TÉCNICAS DE MÍNERIA DE DATOS

REGLAS DE ASOCIACIÓN

Las reglas de asociación sirven para describir una relación de asociación entre los elementos de un conjunto de datos relevantes

Principio de Apiorí: si un conjunto de elementos es frecuente, entonces todos sus subconjuntos también deben ser frecuentes.

El soporte de un conjunto de elementos nunca excede el soporte de sus subconjuntos. Esto se conoce como la propiedad antimonótona de soporte



135 125 200 75 10 60 79 80 W 100 100 110 110 110

Un valor atípico es una observación que es numéricamente distante del resto de los datos.

Técnicas para la detección de valores atípicos • Prueba de Grubbs • Prueba de Dixon • Prueba de Tukey • Análisis de valores • Regresión simple

Una vez detectados los valores atípicos, se pueden eliminar o sustituir si se corrobora que los datos atípicos se deben a un error de captura o en la medición de la variable.

REGRESIÓN

Una regresión es un modelo para determinar el grado de dependencia entre una o más variables. Tipos de regresión: Lineal y lineal multiple.

tiene como objetivo analizar datos de un conjunto y en base a eso predecir lo que ocurrirá en un futuro.

El análisis de la regresión nos permite analizar la relación entre dos o más variables e identificar cuales son los que tienen mayor impacto, además de ayudarnos a tomar mejores decisiones en base a lo obtenido.

CLAUSTERING

El proceso consiste en la división de los datos en grupos de objetos similares. Las técnicas son las que utilizando algoritmos se encargar de agrupar objetos.

Un cluster es una colección de objetos de datos. Similares entre si dentro del mismo grupo.

Análisis de clauster: dado un conjunto de puntos de datos tratar de entender su estructura. De acuerdo con las características presentadas en los datos se encuentran las similitudes de los mismos

PREDICCIÓN

Técnica utilizada para procetar los tipos de datos que se verán en el futuro o realizar una preddicion para el resultado de un evento.

Aplicaciones • Revisar los historiales crediticios • Predecir el tiempo en una entidad • Precio de venta de alguna propiedad • Puntuación de algún partido de futbol

Técnicas • Modelos estadísticos • Estadísticas no lineales • Redes neuronales, RBF

PATRONES

La minería de datos secuenciales, se refere a la extracción de patrones frecuentes relacionados con el tiempo u otro tipo de secuencia. El objetivo es describir de forma concisa relacione temporales que existen entre los valores de atributos del conjunto.

Una subsecuencia también es una secuencia, pero esta se encuentra dentro de otra, con la condición de cumplir ciertas normas.

Análisis de secuencias 1. Bases de datos 2. Secuencia 3. Elemento(transacción) 4. Evento (ítem)

VISUALIZACION DE DATOS

La visualización de datos es la presentación de información en formato gráfico, esto nos ayuda a comprender y ver tendencias valores atípicos y/o patrones en los datos.

Tipos: • Gráficos: gráficos circulare, líneas, columnas, barras aisladas o agrupadas, burbujas, áreas, Diagramas de Dispersión etc. • Mapas • Infografías: Colección de imágenes gráficos y texto simple que resume un tema para que se pueda entender con facilidad. • Cuadros de Mando: Herramienta que nos permite conocer el estado de los indicadores de alguno negocio.

CLASIFICACIÓN

Una clasificación es aquella técnica de la minería de datos encargada de ordenar por clases tomando en cuenta diversas características de sus elementos.

Características: Eficiencia, Robustez, Escalabilidad, Interpretabilidad, Precisión en la predicción.

Métodos en la clasificación de datos: • Análisis discriminante: Se usa para encontrar una combinación lineal de rasgos que separan clases de objetos. • Árboles de decisión: A través de una representación esquemática facilita la toma de decisiones • Reglas de clasificación • Buscan términos no clasificados de forma periódica.

• Redes neuronales artificiales: Modelo de unidades conectadas para transmitir señales.