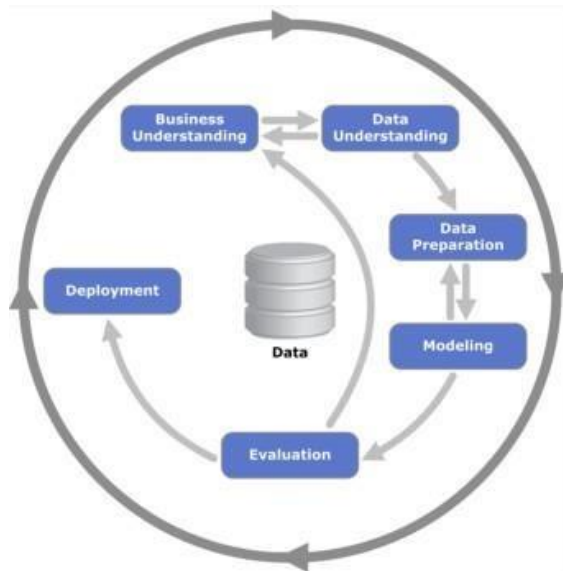


1 Taller02: Modelo CRISP-DM

1.1 Modelo CRISP-DM



El modelo CRISP-DM (*Cross Industry Standard Process for Data Mining*) es un enfoque estándar utilizado en la minería de datos y el análisis de datos. Proporciona una metodología estructurada para guiar el proceso de descubrimiento de conocimiento a partir de datos.

CRISP-DM consta de seis fases principales:

- (1) Comprensión del negocio: Se busca comprender los objetivos y requisitos del negocio, así como determinar qué preguntas deben responderse mediante el análisis de datos.
- (2) Comprensión de los datos: Se recopila y explora el conjunto de datos disponible, evaluando su calidad, familiarizándose con su contenido y buscando patrones o problemas potenciales.
- (3) Preparación de los datos: Se llevan a cabo tareas de limpieza, transformación y selección de los datos para que sean adecuados para el análisis posterior. Esto implica la eliminación de valores atípicos, la normalización de variables y la creación de nuevas características si es necesario.
- (4) Modelado: Se seleccionan y aplican técnicas de modelado apropiadas para descubrir patrones y relaciones en los datos. Esto puede incluir el uso de algoritmos de aprendizaje automático, estadísticas u otras técnicas analíticas.
- (5) Evaluación: Se evalúan los modelos generados en términos de su calidad y relevancia para los objetivos del negocio. Se pueden realizar ajustes o mejoras en los modelos según sea necesario.
- (6) Despliegue: Los resultados del análisis se presentan a las partes interesadas y se implementan en la práctica empresarial. Esto puede implicar la creación de informes, la integración de modelos en sistemas existentes o la toma de decisiones basadas en los resultados obtenidos.

El modelo CRISP-DM proporciona un marco flexible que puede adaptarse a diferentes proyectos y contextos. Permite un enfoque sistemático para abordar problemas complejos de análisis de datos y garantiza un proceso estructurado y repetible para obtener conocimientos valiosos a partir de los datos.

La analítica descriptiva es una rama de la analítica de datos que se centra en la descripción, resumen y visualización de datos para comprender mejor los patrones, tendencias y características presentes en ellos. Su objetivo principal es responder a preguntas como "¿Qué sucedió?" y "¿Qué está ocurriendo actualmente?".

1.2. Analítica Descriptiva

La analítica descriptiva utiliza técnicas y métodos estadísticos para analizar datos históricos y presentarlos de manera comprensible. Estas técnicas pueden incluir medidas de resumen, como promedios, medianas, desviaciones estándar y percentiles, así como gráficos y visualizaciones, como histogramas, gráficos de dispersión y gráficos de líneas.

Al aplicar la analítica descriptiva, se pueden descubrir patrones y tendencias en los datos, identificar valores atípicos, comprender la distribución de los datos y resumir la información de manera concisa. Esto proporciona una base sólida para obtener información valiosa y tomar decisiones fundamentadas.

La analítica descriptiva se utiliza en una amplia gama de industrias y aplicaciones, desde el análisis de datos de ventas y marketing hasta el monitoreo de operaciones y la gestión de recursos humanos. Es una etapa fundamental en el proceso de analítica de datos y sienta las bases para etapas más avanzadas, como la analítica predictiva y la analítica prescriptiva.

Un proyecto de analítica descriptiva típicamente sigue varias fases clave para alcanzar sus objetivos. Estas fases pueden variar según el contexto y la metodología utilizada, pero a continuación se presentan las etapas generales de un proyecto de analítica descriptiva:

- (1) Definición de objetivos: En esta fase inicial, se identifican y establecen los objetivos del proyecto de analítica descriptiva. Esto implica comprender las necesidades y preguntas del negocio que se buscan responder a través del análisis de datos. Es importante definir claramente los objetivos para orientar todo el proceso.
- (2) Recopilación y preparación de datos: En esta fase, se recopilan los datos necesarios para el análisis. Esto puede implicar extraer datos de diversas fuentes, como bases de datos, archivos CSV o sistemas en tiempo real. También se realiza la limpieza y preparación de los datos, incluyendo la eliminación de valores faltantes o erróneos, la estandarización de formatos y la transformación de datos si es necesario.
- (3) Análisis exploratorio de datos: En esta etapa, se lleva a cabo un análisis exploratorio de los datos para comprender su estructura, características y relaciones. Se utilizan técnicas estadísticas y visualizaciones para identificar patrones, tendencias y valores atípicos en los datos. Esto ayuda a obtener una visión inicial de los datos y a generar hipótesis que se abordarán en las siguientes etapas.
- (4) Análisis descriptivo: En esta fase, se aplican técnicas de analítica descriptiva para describir y resumir los datos de manera sistemática. Se utilizan medidas estadísticas, gráficos y visualizaciones para resumir la distribución de los datos, calcular medidas de tendencia central y dispersión, y proporcionar una comprensión más profunda de los datos en relación con los objetivos establecidos.
- (5) Interpretación de resultados: En esta etapa, se interpretan los resultados obtenidos en el análisis descriptivo y se vinculan con los objetivos del proyecto. Se buscan respuestas a las preguntas iniciales y se extraen conclusiones significativas a partir de los hallazgos. Se pueden generar informes y visualizaciones para comunicar los resultados a las partes interesadas.
- (6) Presentación y comunicación: En esta fase final, se presenta y comunica el análisis descriptivo a las partes interesadas. Esto implica la preparación de informes, presentaciones y visualizaciones que resuman los hallazgos clave de manera clara y comprensible. Se facilita la discusión y se brinda oportunidad para aclarar dudas y obtener retroalimentación.

Estas fases no son necesariamente lineales y pueden requerir iteraciones y ajustes a lo largo del proyecto. Además, la analítica descriptiva sienta las bases para etapas más avanzadas, como la analítica predictiva y la analítica prescriptiva, que buscan utilizar los conocimientos obtenidos para predecir y prescribir acciones futuras.

Bibliografía: Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of data warehousing*, 5(4), 13-22

1.3. Definición de un proyecto de Analítica Descriptiva:

(1) Problema a resolver:

Analizar y tratar de predecir el rendimiento de diferentes sectores y productos de exportación ecuatorianos en el mercado de Estados Unidos, considerando para esto las variables económicas, políticas y comerciales que podrían afectar. Utilizar esto para desarrollar un modelo predictivo de Machine Learning que pueda mínimamente predecir o graficar la demanda futura de productos específicos exportados a Estados Unidos e identificar los factores clave que influyen en las exportaciones.

(2) Objetivos:

-Analizar el rendimiento histórico de los sectores y productos de exportación ecuatorianos hacia Estados Unidos: Mediante el análisis descriptivo de los datos históricos, se busca comprender el rendimiento pasado de los sectores y productos de exportación ecuatorianos en el mercado estadounidense. Esto implica examinar las tendencias, patrones y variaciones en los volúmenes de exportación, ingresos generados y otras métricas relevantes.

-Identificar los factores económicos, políticos y comerciales que influyen en las exportaciones: Mediante el análisis de las variables económicas, políticas y comerciales disponibles o tratar de inferir, se pretende identificar los factores clave que tienen un impacto significativo en las exportaciones ecuatorianas hacia Estados Unidos. Esto implica determinar cómo las tasas de cambio, las políticas comerciales, los acuerdos internacionales y otros factores pueden influir en el rendimiento de los sectores y productos de exportación. Usando Data Mining.

-Desarrollar un modelo predictivo de Machine Learning para predecir la demanda futura de productos exportados a Estados Unidos: El objetivo es construir un modelo predictivo utilizando técnicas de Machine Learning y Data Mining que pueda estimar la demanda futura de productos específicos exportados desde Ecuador hacia Estados Unidos. Este modelo se basará en datos históricos y variables relevantes para realizar pronósticos y ayudar en la toma de decisiones.

-Evaluar el rendimiento del modelo predictivo y su precisión: Se llevará a cabo una evaluación del modelo predictivo utilizando métricas apropiadas para determinar su capacidad de hacer predicciones precisas. Esto permitirá evaluar la efectividad del modelo y su utilidad para la toma de decisiones relacionadas con las exportaciones.

-Identificar estrategias y oportunidades para mejorar la competitividad en el mercado de Estados Unidos: Basándose en los resultados del análisis descriptivo y el modelo predictivo, se buscará identificar estrategias y oportunidades para mejorar la competitividad de los sectores y productos de exportación ecuatorianos en el mercado de Estados Unidos. Esto podría implicar recomendaciones sobre productos con alto potencial de crecimiento, estrategias comerciales específicas o medidas para enfrentar desafíos identificados.

-Tratar de cumplir los máximo objetivos posibles.

(3) Datos:

Aquí a continuación daremos pequeños ejemplos de lo que se van a usar para el proyecto ya que se obtuvo del gobierno.

a. Fuentes de datos:

- Nombre de la fuente de datos:
MPCEIP_DEECO_EXPORTACIONES_EEUU_2023_02.csv
- Origen de la fuente de datos: Ministerio de Producción, Comercio Exterior, Inversiones y Pesca (MPCEIP).
- Descripción de la fuente de datos: La fuente de datos proporciona información

histórica de las exportaciones desde Ecuador hacia Estados Unidos desde el año 2001. Contiene datos sobre los diferentes sectores y productos de exportación, así como variables económicas, políticas y comerciales relacionadas.

b. Tipos de datos:

- Datos numéricos o cuantitativos: Toneladas exportadas, códigos export FOB (miles), valores económicos, tasas de cambio, entre otros.
- Datos categóricos: País destino, capítulo, producto principal nivel 4, producto tradicional o no tradicional, producto petrolero o no petrolero, entre otros.
- Datos de tiempo/temporales: Año, mes.
- Datos espaciales: Estos datos representan información geográfica, Ecuador & Estados Unidos. No obstante, estos ya entrarían en categóricos.

c. Procesamiento:

- i. Limpieza de datos: Realizar la limpieza de datos para manejar valores nulos, valores atípicos y posibles inconsistencias en los datos. Esto puede incluir técnicas como el reemplazo de valores faltantes, la eliminación de registros duplicados y la corrección de errores.
- ii. Integración de datos: En caso de tener más de una fuente de datos, se debe realizar la integración de las diferentes fuentes para consolidar la información en un único conjunto de datos.
- iii. Transformación de datos: Realizar transformaciones en los datos según sea necesario, como la normalización de variables numéricas, codificación de variables categóricas y agregación de datos a diferentes niveles de granularidad.
- iv. Análisis exploratorio de datos: Realizar análisis estadísticos y visualizaciones para comprender la distribución de los datos, identificar relaciones entre variables y explorar patrones y tendencias en los datos.
- v. Selección de características: Identificar las variables más relevantes para el análisis y modelado posterior, utilizando técnicas como la correlación o la importancia de las variables en modelos de Machine Learning.
- vi. Análisis de datos: Utilizar técnicas de análisis descriptivo para obtener insights sobre los diferentes sectores y productos de exportación, explorar relaciones entre variables y examinar el rendimiento histórico.
- vii. Preparación de datos para el modelado: Preparar los datos para el desarrollo del modelo predictivo, dividiendo el conjunto de datos en conjuntos de entrenamiento y prueba, y realizando cualquier transformación adicional requerida por el algoritmo de Machine Learning seleccionado.