita prestar servicio a los clientes | | | | | |
| 15 | Podrá obtener información sobre la eficacia de su aplicación cuando necesita corroborar la información de satisfacción del cliente | | | | | |
| 16 | Podrá brindar información sobre los futuros precios del cobre cuando necesita construir el modelo de predicción del precio de cobre | | | | | |
| 17 | Podrá cumplir las expectativas por la falta de personalización cuando necesita medir la satisfacción del cliente | | | | | |
| 18 | Podrá analizar la importancia de los costos de producción cuando necesita mencionar los costos de operación | | | | | |

| | | | | | | |
|----|---|--|--|--|--|--|
| 19 | Podrá ampliar la información sobre el precio del cobre cuando necesita digitalizar el proceso de predicción del precio de cobre | | | | | |
| 20 | Podrá promover la mejora continua de los procesos de la organización cuando necesita realizar el proceso de auditoria | | | | | |
| 21 | Podrá verificar si el modelo hace buenas predicciones cuando necesita validar el modelo de predicción de precio del cobre | | | | | |

C. Confiabilidad: Preparación

a) Análisis estadístico

The screenshot shows the Minitab software interface with the 'Estadísticas' menu open. The 'Análisis de elementos...' option is highlighted, and a tooltip is visible over it. The tooltip text reads: 'Análisis de elementos. Evaluar la confiabilidad de múltiples elementos de un sondeo para medir la misma característica.'

The background shows a data table with the following structure:

| | C21 | C22 | | | | | C27 | C28 | |
|---|---------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|------|
| | Piloto Ite 21 | Dim 1 Ite 1 | Dim 1 Ite 2 | Dim 1 Ite 3 | Dim 1 Ite 4 | Dim 1 Ite 5 | Dim 1 Ite 6 | Prueba Ite 7 | Prue |
| 1 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | | |
| 2 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | | |
| 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | | |
| 4 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | | |
| 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | | |
| 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | | |
| 7 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | | |
| 8 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | | |
| 9 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | | |



b) Interpretación estadística

HOJA DE TRABAJO 1

Análisis de elementos de Dim 1 Ite 1; Dim 1 Ite 2; Dim 1 Ite 3;

| | | | |
|-------------|----|--------|-------|
| Dim 1 Ite 5 | 36 | 4.444 | 1.229 |
| Dim 1 Ite 6 | 36 | 4.417 | 1.360 |
| Total | 36 | 27.556 | 5.940 |

Alfa de Cronbach

Alfa
0.9448

Estadísticas de elementos omitidas

| Variable omitida | Media total ajustada | Desv.Est. total ajustada | Correlación total ajustada por elemento | Correlación múltiple cuadrada | Alfa de Cronbach |
|------------------|----------------------|--------------------------|---|-------------------------------|------------------|
| Dim 1 Ite 1 | 22.889 | 5.007 | 0.9039 | 0.9716 | 0.9272 |
| Dim 1 Ite 2 | 22.889 | 5.007 | 0.9039 | 0.9727 | 0.9272 |

En la teoría según Sampieri (2014) se muestra la tabla de rangos para los resultados de confiabilidad

| Rangos | Magnitud |
|-------------|----------|
| 0,81 a 1,00 | Muy Alta |
| 0,61 a 0,80 | Alta |
| 0,41 a 0,60 | Moderada |
| 0,21 a 0,40 | Baja |
| 0,01 a 0,20 | Muy Baja |

Con base en Sampieri (2014), al obtener una alfa de Cronbach de 0.9448, la encuesta es confiable y en esta etapa mantenemos las respuestas de 36 encuestas que se necesitaron.

D. Confiabilidad: Exploración

a) Análisis estadístico

The screenshot shows the Minitab software interface. The 'Estadísticas' menu is open, and 'Análisis de elementos...' is selected. A tooltip for 'Análisis de elementos' is displayed, stating: 'Evaluar la confiabilidad de múltiples elementos de un sondeo para medir la misma característica.'

| | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 |
|---|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | Piloto Ite 21 | Dim 1 Ite 1 | Dim 1 Ite 2 | Dim 1 Ite 3 | Dim 1 Ite 4 | Dim 1 Ite 5 | Dim 1 Ite 6 | Dim 2 Ite 7 | Dim 2 Ite 8 |
| 1 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 2 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 4 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

| C23 | C24 | C25 | C26 | C27 | C28 | C29 |
|-----------|-------------|-------------|-------------|-------------|-------------|-------------|
| n 1 Ite 2 | Dim 1 Ite 3 | Dim 1 Ite 4 | Dim 1 Ite 5 | Dim 1 Ite 6 | Dim 2 Ite 7 | Dim 2 Ite 8 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 |

b) Interpretación estadística

Análisis de elementos de Piloto I...

Análisis de elementos de Dim 1 I...

Análisis de elementos de Dim 2 I...

HOJA DE TRABAJO 1

Análisis de elementos de Dim 2 Ite 7; Dim 2 Ite 8

Matriz de correlación

Correlación de Pearson de Dim 2 Ite 7 y Dim 2 Ite 8 = 0.841

Estadísticas totales y de elementos

| Variable | Conteo total | Media | Desv.Est. |
|-------------|--------------|--------|-----------|
| Dim 2 Ite 7 | 36 | 4.6667 | 1.0142 |
| Dim 2 Ite 8 | 36 | 4.6111 | 1.0496 |
| Total | 36 | 9.2778 | 1.9801 |

Alfa de Cronbach

Alfa
0.9134

En la teoría según Sampieri (2014) se muestra la tabla de rangos para los resultados de confiabilidad

| Rangos | Magnitud |
|-------------|----------|
| 0,81 a 1,00 | Muy Alta |
| 0,61 a 0,80 | Alta |
| 0,41 a 0,60 | Moderada |
| 0,21 a 0,40 | Baja |
| 0,01 a 0,20 | Muy Baja |

Con base en Sampieri (2014), al obtener una alfa de Cronbach de 0.9134, la encuesta es confiable y en esta etapa mantenemos las respuestas de 36 encuestas que se necesitaron.

E. Confiabilidad: Visualización

a) Análisis estadístico

The screenshot shows the Minitab software interface. The 'Estadísticas' menu is open, and 'Análisis de elementos...' is selected. A tooltip for 'Análisis de elementos' is displayed, stating: 'Evaluar la confiabilidad de múltiples elementos de un sondeo para medir la misma característica.'

| | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 |
|---|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | Dim 3 Ite 13 | Dim 3 Ite 14 | Dim 3 Ite 15 | Dim 3 Ite 16 | Dim 3 Ite 17 | Dim 3 Ite 18 | Dim 3 Ite 19 | Dim 3 Ite 20 | Dim 3 Ite 21 |
| 1 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 2 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

| C36 | C37 | C38 | C39 | C40 | C41 | C42 |
|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Dim 3 Ite 15 | Dim 3 Ite 16 | Dim 3 Ite 17 | Dim 3 Ite 18 | Dim 3 Ite 19 | Dim 3 Ite 20 | Dim 3 Ite 21 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 |

b) Interpretación estadística

| Dim 3 Ite 21 | 50 | 4.833 | 0.731 |
|--------------|----|--------|-------|
| Total | 36 | 61.361 | 9.631 |

Alfa de Cronbach

Alfa
0.9612

Estadísticas de elementos omitidas

| Variable omitida | Media total ajustada | Desv.Est. total ajustada | Correlación total ajustada por elemento | Correlación múltiple cuadrada | Alfa de Cronbach |
|------------------|----------------------|--------------------------|---|-------------------------------|------------------|
| Dim 3 Ite 9 | 56.472 | 9.063 | 0.8412 | 1.0000 | 0.9579 |
| Dim 3 Ite 10 | 56.611 | 8.875 | 0.8525 | 1.0000 | 0.9567 |
| Dim 3 Ite 11 | 56.750 | 8.729 | 0.8436 | 1.0000 | 0.9570 |
| Dim 3 Ite 12 | 56.733 | 8.700 | 0.8531 | 1.0000 | 0.9566 |

En la teoría según Sampieri (2014) se muestra la tabla de rangos para los resultados de confiabilidad

| Rangos | Magnitud |
|-------------|----------|
| 0,81 a 1,00 | Muy Alta |
| 0,61 a 0,80 | Alta |
| 0,41 a 0,60 | Moderada |
| 0,21 a 0,40 | Baja |
| 0,01 a 0,20 | Muy Baja |

Con base en Sampieri (2014), al obtener una alfa de Cronbach de 0.9612, la encuesta es confiable y en esta etapa mantenemos las respuestas de 36 encuestas que se necesitaron.

b) Cuadro de confiabilidad

| Dimensión 01: | N° ítem | Confiabilidad |
|---|---------|-----------------|
| Preparación | | |
| Podrá utilizar mecanismos simples para la planificación de tareas cuando necesita automatizar tareas sencillas de manufactura | 1 | 94.48% muy alta |
| Podrá buscar a partir de información disponible cuando necesita digitalizar KPI's | 2 | |
| Podrá ayudar en la automatización laboral cuando necesita automatizar tareas | 3 | |
| Podrá brindar una imagen clara de sus logros esperados cuando necesita establecer un número de indicadores clave | 4 | |

| | | |
|---|----------------|----------------------|
| Podrá identificar los elementos que intervendrán en el modelo predictivo cuando necesita desarrollar técnicas de análisis de costos | 5 | |
| Podrá ejecutar instrumentos de capacitación en la empresa cuando necesita formar un equipo de auditoria | 6 | |
| Dimensión 02: | N° ítem | Confiabilidad |
| Exploración | | |
| Podrá detallar la comprobación de sus resultados cuando necesita elaborar una lista de verificación | 7 | 91.34% muy alta |
| Podrá realizar un análisis de las predicciones de demanda cuando necesita fijar un precio por cada predicción realizada | 8 | |
| Dimensión 03: | N° ítem | Confiabilidad |
| Visualización | | |
| Podrá explicar los precios del cobre cuando necesita establecer un número de indicadores clave | 9 | 96.12% muy alta |
| Podrá diseñar un servicio que incluya parámetros de calidad cuando necesita cuantificar la satisfacción del cliente | 10 | |
| Podrá generar evidencia que proporcione mayor fiabilidad a la empresa cuando se necesita comprobar de forma exhaustiva los procesos | 11 | |
| Podrá poner a disposición de la empresa los balances de la situación cuando necesita realizar una auditoria | 12 | |

| | | |
|---|-----------|--|
| Podrá pronosticar la demanda de efectivo cuando necesita la aplicación del pronóstico | 13 | |
| Podrá conocer el comportamiento de los clientes cuando necesita prestar servicio a los clientes | 14 | |
| Podrá obtener información sobre la eficacia de su aplicación cuando necesita corroborar la información de satisfacción del cliente | 15 | |
| Podrá brindar información sobre los futuros precios del cobre cuando necesita construir el modelo de predicción del precio de cobre | 16 | |
| Podrá cumplir las expectativas por la falta de personalización cuando necesita medir la satisfacción del cliente | 17 | |
| Podrá analizar la importancia de los costos de producción cuando necesita mencionar los costos de operación | 18 | |
| Podrá ampliar la información sobre el precio del cobre cuando necesita digitalizar el proceso de predicción del precio de cobre | 19 | |
| Podrá promover la mejora continua de los procesos de la organización cuando necesita realizar el proceso de auditoria | 20 | |
| Podrá verificar si el modelo hace buenas predicciones cuando necesita validar el modelo de predicción de precio del cobre | 21 | |

7. Resultados y discusión

7.1. Metodología propuesta

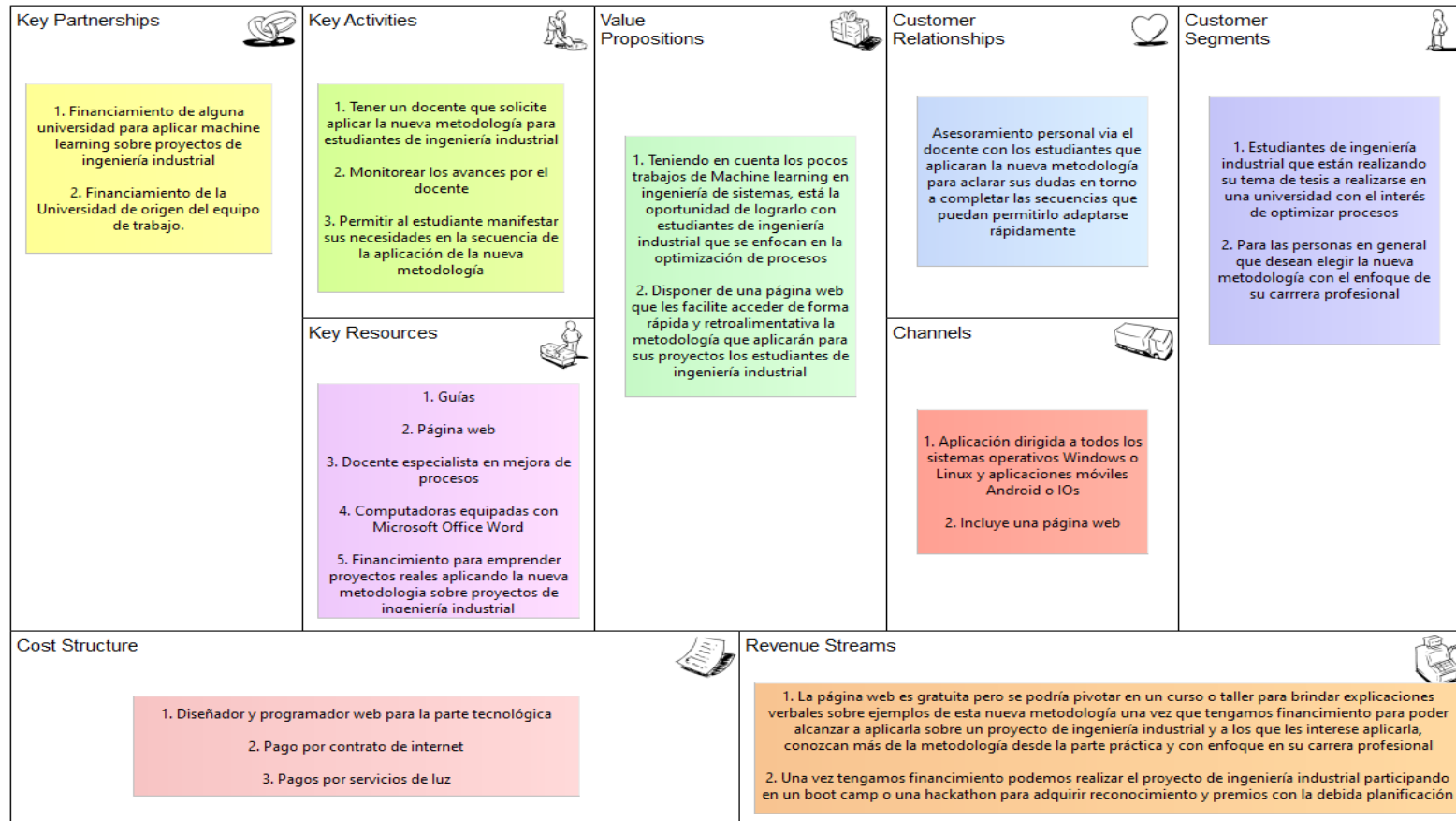


Fig. 11. Modelo CANVAS de la metodología propuesta

7..2. Desarrollo de las encuestas

Con el objetivo de identificar el grado de satisfacción que permitirá mejorar la nueva metodología en el futuro y la confiabilidad muy alta por Alfa de Cronbach, se ha procedido a mostrar los resultados a continuación.

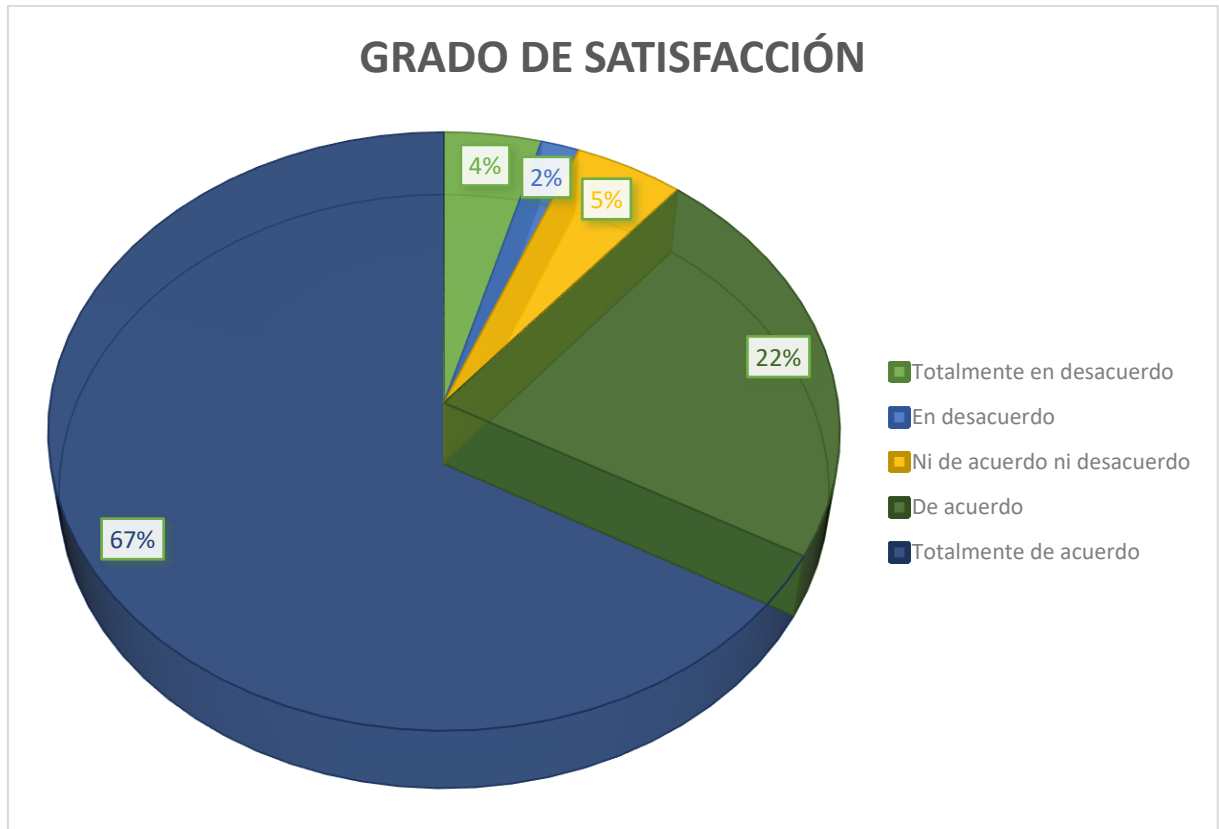


Fig. 11. Grado de satisfacción

Como se visualiza, el 67% de los encuestados califica como totalmente de acuerdo con la nueva metodología CSKT (CRISP-DM, SEMMA, KDD y TDSP), en tanto que el 22% opina estar de acuerdo, mientras que el 5% expresa estar ni de acuerdo ni desacuerdo, en tanto que el 2% manifiesta estar en desacuerdo y el 4% indica estar totalmente en desacuerdo.

8. Conclusiones y trabajo futuro

Tras realizar todo el trabajo de investigación hay un punto de referencia de este proyecto que son las metodologías CRISP-DM, SEMMA, KDD y TDSP, ellas pueden ser implementadas por separado, o donde un proyecto solo puede elegir una de estas metodologías para mejorar un proceso aplicando machine learning. Todas estas metodologías son buenas, sin embargo, existen otros campos como la ingeniería industrial y relacionado al objetivo propuesto que es diseñar y desarrollar una nueva metodología llamada CSKT que se ajuste a la implementación de proyectos de Machine Learning desarrollados en el campo de la Ingeniería Industrial, con los resultados podemos evidenciar hasta qué punto la metodología propuesta le facilita o facilitará la implementación de machine learning para un proceso que requiere ser automatizado por estudiantes que han señalado su grado de satisfacción respecto a esta presentación de la metodología CSKT que requiere ser mejorada en la manera en que actúa una metodología que se ha aplicado llamada Design Thinking que nos permitió obtener ideas interesantes sobre esta metodología en pro de mejorada a futuro como por ejemplo:

- Utilizar la estrategia en numerosas ocasiones para ayudar a los estudiantes de Ingeniería industrial a generar nuevas ideas para sus proyectos de tesis.
- Cuando hablamos de metodología podemos aplicarlo al área de desarrollo y enseñanza para los proyectos considerados adecuados.
- Todas las diferencias con respecto a otras metodologías de Machine Learning deben usar para beneficiar a todos los estudiantes de Ingeniería Industrial para los proyectos que consideran adecuados.

Y otras ideas nuevas que sobre la marcha y con la consecución de las actividades clave del proyecto contempladas en el diagrama canvas aunado a cambios en el desarrollo de los sprints conforme a la metodología Scrum para sumar a este emprendimiento las siguientes ideas nuevas:

- Aplicar la enseñanza aprendizaje por investigación ligada a la metodología.
- Las metodologías educativas suelen girar en torno a las teorías de aprendizaje.
- Cuando el estudiante pone en juego sus conocimientos es cuando verdaderamente descubre el significado.

Como recomendación tras haber realizado toda la investigación es emprender una relación causal entre la educación y la nueva metodología, puesto que en el país de Perú falta más formación en informática que puede resultar difícil cuando se aplica esta nueva metodología con las consideraciones pedagógicas necesarias si no se hacen conocer desde la explicación metodológica pasando por la verbal hasta la práctica de:

- SQL
- R
- Python
- Tableau
- Power BI
- Excel avanzado

Debido a que la verdadera garantía de que se aplique la metodología propuesta en el flujo o secuencia especificada en la Tabla 9 llamada, Producto Mínimo Viable (MPV) es que constantemente recibamos críticas constructivas sobre el proceso provenientes de los estudiantes con enfoque en su formación profesional, tomando en cuenta que necesitan probar algo que funcione, y que ese algo tenga éxito en la optimización de procesos, aunado a que el esfuerzo que requiera para aplicarlo no sea muy alto o costo. Por todas estas consideraciones, es importante que futuros trabajos analicen las siguientes preguntas en aras de emprenderse a crear sus propias metodologías, de pronto, para otras especialidades y no necesariamente la ingeniería industrial que se enfoca en optimización de procesos y que además se trató en este documento, sino otras carreras que compartan un enfoque similar, respondiendo a las interrogantes a continuación:

- ¿Encontrar metodologías nuevas más significativas disminuirá la deserción o fracaso en el proceso de elaboración de tesis en Ingeniería Industrial?
- ¿Cuándo algún estudiante me pregunte qué metodología hacer para el proyecto de investigación, la primera pregunta que le hago es: usarás Machine Learning?

- ¿La nueva metodología se presenta como un modelo educativo para Ingeniería Industrial porque la propuesta tiene en cuenta que la carrera se dedica a la optimización de procesos y las tesis no presentan de manera explícita un proyecto con Machine Learning?
- ¿Los estudiantes que cursan metodologías de investigación poseen los requisitos necesarios para abordar la propuesta?

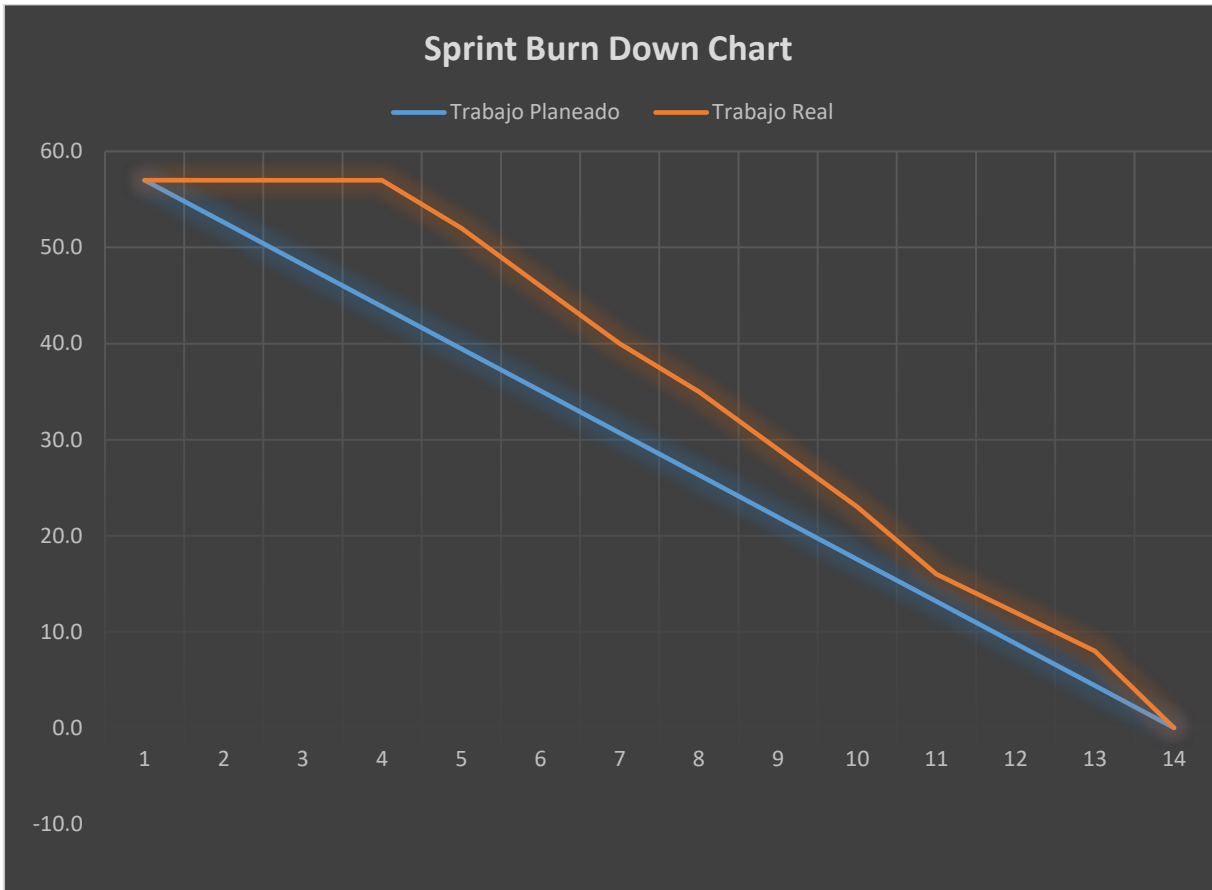
Referencias bibliográficas

1. Al-Anqoudi, Y., Al-Hamdani, A., Al-Badawi, M., & Hedjam, R. (2021). Using machine learning in business process re-engineering. *Big Data and Cognitive Computing*, 5(4), 61.
2. Cabello, J. (Recuperado el 22 de agosto de 2022). *¿Cuánto cuesta realmente una página web? Descubre su precio justo*. José Cabello. <https://josecabello.net/disenio/cuanto-cuesta-realmente-una-pagina-web/>
3. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories
4. Darwish, S. (2021). *Mapping the job potential of the industrial engineer: a web-investigation* (Doctoral dissertation, North-West University (South Africa).).
5. Dias, R., Senna, P., Gonçalves, A., Reis, J., Michalaros, N., Alexopoulos, K., & Gomes, M. (2021). PREFAB Framework-Product quality towards zero defects for melamine surface Boards industry. *IFAC-PapersOnLine*, 54(1), 570-575.
6. Fidalgo, A., (noviembre de 2017). *Metodologías educativas*. Innovación educativa. <https://innovacioneducativa.wordpress.com/2007/10/08/metodologias-educativas/>
7. García Sevilla, Julia. (2006). La metodología del Aprendizaje Basado en Problemas [Manuscrito enviado para su publicación]. Universidad de Murcia http://www.ub.edu/dikasteia/LIBRO_MURCIA.pdf
8. Gómez Moline, M. (2003). Algunos factores que influyen en el éxito académico de los estudiantes universitarios en el área de química. [Manuscrito para su publicación]. Universidad Autónoma de Barcelona. <https://www.tdx.cat/bitstream/handle/10803/4700/mgm1de1.pdf>

9. González, D., (15 de junio de 2020). *APRENDO EN CASA SEMANA 12: "Aplicamos la metodología Design Thinking: DEFINIR"*. Aprendo en casa. <https://donny-ept-2cd.blogspot.com/2020/06/aplicamos-la-metodologia-design.html?m=1>
10. González, D., (22 de junio de 2020). *APRENDO EN CASA SEMANA 12: "Aplicamos la metodología Design Thinking: idear"*. Aprendo en casa. <https://donny-ept-2cd.blogspot.com/2020/06/aprendo-en-casa-semana-12-aplicamos-la.html>
11. González, D., (28 de junio de 2020). *APRENDO EN CASA SEMANA 13: "Aplicamos la metodología Design Thinking: prototipar"*. Aprendo en casa. <https://donny-ept-2cd.blogspot.com/2020/06/aprendo-en-casa-semana-13-aplicamos-la.html>
12. González, D., (6 de julio de 2020). *Aprendo en casa semana 14: "Aplicamos la metodología Design Thinking: evaluar"*. Aprendo en casa. <https://donny-ept-2cd.blogspot.com/2020/07/aprendo-en-casa-semana-14-aplicamos-la.html>
13. González, D., (8 de junio de 2020). *APRENDO EN CASA SEMANA 10: "Aplicamos la metodología Design Thinking: empatizar"*. Aprendo en casa. <https://donny-ept-1abc.blogspot.com/2020/06/semana-10-aprendo-en-casa-aplicamos-la.html>
14. <https://docs.microsoft.com/es-ES/azure/architecture/data-science-process/overview>
15. Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications—a holistic extension to the CRISP-DM model. *Procedia Cirp*, 79, 403-408.
16. Ministerio de Educación Nacional. (2010). Manual de Implementación Escuela Nueva [Manuscrito enviado para publicación]. Ministerio de Educación Nacional, Republica de Colombia. https://www.mineducacion.gov.co/1621/articles-340089_archivopdf_orientaciones_pedagogicas_tomol.pdf
17. Nagashima, H., & Kato, Y. (2019, March). APREP-DM: a Framework for Automating the Pre-Processing of a Sensor Data Analysis based on CRISP-DM. In 2019 *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)* (pp. 555-560). IEEE.
18. Proceso de Conceptualización del Entendimiento del Negocio para Proyectos de Explotación de Información

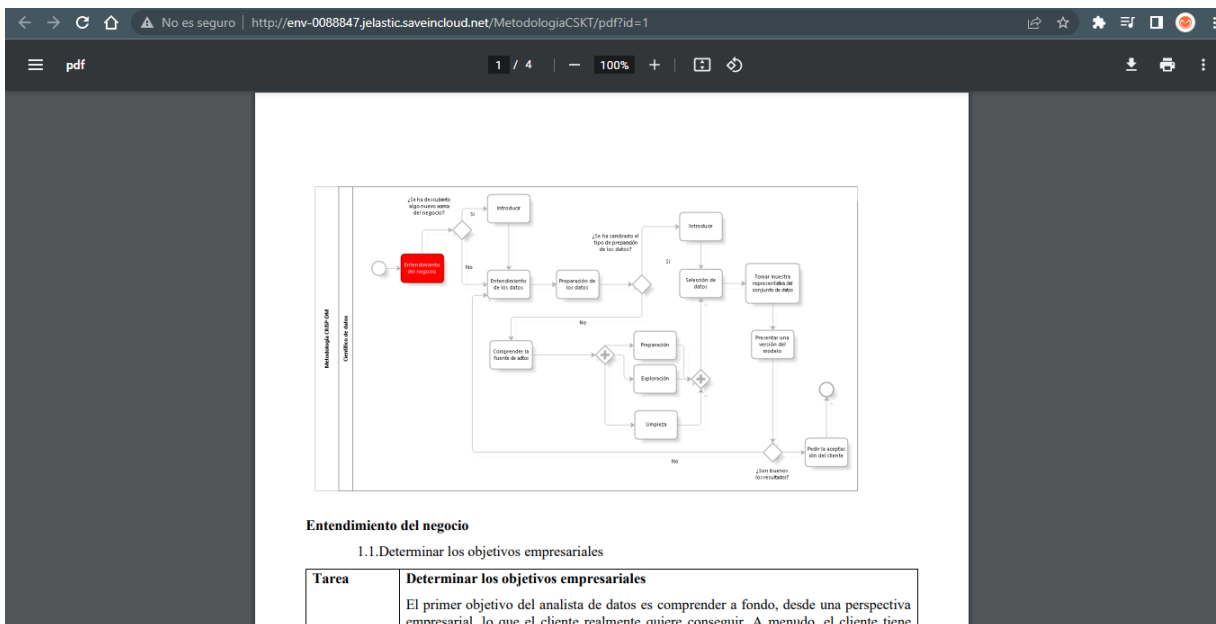
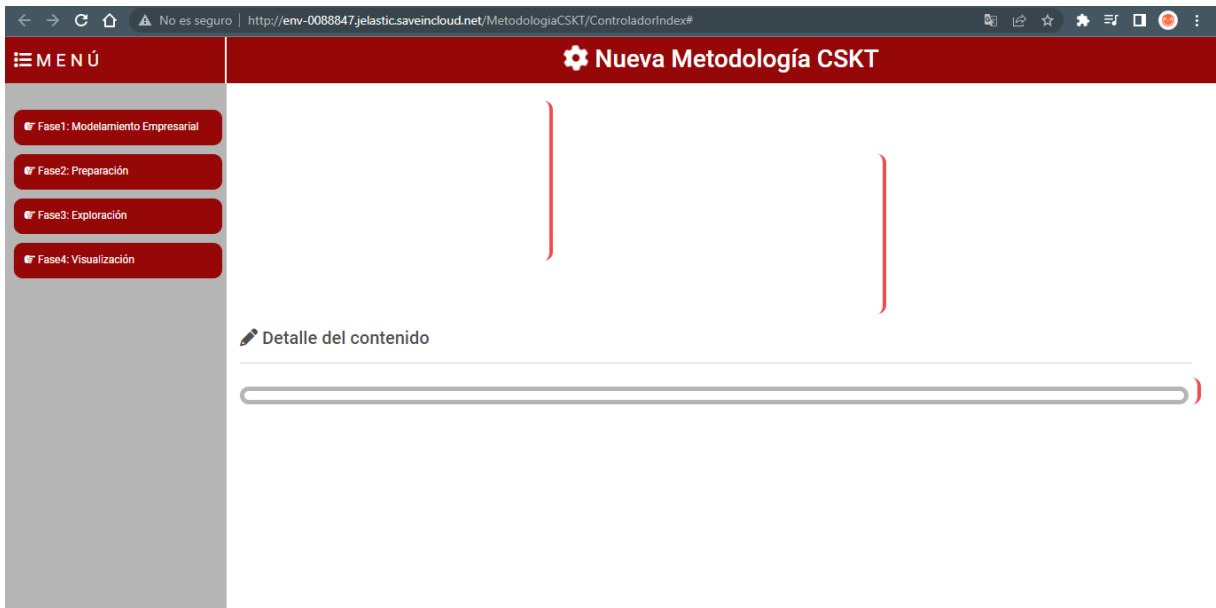
19. Rodríguez, A. *et al.* (2012). Desarrollo del pensamiento científico: proyecto innovación en formación científica. [Manuscrito para su publicación.] Instituto para la Investigación Educativa y el Desarrollo Pedagógico, IDEP <http://biblioteca.clacso.edu.ar/Colombia/idep/20151026052301/DesarrolloPensamiento Cientifico.pdf>
20. Rodríguez, B. (2020). Primera charla de DABIA: "Ciencia de datos e Ingeniería Industrial se parecen más de lo que imaginan". <https://www.grupodabia.com/post/2020-03-05-primera-charla-ingenieria-industrial-y-ciencia-de-datos/>
21. Rohanizadeh, S., & BAMENI, M. (2009). A proposed data mining methodology and its application to industrial procedures.
22. Sampieri, R. (2014). Metodología de la Investigación. 6ta Edición MrGraw-Hill. Recuperado de: <http://observatorio.epacartagena.gov.co/wp-content/uploads/2017/08/metodologia-de-lainvestigacion-sexta-edicion.compressed.pdf>.
23. Solarte, J. (2002). A proposed data mining methodology and its application to industrial engineering.
24. Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y Alvarado Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. En Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional (pp. 63-86). Bogotá: Ediciones Universidad Cooperativa de Colombia. doi: <http://dx.doi.org/10.16925/9789587600490>
25. Vara Horna, A. (2008). 7 pasos para una tesis exitosa, desde la idea inicial hasta la sustentación. [Manuscrito enviado para su publicación]. Facultad de Ciencias Administrativas y Recursos Humanos. Universidad de San Martín de Porres. <https://www.administracion.usmp.edu.pe/investigacion/files/7-PASOS-PARA-UNA-TESIS-EXITOSA-Desde-la-idea-inicial-hasta-la-sustentaci%C3%B3n.pdf>

Anexo A. Actualización del Sprint Burn Down Chart



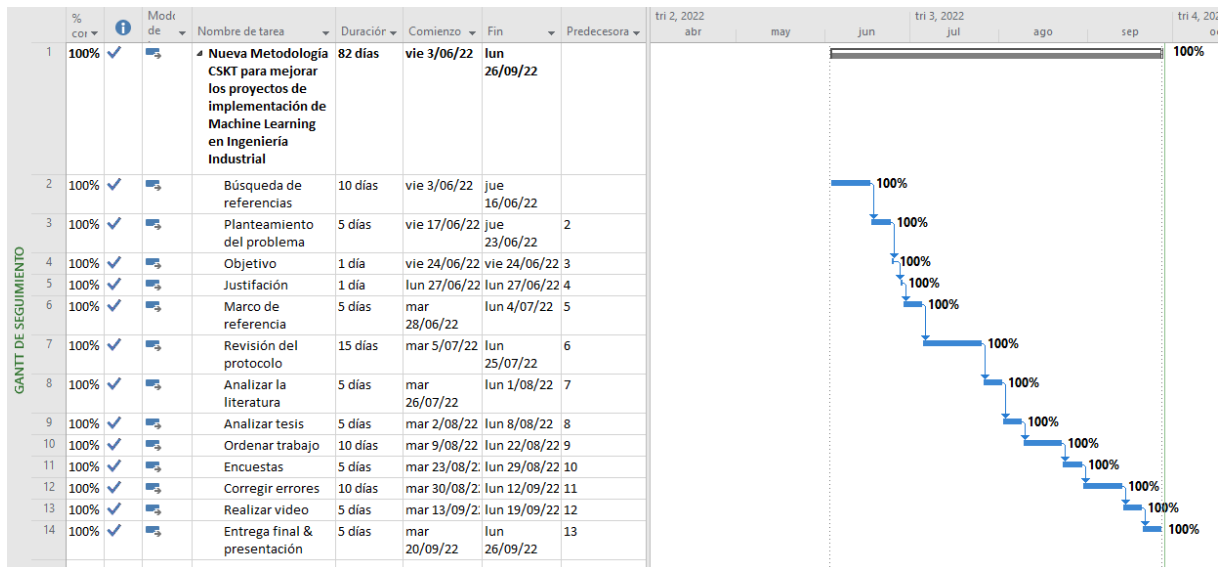
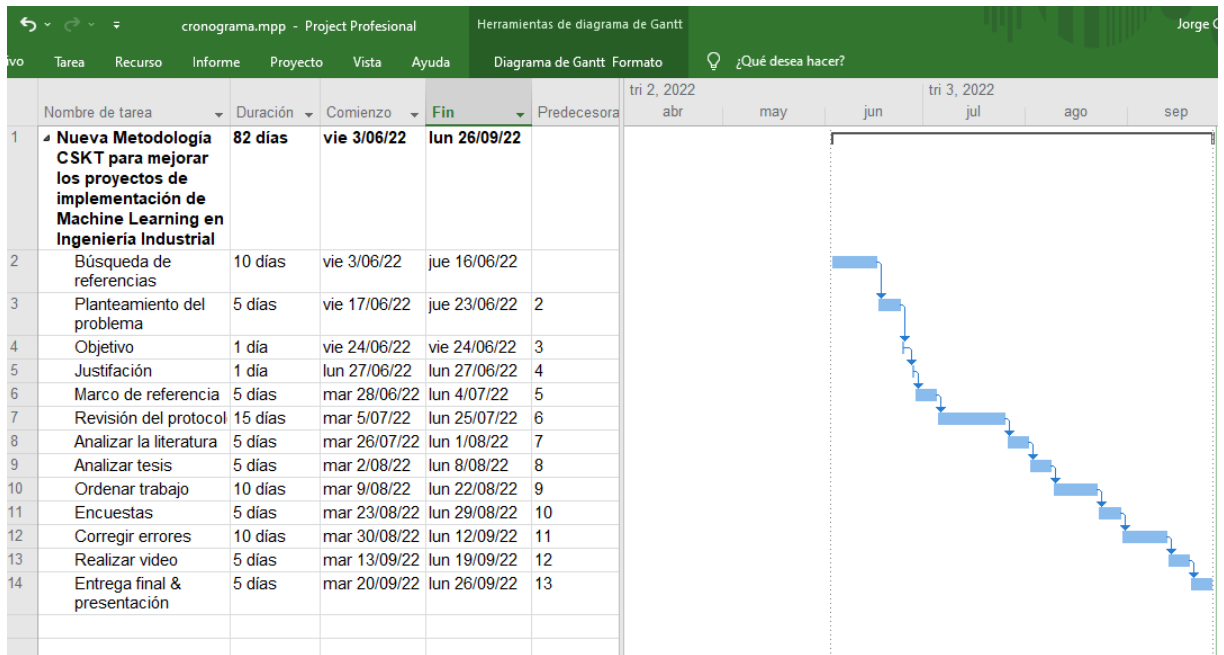
Explicación: Los primeros 4 días el equipo no trabajó, y quedaban 10 días, donde el equipo tuvo que recuperar en esos 10 días donde fuimos acercándonos al trabajo planeado respecto al trabajo real, pero precisamente el Scrum que se ha aplicado para este proyecto está medio automatizado y nos facilitó la adaptación al escenario del tiempo y con ello pudimos ir cambiando el estado de las actividades que pasaban a terminadas, lo cual se ajustó al proyecto y es recomendable para situaciones donde empiezas tarde pero puedes ir bajando los diferentes puntos de historia de forma muy sencilla.

Anexo B. Página web



Explicación: La parte tecnológica está desarrollada con PHP y de forma muy sencilla permite visualizar las guías que se compartieron para visualización del público en general en lo cual el link para acceder a la página web es <http://env-0088847.jelastic.saveincloud.net/MetodologiaCSKT/ControladorIndex#>

Anexo C. Cronograma del proyecto de investigación



Anexo D. Evidencia de encuestas

Mejoras para la Nueva Metodología CSKT (CRISP-DM + SEMMA + KDD + TDSP) - Guardado

Vista previa Tema Recopilar respuestas

Preguntas Respuestas 36

Mejoras para la Nueva Metodología CSKT (CRISP-DM + SEMMA + KDD + TDSP)

El propósito de esta encuesta es ayudarnos a entender hasta qué punto la metodología propuesta le facilita o facilitará la implementación de machine learning para un proceso que requiere ser automatizado. No hay respuestas 'correctas' o 'equivocadas', nosotros estamos interesados sólo en su opinión. Sus respuestas serán tratadas con alto grado de confidencialidad y no afectarán su evaluación. Sus respuestas, pensadas cuidadosamente, nos ayudarán a mejorar la manera de presentar esta nueva metodología en el futuro. Muchas gracias.

Todas las respuestas son necesarias y deben ser contestadas.

Sección 1

Mejoras para la Nueva Metodología CSKT (CRISP-DM + SEMMA + KDD + TDSP) - Guardado

Vista previa Tema Recopilar respuestas

Preguntas Respuestas 36

Mejoras para la Nueva Metodología CSKT (CRISP-DM + SEMMA + KDD + TDSP)

| | | |
|------------------|--------------------------------------|------------------|
| 36 Respuestas | 02:42 Tiempo medio para finalizar | Activo Estado |
|------------------|--------------------------------------|------------------|

Revisar respuestas Publicar puntuaciones Abrir en Excel

1. Podrá utilizar mecanismos simples para la planificación de tareas cuando necesita automatizar tareas sencillas de manufactura (0 punto)

Más detalles Información

Totalmente en desacuerdo 2

Anexo E. Evidencia de video

