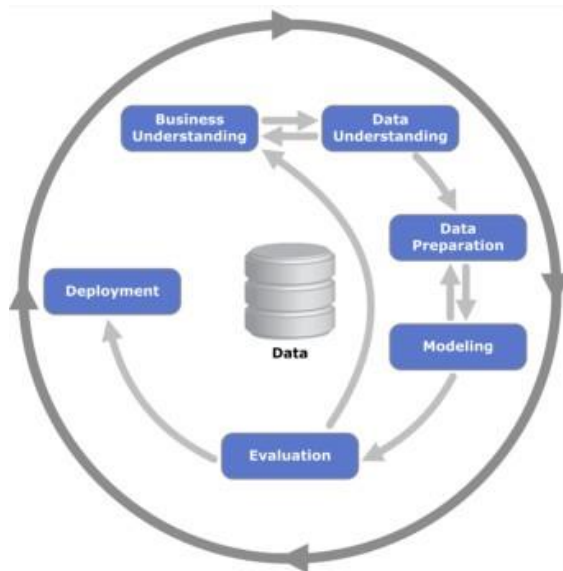


## 1 Taller02: Modelo CRISP-DM

### 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:

Explicación del “Valle Inquietante” y del Neoludismo con su convergencia y concurrencia en la sociedad de las tecnologías emergentes con los posibles cambios de paradigma.

(2) Objetivos:

- Explicar el Valle Inquietante y el Neoludismo.
- Ver causas, hipótesis, posibles consecuencias.
- Establecer y analizar conexiones y semejanzas.
- Proponer posibles “soluciones” a ambos después de una síntesis.
- Entender los paradigmas que representan con la Inteligencia Artificial y otras tecnologías.

(3) Datos:

Aquí a continuación daremos pequeños ejemplos de lo que se van a usar para el proyecto ya que se irán incluyendo después y con su respectiva bibliografía.

a. Fuentes de datos:

- Libros: "El valle inquietante" de Mori Hiroshi será usado al igual que otros.
- Artículos científicos: "The Uncanny Valley in Virtual Reality: Examining the Perceived Realism of Humans, Robots, and Zombies" de Nicole Lazzeri y colaboradores.
- Novelas: como Asimov, Clarke y/o Ellison.
- Internet: impacto y reacciones sobre las tecnologías emergentes.
- Blogs
- Películas: sobre efectos del valle y del neoludismo.
- Gráficos: La curva describe cómo, a medida que la apariencia se vuelve más realista y humana. Las infografías son representaciones visuales que combinan imágenes, gráficos y texto para presentar información de manera concisa y atractiva.

b. Tipos de datos:

- Datos numéricos o cuantitativos: Estos datos representan cantidades o medidas numéricas. Pueden ser datos continuos o datos discretos.
- Datos categóricos: Estos datos representan categorías o grupos distintos. Pueden ser datos nominales, donde no hay un orden o jerarquía específica, como el género o el estado civil, o datos ordinales, donde hay un orden o jerarquía, como el nivel educativo o la clasificación de satisfacción (baja, media, alta). Los datos categóricos se pueden utilizar para realizar conteos y análisis de frecuencias.
- Datos de tiempo: Estos datos representan información temporal. Pueden ser incluso fechas.
- Datos espaciales: Estos datos representan información geográfica.

c. Procesamiento:

- i. Extracción de datos: Consiste en obtener los datos de diversas fuentes, como bases de datos, archivos, sistemas en línea, sensores, entre otros. La extracción puede implicar la recopilación de datos en tiempo real o la obtención de datos históricos.
- ii. Limpieza de datos: Los datos recopilados pueden contener errores, valores faltantes, duplicados u otros problemas. La limpieza de datos implica detectar y corregir estos errores para asegurar que los datos sean precisos y confiables.

- iii. Transformación de datos: En esta etapa, los datos se pueden reestructurar o modificar para adaptarlos a los requisitos específicos del análisis. Esto puede incluir la normalización de datos, la agregación de información, la creación de variables derivadas, la codificación de variables categóricas, entre otros procesos.
- iv. Integración de datos: En ocasiones, los datos provienen de múltiples fuentes y es necesario combinarlos para obtener una visión completa. La integración de datos implica fusionar y unificar conjuntos de datos heterogéneos en un formato coherente.
- v. Análisis y modelado: Una vez que los datos están procesados y preparados, se pueden aplicar diversas técnicas de análisis, como estadísticas descriptivas, minería de datos, aprendizaje automático (machine learning), entre otros, para descubrir patrones, tendencias y relaciones dentro de los datos.
- vi. Visualización de datos: La presentación visual de los datos puede ser una parte importante del procesamiento, ya que las visualizaciones pueden ayudar a comprender mejor los patrones y tendencias presentes en los datos. Las gráficas, diagramas y mapas son ejemplos comunes de herramientas de visualización utilizadas en el procesamiento de datos.