

**UNIVERSIDAD LAICA ELOY ALFARO DE MANABI**  
**FACULTAD DE CIENCIAS INFORMATICAS**  
**CARRERA DE TECNOLOGIA DE LA INFORMACION**

**TEMA:**

METODOLOGÍAS Y MODELOS DE PROYECTOS DE MINERÍA DE DATOS

**NOMBRE:**

MACIAS PICO JOSSELYN STEFANY

**CURSO:**

SEXTO “B”

**MATERIA:**

MINERIA DE DATOS

**DOCENTE:**

ING. FABRICIO JAVIER RIVADENEIRA ZAMBRANO

**FECHA:**

25-05-2021

**MANTA-MANABI-ECUADOR**

## INDICE

INTRODUCCION .....	3
Metodologías y Modelos de proyectos de Minería de Datos .....	4
CRISP-DM: .....	4
Fases: .....	4
TDSP: .....	9
Fases: .....	10
CATALYST .....	12
Fases: .....	12
ESCENARIOS .....	14
Ventajas Y Desventajas de Metodologías y Modelos .....	15
Semejanzas de Metodologías y Modelos .....	16
Diferencias de Metodologías y Modelos.....	17
CONCLUSIONES .....	18
RECOMENDACIONES .....	18
BIBLIOGRAFÍA.....	19

## Índice de imágenes

IMAGEN 1: Modelo del proceso CRISP-DM .....	4
IMAGEN 2: Tareas de la Comprensión de negocio.....	5
IMAGEN 3: Comprensión de los datos .....	5
IMAGEN 4: Preparación de los datos.....	6
IMAGEN 5: Modelado.....	7
IMAGEN 6: Evaluación.....	8
IMAGEN 7: Implantación.....	9
IMAGEN 8: Ciclo de vida TDSP.....	11
IMAGEN 9:Ejemplo de trabajo en varios proyectos compartiendo componentes .....	11
IMAGEN 10: Relaciones P3TQ.....	12
IMAGEN 11: Cadena de valor empresarial .....	12

## INTRODUCCION

La minería de datos es una manera de aprender del pasado para tomar mejores decisiones en el futuro y así tener excelentes resultados. Cada una de las metodologías esta basada en mejores prácticas estas están realizadas mediante análisis estadísticos.

A finales de los 80 las estadísticas se ampliaron a varias técnicas como difusa, razonamiento heurístico y redes neuronales. Actualmente estas técnicas generan mayor conocimiento. La minería de datos permite extraer patrones tendencias y regularidades para describir y comprender mejor cada uno de los datos.

Diariamente se almacenan gigantescas cantidades de datos, esta información que es almacenada no representa mayores costos y son almacenados con el fin de poder analizarlos mas adelante. En minería de datos no existe un solo enfoque sino un conjunto de técnicas que se pueden utilizar de manera independiente o en combinación. Mediante consultas simples sobre estos datos se pueden obtener variedades de resultados a medida que crece la complejidad de la base de datos y el número de registros

En este documento se explicará sobre algunas de las metodologías como CRISP-DM la cual es un estándar en el mercado, TDSP esta es una metodología ágil e interactiva además esta incluye ls mejores practicas y estructuras de Microsoft. Y por ultimo CATALYST se dará a conocer más informacion durante el informe.

## Metodologías y Modelos de proyectos de Minería de Datos

### CRISP-DM:

Para implementar una tecnología en un negocio, se requiere de una metodología. La mayoría de las consultoras especializadas en alguna tecnología cuentan, con por lo menos, una metodología, según los tipos de proyectos que aborden. Estos métodos son definidos a partir de sus experiencias y tomando lo mejor de los procedimientos más exitosos o populares. Contar con una metodología, se ha convertido tan importante y necesario como la carta de presentación de las empresas.

Esta es una metodología utilizada para minería de datos para la salud en su siglas en ingles significan **Cross Industry Standard Process for Data Mining** esta se encuentra dividida en cuatro niveles de abstracción organizados de forma jerárquica en tareas que van desde el nivel mas general hasta la tarea más específica y así puede organiza un proyecto de Data Mining utilizando seis fases.

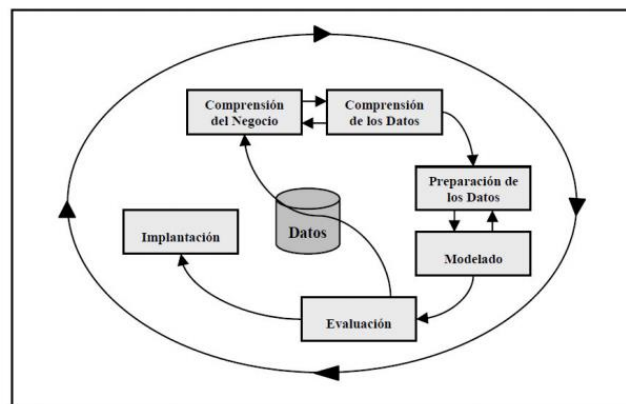


IMAGEN 1: Modelo del proceso CRISP-DM

### Fases:

- **Comprensión del negocio:**

En esta fase lo que se realiza es alinear los objetivos del proyecto con los objetivos de negocio en esta fase se debe tener en cuenta lo siguiente:

- Comprensión del negocio.
- Comprensión de los datos.
- Preparación de los datos.
- Modelado.
- Evaluación.
- Despliegue.

Una vez que se establecen y se alinean los objetivos se creará un plan de proyecto que incluya las fases, tareas y actividades necesarias para lograr conseguir los objetivos.

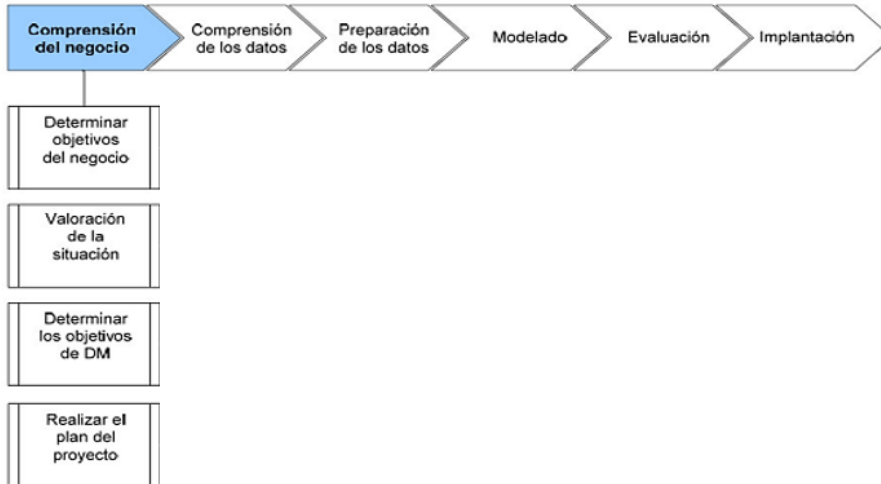


IMAGEN 2: Tareas de la Comprensión de negocio

- **Compresión de los datos.**

En esta parte se logra conocer los datos la estructura y su distribución y así poder mejorar la calidad de estos. En esta fase se debe ser capaz de:

- Ejecutar procesos de captura de datos.
- Proporcionar una descripción del juego de datos.
- Realizar tareas de exploración de datos.
- Gestionar la calidad de los datos, identificando problemas y proporcionando soluciones.

El proyecto está formado por un conjunto de 569 observaciones con 30 variables de tipo real. El conjunto de datos no presenta valores nulos. Las variables son las siguientes: radio, textura, perímetro, área, suavidad, compacidad, concavidad, puntos cóncavos, simetría y dimensión fractal

El conjunto de datos está formado por la media, el error estándar y el peor de los valores de las variables para cada una de las imágenes.

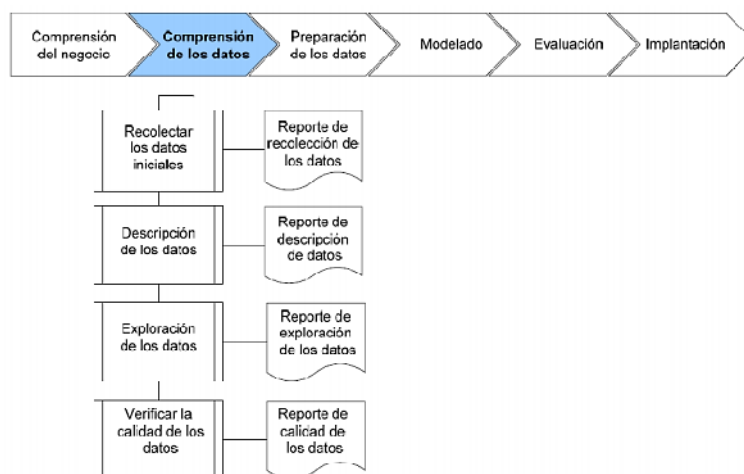


IMAGEN 3: Compresion de los datos

- **Preparación de los datos:**

Una vez efectuada la recolección de datos se procede a la preparación de la adaptación a las técnicas de Data Mining así como la visualización de datos de búsqueda para las relaciones entre variables y exploración de datos, a esta preparación de datos se le incluye tareas generales que se aplican a técnicas de modelado, limpieza de datos, generación de variables adicionales, integración de orígenes de datos y cambios de formato.

Esta se relaciona con la fase de modelado, ya que los datos deben ser procesados de diferentes formas, es así como tanto la fase de preparación y la de modelado realizan trabajos permanentes.

En esta fase debemos ser capaces de:

- Establecer el universo de datos con los que trabajar.
- Realizar tareas de limpieza de datos.
- Construir un juego de datos apto para ser usado en modelos de minería de datos.
- Integrar datos de fuentes heterogéneas si es necesario.

Otro análisis relevante que nos permitiría descartar atributos redundantes para la construcción del modelo sería la matriz de correlación (Arancibia)

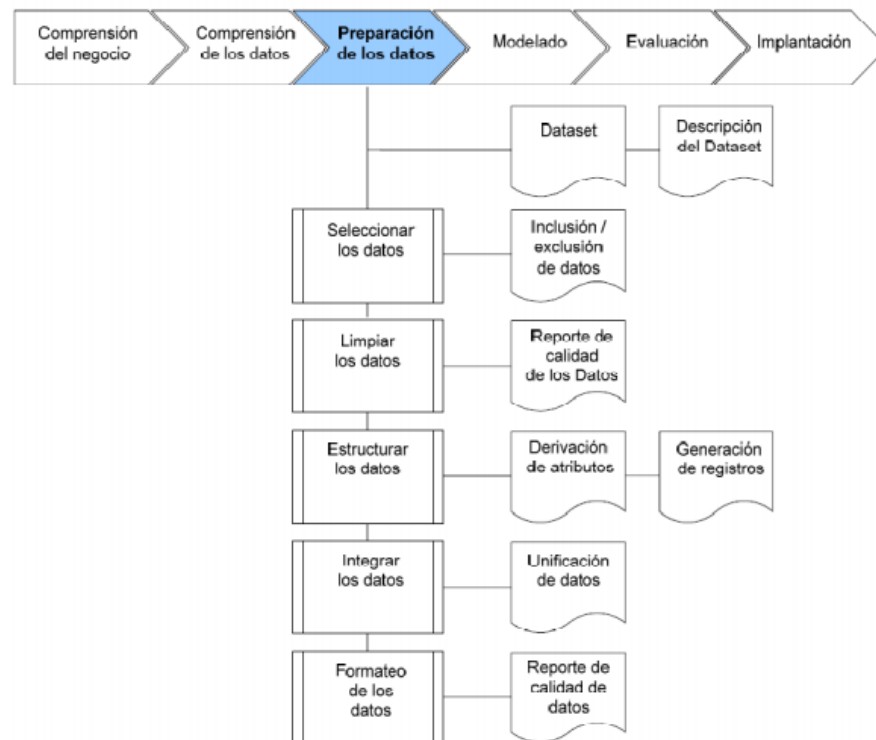


IMAGEN 4: Preparación de los datos

- **Modelado.**

En esta fase ñp que se busca es construir un modelo el cual permita alcanzar los objetivos de los proyectos, seleccionando las técnicas de modelado más apropiadas para el proyecto donde se deben tener en cuenta los siguientes criterios:

- Ser apropiada al problema

- Disponer de datos adecuados
- Cumplir los requisitos del problema
- Tiempo adecuado para obtener un modelo
- Conocimiento de la técnica

Los parámetros utilizados en la generación del modelo, dependen de las características de los datos y de las características de precisión que se quieran lograr con el modelo.

Las principales tareas que se deben realizar en esta fase se nombran a continuación:

- **Selección de la técnica de modelado:** Esta tarea consiste en la selección de la técnica de DM más apropiada al tipo de problema a resolver. En esta se debe considerar el objetivo principal del proyecto y la relación con las herramientas de DM existentes
- **Generación del plan de prueba:** Una vez construido un modelo, se debe generar un procedimiento destinado a probar la calidad y validez de este.
- **Construcción del modelo:** Después de haber seleccionado la técnica, se ejecuta sobre los datos que ya han sido preparados previamente para generar uno o más modelos. La selección de los mejores parámetros es un proceso iterativo y se basa exclusivamente en los resultados generados.
- **Evaluación del modelo:** los ingenieros de DM interpretan los modelos de acuerdo con el conocimiento preexistente del dominio y los criterios de éxito preestablecidos. Algunos de los expertos en el dominio del problema juzgan los modelos del dominio y expertos en Data Mining aplican sus propios criterios. (Rueda, s.f.)

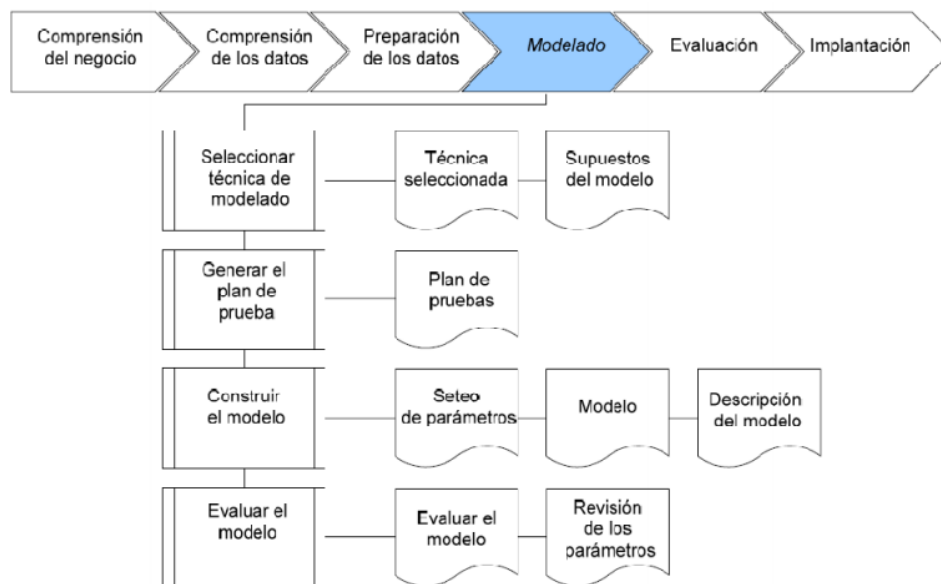


IMAGEN 5: Modelado

### • Evaluación.

En esta fase nos centramos en evaluar el grado de acercamiento del modelo a los objetivos de negocio. Para continuar con esta fase se debe ser capaz de:

- Evaluar el modelo o modelos generados hasta el momento.
- Revisar todo el proceso de minería de datos que nos ha llevado hasta este punto.

- Establecer los siguientes pasos a tomar, tanto si se trata de repetir fases anteriores como si se trata de abrir nuevas líneas de investigación.

Si la precisión alcanzada por nuestro modelo no fuese la adecuada para satisfacer los objetivos del proyecto entraríamos en un proceso iterativo con el objetivo de conseguir un modelo mejor y esto implicaría:

- Tuning de los hiperparámetros del modelo.
- Seleccionar otra técnica de construcción de modelos: SVM, redes neuronales, etc...

Se debe considerar que se puedan emplear múltiples herramientas para la interpretación de los resultados.

Las tareas involucradas en esta fase del proceso son las siguientes:

- **Evaluación de los resultados:** Esta tarea involucra la evaluación del modelo en relación con los objetivos del negocio y busca determinar si hay alguna razón de negocio para la cual, el modelo sea deficiente, o si es aconsejable probar el modelo, en un problema real si el tiempo y restricciones lo permiten. (Rueda, s.f.)
- **Proceso de revisión:** El proceso de revisión, se refiere a calificar al proceso entero de DM, al objeto de identificar elementos que pudieran ser mejorados.

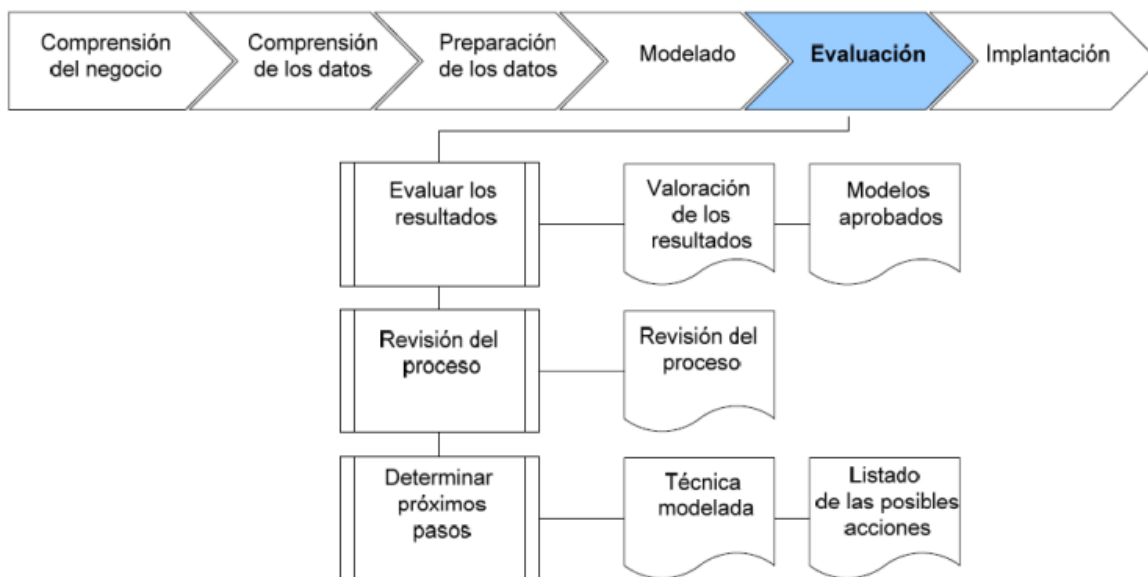


IMAGEN 6: Evaluación

### • **Implementación.**

Una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio, ya sea que el analista recomiende acciones basadas en la observación del modelo y sus resultados, ya sea aplicando el modelo a diferentes conjuntos de datos o como parte del proceso. Generalmente un proyecto de Data Mining no concluye en la implantación del modelo, pues se deben documentar y presentar los resultados de manera comprensible para el usuario, con el objetivo de lograr un incremento del conocimiento.

Se debe ser capaz de:



- Diseñar un plan de despliegue de modelos y conocimiento sobre nuestra organización.
- Realizar seguimiento y mantenimiento de la parte más operativa del despliegue.
- Revisar el proyecto en su globalidad con el objetivo de identificar lecciones aprendidas.

Y se deben realizar las siguientes tareas:

- Implantación de un sistema de captura digital de la imagen de anatomía patológica.
- Implantación de software de clasificación automática de tejidos de cáncer de mama.
- Integración con el software del laboratorio de anatomía patológica.
- Implantación de nuevos procedimientos operativos del laboratorio de anatomía patológica.
- Monitorización del funcionamiento del nuevo sistema y grado de adherencia de los usuarios al mismo.
- Medición del grado de precisión diagnóstica y la eficiencia del laboratorio tras la puesta en marcha del sistema. (Arancibia)

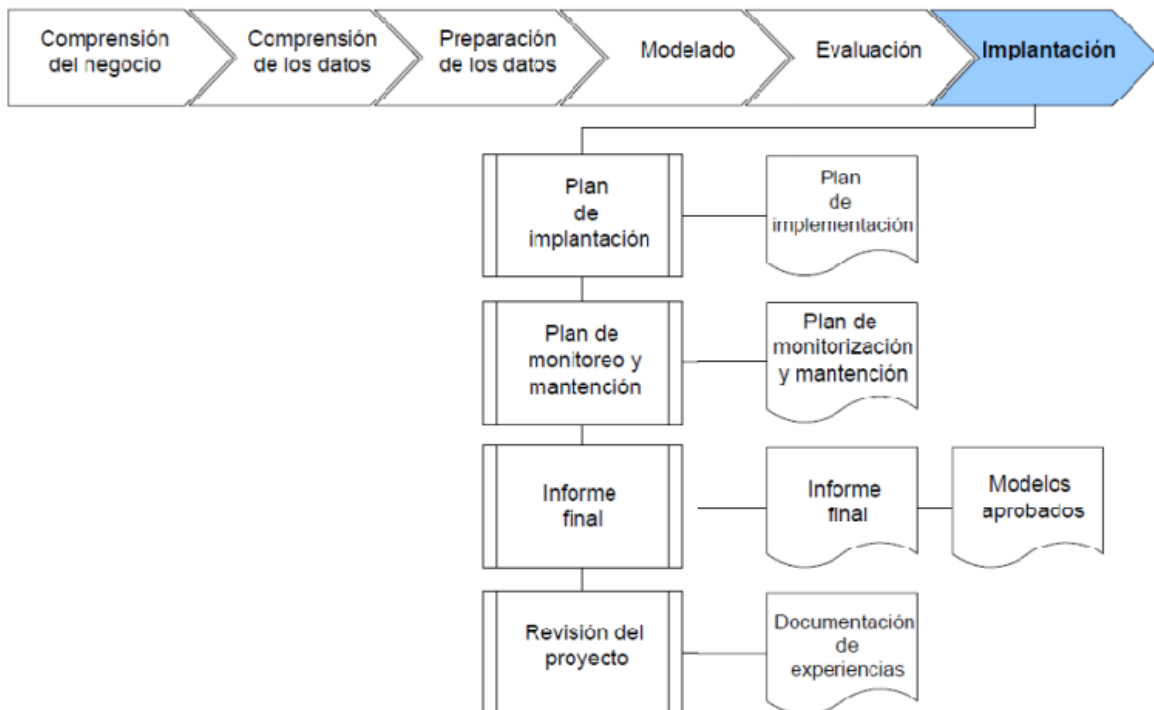


IMAGEN 7: Implantación

## TDSP:

La TDSP es una metodología de ciencia de datos ágil e iterativa utilizada para ofrecer soluciones de análisis predictivo y aplicaciones inteligentes de manera eficiente. Además, ayuda a mejorar la colaboración y el aprendizaje del equipo al sugerir cómo los roles del equipo funcionan mejor juntos. También incluye las mejores prácticas y estructuras de Microsoft y otros líderes de la industria para lograr una implementación exitosa de las iniciativas de ciencia de datos. El objetivo es ayudar a empresas para que aprovechen los beneficios de su programa de análisis.

Más adelante conoceremos algunas descripciones genéricas del proceso que se puede implementar con diferentes tipos de herramientas. También se proporcionará una descripción más detallada de las

tareas y roles del proyecto involucrados en el proceso. También se proporciona orientación sobre cómo implementar el TDSP utilizando un conjunto específico de herramientas e infraestructura de Microsoft que usamos para implementar el TDSP en nuestros equipos.

Esta metodología cuenta con algunos componentes como:

- Una definición del 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 del proyecto

#### *Ciclo de vida de la ciencia de datos*

El proceso de ciencia de datos en equipo (TDSP) proporciona un ciclo de vida para estructurar el desarrollo de sus proyectos de ciencia de datos.

Si está utilizando otro ciclo de vida de ciencia de datos, como CRISP-DM, KDD o el propio proceso personalizado de su organización, aún puede usar el TDSP basado en tareas en el contexto de esos ciclos de vida de desarrollo.

Este ciclo de vida esta sido diseñado para proyectos de ciencia de datos que se envían como parte de aplicaciones inteligentes las cuales implementan modelos de aprendizaje automático o inteligencia artificial para análisis predictivos. Los proyectos de ciencia de datos exploratorios o los proyectos de análisis improvisados también pueden beneficiarse del uso de este proceso. Pero en tales casos, es posible que algunos de los pasos descritos no sean necesarios. (Data Science, s.f.)

#### **Fases:**

Las fases de esta metodología tienen alguna similitud a la CRISP-DM

- ***Comprensión Empresarial***

En esta fase se define cada uno de los objetivos e identifica alguna de las fuentes de datos.

- ***Adquisición y comprensión de datos***

Se debe ingerir datos y determinar si pueden responder a la pregunta de presentación (para hacerla más efectiva se puede realizar una combinación la comprensión y la limpieza de datos de CRISP-DM)

- ***Modelado***

En esta se realiza una ingeniería de características y entrenamiento de modelos (se puede realizar una combinación de modelado y evaluación)

- ***Despliegue***

Se debe implementar en un entorno de producción.

- ***Aceptación del cliente***

En esta parte se realiza una validación del cliente tomando en cuenta si el sistema satisface las necesidades comerciales (una fase no cubierta por CRISP-DM)

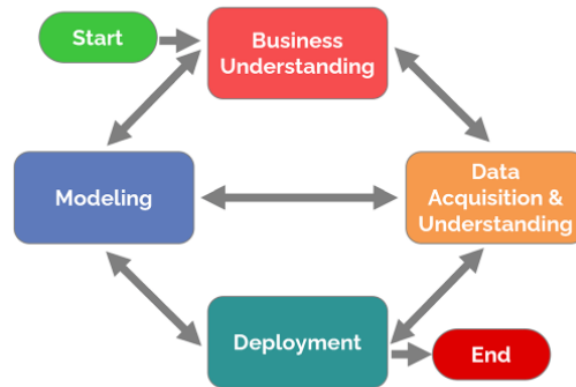


IMAGEN 8: Ciclo de vida TDSP

TDSP aborda la debilidad de la falta de definición de equipo de CRISP-DM mediante la definición de seis roles:

- Arquitecto de soluciones
- Gerente de proyecto
- Ingeniero de datos
- Científico de datos
- Desarrollador de aplicaciones
- Líder del Proyecto

Microsoft es el que define las tareas y los artefactos relevantes para muchos de los roles del equipo durante cada fase del ciclo de vida del proyecto. (Microsoft Build , 2020)

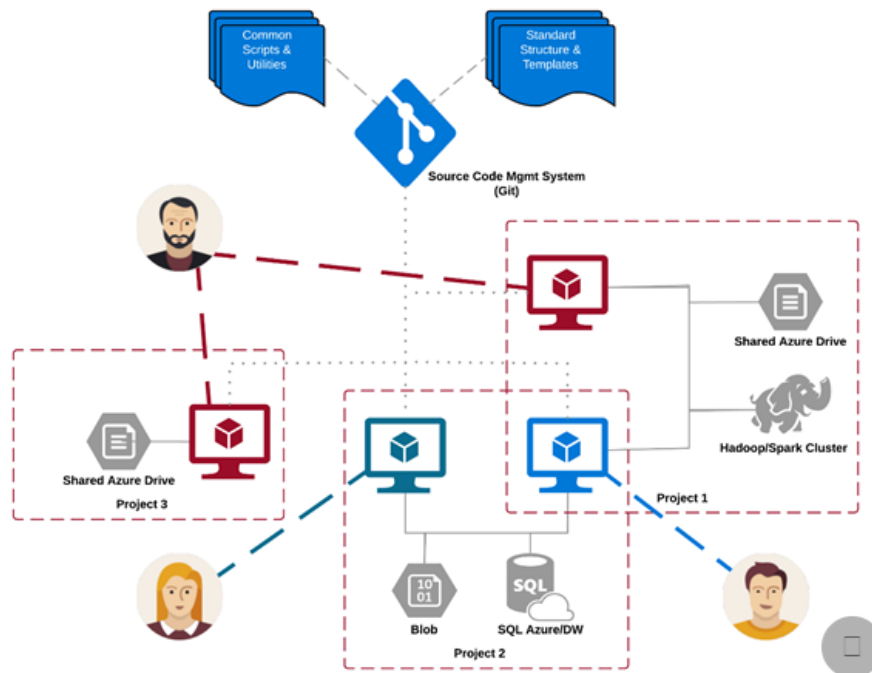


IMAGEN 9: Ejemplo de trabajo en varios proyectos compartiendo componentes

## CATALYST

Esta es una metodología utilizada para el proceso de extracción de conocimiento en bases de datos a pesar de ser una metodología muy completa no tiene tanto éxito como CRISP-DM

Para proyectos donde el problema u oportunidad del negocio en el cual se va a trabajar no está definido, se recomienda comenzar analizando las relaciones de P3TQ que son:

- Product (Producto)
- Place (Lugar)
- Price (Precio)
- Time (Tiempo)
- Quantity (Cantidad)

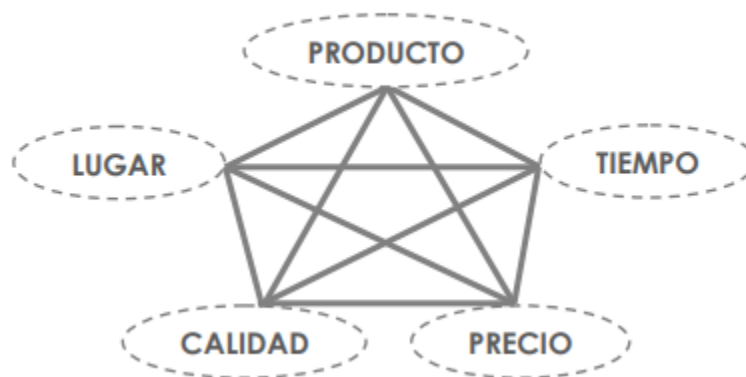


IMAGEN 10: Relaciones P3TQ

Estas se refieren a tener el producto correcto, en el lugar adecuado, en el momento adecuado, en la cantidad correcta y con el precio correcto.

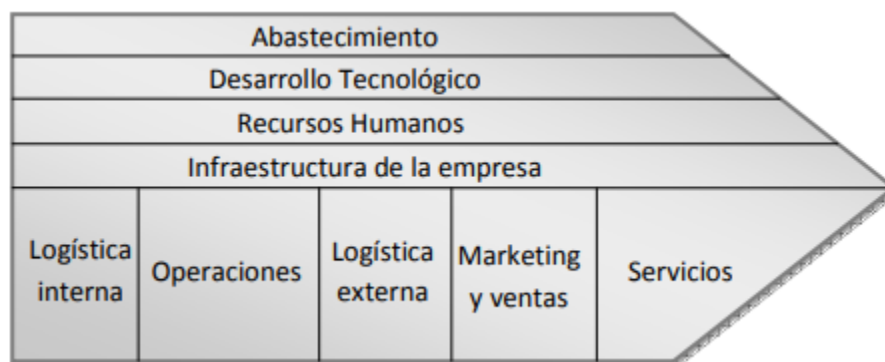


IMAGEN 11: Cadena de valor empresarial

La cadena de valor empresarial mostrada en la imagen anterior define cada una de las actividades de la empresa que añaden valores al producto a medida que pasa por cada una:

### Fases:

- **Preparación de los datos**

En la preparación de los datos se va a tener en cuenta lo siguiente:



1. Evaluar las variables en estudio (medidas de posición y dispersión, outliers, datos faltantes, etc).
2. Chequear problemas básicos en las variables (variables con muchas categorías por ejemplo).
3. Chequear problemas en la base de datos completa, análisis multivariado (CHAID analysis).
4. Chequear variables anacrónicas (que no aportan información).
5. Chequear que haya suficientes datos.
6. Chequear que se cubran todos los valores posibles de las variables, aun los que no son de interés.
7. Chequear la necesidad de recodificar variables.

- ***Selección de herramientas y modelado inicial***

En la segunda fase lo que se hace es seleccionar las herramientas y el modelo inicial para poder continuar a la siguiente fase esto se lo realiza de la siguiente manera:

1. Estructurar los datos para el proceso de minería (dividir los datos de entrenamiento, prueba y evaluación).
2. Caracterizar las variables de entrada y salida.
3. Seleccionar el algoritmo de minería.
4. Evaluar el impacto de los datos faltantes mediante un MVCN (Missing Value Check Model).
5. Crear un modelo inicial: este modelo inicial se lo puede hacer de algunas maneras las cuales serán mostradas a continuación:
  - a. Exploratorio / Descriptivo: Si se realiza minería para entender la situación del negocio.
  - b. De Clasificación: si se realiza minería para clasificar.
  - c. De Predicción: si se realiza minería para predecir.

- ***Refinar el modelo***

Para continuar con esta fase se necesita tener en cuenta el modelo inicial en que esta realizado:

1. Si el método es exploratorio, describir los resultados encontrados sobre la situación actual.
2. Si el modelo es predictivo o de clasificación, verificar la capacidad predictiva del modelo (por ejemplo, con matrices de confusión o gráficos de “valores predichos vs observados”, según el caso).
3. Verificar el modelo con el personal interesado.

- ***Implementar el modelo***

1. Si el modelo es exploratorio: se deben revisar los requerimientos del problema y así poder elaborar un informe con los resultados del descubrimiento y así poder contabilizar los valores extremos y así poder incorporar la evidencia negativa e incorporar evidencia empírica/experimental y obtener realimentación de los usuarios.
2. Si el modelo es de clasificación/predicción: se deben revisar los requerimientos de la implementación planteados antes de la minería, así también revisar los requerimientos del problema y poder preparar una explicación del modelo y así poder revisar los requerimientos finales de implementación.
3. Por último, se desarrolla la comunicación y la difusión de resultados.

## **ESCENARIOS**

### **ESCENARIO 1: DATOS**

Este escenario se realizó con el fin de explorar los datos en búsqueda de relaciones útiles e interesantes.

1. Determinar las fuentes de donde se recolectarán los datos.
2. Identificar al personal interesado (stakeholders) en el proyecto.
3. Discutir el proyecto original con el personal interesado.
4. Caracterizar el conjunto de datos en función de las relaciones P3TQ por las cuales fueron recolectados.
5. Caracterizar la motivación del negocio para recolectar y almacenar los datos.
6. Descubrir quién o qué departamento originó el proyecto y que expectativas tienen sobre el mismo.
7. Descubrimiento del problema: Al descubrir un problema se necesita realizar los siguientes pasos
  - a. Identificar las principales relaciones P3TQ que dan origen a los datos.
  - b. Identificar y caracterizar al personal interesado.
  - c. Identificar los objetos organizacionales que los datos representan.
  - d. Enmarcar el problema u oportunidad.
  - e. Preparar un esbozo del caso de negocio.
  - f. Presentar el nuevo proyecto al personal interesado.
  - g. Armar el caso de negocio completo, si es necesario.
  - h. Enmarcar y describir la situación del negocio.
  - i. Definir los requerimientos de la implementación.

### **ESCENARIO 2: PROBLEMA/OPORTUNIDAD**

Este escenario esta realizado con la finalidad al haberse dado algún problema u oportunidad de negocio y así poder ver cómo la minería de datos puede colaborar con la misma.

1. Identificar y caracterizar al personal interesado relevante.
2. Explorar la situación de negocio con el personal interesado.
3. Enmarcar y describir la situación del negocio.
4. Identificar los objetivos de negocio relevantes para el proyecto.
5. Buscar los datos que se explotarán.
6. Armar el caso de negocio.
7. Presentar el caso de negocio al personal interesado.
8. Describir la situación del negocio para el proceso de minería.
9. Definir los requerimientos de implementación.

### **ESCENARIO 3: PROSPECCION**

Este escenario está diseñado para descubrir dónde la minería de datos puede aportar valor en la organización.

1. Caracterizar las relaciones P3TQ claves de la organización.
2. Identificar el flujo de los principales procesos de la organización.
3. Identificar el personal interesado.
4. Entrevistar al personal interesado.
5. Descubrir qué “cambios estratégicos” pueden resultar de mayor interés para cada usuario.

6. Caracterizar los modelos de minería que pueden dar soporte a los cambios estratégicos.
7. Explorar las fuentes de datos.
8. Preparar un borrador del caso de negocio para cada oportunidad significativa.
9. Presentar los casos de negocio al personal interesado.
10. Enmarcar la situación de negocio que se abordará.
11. Definir los requerimientos de implementación.

## Ventajas Y Desventajas de Metodologías y Modelos

COMPARACIONES		
Metodologías y Modelos de Proceso	VENTAJAS	DESVENTAJAS
CRISP-DM:	<ul style="list-style-type: none"> <li>- Mayor competitividad</li> <li>- Satisfacción del cliente</li> <li>- Disminución y cumplimiento de los plazos</li> <li>- Reducción de tiempo</li> </ul>	<ul style="list-style-type: none"> <li>- Puede aparecer algún inconveniente a la hora de utilizar esta metodología</li> <li>- Dependiente del tipo de datos que se quieran recopilar, nos puede llevar mucho trabajo</li> </ul>
TDSP	<ul style="list-style-type: none"> <li>- Enfatiza la necesidad de entregables incrementales.</li> <li>- TDSP reconoce que la ciencia de datos y la ingeniería de software está diseñado para equipos que trabajan en proyectos de producción.</li> <li>- TDSP se puede implementar tal como se define o junto con otros enfoques como CRISP-DM.</li> <li>-</li> </ul>	<ul style="list-style-type: none"> <li>- TDSP aprovecha los sprints de planificación de duración fija con los que luchan muchos científicos de datos.</li> <li>- No toda la documentación de Microsoft es coherente.</li> </ul>
Catalyst	<ul style="list-style-type: none"> <li>- Posee buena documentación.</li> <li>- Es multiplataforma.</li> <li>- Posee una numerosa comunidad de desarrolladores y usuarios</li> </ul>	<ul style="list-style-type: none"> <li>- Lentitud al inicio de su ejecución.</li> <li>- Sin control de excepciones</li> </ul>



## Semejanzas de Metodologías y Modelos

<b>SEMEJANZAS ENTRE METODOLOGIAS Y MODELOS DE PROYECTOS DE MINERIA DE DATOS</b>			
	<b>CRISP-DM</b>	<b>TDSP</b>	<b>CATALYST</b>
ETAPAS INTERACTIVAS	SI	SI	SI
EVALUACION DEL RESULTADO	Basado en los objetivos del proyecto	Basado en los objetivos del proyecto	Basado en los objetivos del proyecto
ORIGINADA PARA MINERIA DE DATOS	SI	SI	SI
ELECCION DE HERRAMIENTAS	Libres y comerciales	Libres y comerciales	Libres y comerciales
ENFOQUE	proporcionar a los equipos de ciencia de datos un enfoque sistemático	proporcionar a los equipos de ciencia de datos un enfoque sistemático,	proporcionar a los equipos de ciencia de datos un enfoque sistemático
COMPRESION	comprensión del negocio	comprensión del negocio	
ACTIVIDADES ESPECIFICAS	Se definen actividades específicas para cada fase del proceso		Se definen actividades específicas para cada fase del proceso

(Cáceres, 2016)



## Diferencias de Metodologías y Modelos

DIFERENCIAS DE LAS METODOLOGIAS Y MODELOS DE PROYECTOS DE MINERIA DE DATOS			
DETALLE	CRISP-DM	TDSP	CATALYST
<b>FASES</b>	Entendimiento del negocio Entendimiento de los datos Preparación de los datos Modelado Evaluación Despliegue	Si tienen etapas interactivas Originada para minería de datos	Preparación de los datos Modelado Refinar el modelo Implementar el modelo Comunicación de resultados
<b>ETAPAS INTERACTIVAS</b>	Año de creación 1999	Año de creación 2016	Año de creación 2003
<b>EVALUACION DE RESULTADO</b>	Basado en el modelo y los objetivos del proyecto	Toma en cuenta mucho mas la satisfacción del cliente.	Basado en los objetivos del proyecto
<b>AÑOS DE CREACION</b>	Año de creación 1999	Año de creación 2016	Año de creación 2003
<b>ANALISIS Y COMPRESION DEL NEGOCIO</b>	Comprensión del negocio	Modelado del negocio	Modelado del negocio
<b>EVALUACION</b>	Evaluación		Refinamiento del modelo
<b>IMPLEMENTACION</b>	Despliegue	Aceptación	Comunicación
<b>METODOLOGIA</b>	Es una metodología abierta, no está ligada a ningún producto comercial,	Metodología ágil e iterativa de ciencia de datos basada en las mejores prácticas de Microsoft	Una metodología para el proceso de extracción de conocimiento en bases de datos
<b>OTROS</b>		Aprovechar marcos y herramientas bien conocidos como el control de versiones de git.	Selección de herramientas y modelado inicial

## **CONCLUSIONES**

Algunas de las distintas metodologías y modelos de proceso que existen en la actualidad se encuentran vigentes para proyectos de minería de datos de las cuales con mayor exaltación han sido detalladas en este informe. Ya que estos retornan una serie de beneficios y soluciones para el cliente como para el desarrollador.

Se estudia y practica el nivel de especificación de cada una de las tareas plantadas, los escenarios de aplicación, las actividades que componen cada fase del proceso y la incorporación de actividades para la dirección del proyecto.

Durante el estudio se ha logrado conocer y entender un poco más de las metodologías y modelos de proyectos para minería de datos lo cual nos ayudara a seguir desarrollando actividades de acuerdo a cada proyecto.

## **RECOMENDACIONES**

Se recomienda que el proceso de minería de datos siempre deba colaborar con una situación organizacional, como un problema u oportunidad.

No se debe trabajar directamente con los datos sino establecer de antemano la problemática que se va a abordar, el personal involucrado y las expectativas y necesidades de los usuarios.

Este punto resulta de gran importancia para justificar la realización del proyecto, ya que difícilmente una organización compre una herramienta si no sabe la función que cumplirá.

## BIBLIOGRAFÍA

Arancibia, J. A. (s.f.). *Metodología para el Desarrollo de Minería de Datos CRISP-DM*.

Cáceres, R. R. (18 de 10 de 2016). *rscs*. Obtenido de  
[https://www.rcs.cic.ipn.mx/2016\\_124/Mineria%20de%20datos%20centrada%20en%20el%20usuario%20para%20el%20analisis%20de%20la%20supervivencia%20y%20mortalidad.p  
df](https://www.rcs.cic.ipn.mx/2016_124/Mineria%20de%20datos%20centrada%20en%20el%20usuario%20para%20el%20analisis%20de%20la%20supervivencia%20y%20mortalidad.pdf)

*Data Science*. (s.f.). Obtenido de <https://www.datascience-pm.com/tdsp/>

*Microsoft Build*. (17 de 11 de 2020). Obtenido de [https://docs.microsoft.com/en-us/azure/machine-  
learning/team-data-science-process/overview](https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview)

Rueda, J. F. (s.f.). *Health Data Miner*. Obtenido de Data Mining : [thdataminer.com/healdata-  
mining/crisp-dm-una-metodologia-para-mineria-de-datos-en-salud/](https://thdataminer.com/healdata-mining/crisp-dm-una-metodologia-para-mineria-de-datos-en-salud/)