



30 DE SEPTIEMBRE DE 2020

RESUMEN PRESENTACIONES

MINERÍA DE DATOS

IGNACIO CONTRERAS RUIZ

1887848




Tabla de contenido

| | |
|-----------------------------|---|
| Resumen de Técnicas | 2 |
| DESCRIPTIVAS..... | 2 |
| Clustering | 2 |
| Regla de Asociación..... | 2 |
| Detección de outliers | 3 |
| Visualización..... | 3 |
| PREDICTIVAS..... | 4 |
| Regresión..... | 4 |
| Clasificación..... | 4 |
| Patrones Secuenciales..... | 5 |
| Predicción..... | 5 |

Resumen de Técnicas

DESCRIPTIVAS

Clustering

El clustering consiste en crear grupos dentro de la base de datos, los cuales buscan contener elementos con características similares, o bien con una cercanía relativa dentro del área de trabajo, se utilizan diferentes métodos para poder identificar el número adecuado de clusters y/o centroides, como puede ser el método elbow, dentro del mismo procedimiento se utiliza un proceso muy conocido en minería de datos, el de K-medias,

Este tipo de técnica se puede emplear cuando se hace una investigación de mercado y lo que queremos hacer es una segmentación de clientes.

Observando las métricas de los clientes históricos podemos definir los diferentes grupos que tenemos y así poder enfocarnos en áreas de mejora que se encuentren en el proceso.

Regla de Asociación

Se encarga de buscar y analizar relaciones relevantes entre los datos, para así poder definir y predecir lo que puede seguir a continuación después de que un elemento aleatorio es añadido a la base de datos.

Podemos encontrar este tipo de técnica en nuestro propio celular o en aplicaciones actualmente utilizadas como es Netflix, Youtube, Spotify.

Se genera una asociación dentro de nuestras selecciones al utilizar dichas plataformas y así se especula lo que puede ser reproducido más adelante, esto con ayuda de comportamientos anteriores de otros usuarios.

Detección de outliers

Consiste en analizar aquellos factores que sobresalen dentro de la información, todo dato que llegue a ser muy lejano a una media por ejemplo se convierte en un outlier, estos generan un sesgo dentro de la información.

Estos datos generan conflictos durante el proceso de análisis, lo que resulta en información poco confiable e inconsistente. Sin embargo, si bien los datos anómalos son irrelevantes para encontrar patrones en la cotidianidad de los datos, son útiles como objeto de estudio en casos en donde, mediante estos es posible identificar mediante un proceso no controlado, problemas tales como fraudes financieros.

Poder detectarlos es conveniente para poder identificar algún error dentro de la información que merezca ser analizado con mayor detenimiento.

Visualización

En el caso de la visualización se pueden involucrar cualquier de las técnicas que se mencionen o sean mencionadas en este texto, ya que es un apoyo casi indispensable para el análisis completo de la información y que permite crear una conciencia mayor de la ubicación o comportamiento de los datos.

Así como podemos generar elementos como cajas de bigotes, gráficas de pie, histogramas, así como diferentes gráficas de dispersión ayudan a analizar la información desde diferentes planos, como puede ser en dos dimensiones, así como en tres o más si fuera necesario.

El poder entender la información es parte esencial a la hora de crear soluciones o administrar datos para crear una idea general del comportamiento de la información y así adaptar los procesos necesarios según lo que se observe.

PREDICTIVAS

Regresión

Para esta técnica comenzamos a observar el análisis que propicia un modelo predictivo capaz de adecuar la información y brindarnos un valor lo suficientemente cercano a la realidad para poder preparar las acciones que se deberán tomar en el momento que el dato a predecir finalmente llegue.

La regresión puede ser desde sencilla con pocas variables hasta una regresión múltiple, a su vez, existen regresiones lineales, cúbicas, logarítmicas y más, las cuales en su mayoría con diversas adecuaciones se logra asemejar el modelo a un lineal para poder así trabajar de forma más sencilla la información.

Los datos pueden ser de diferentes índoles y según sea el caso lo único que se requiere es utilizar los datos según la variable a predecir, así como la que ayuda a crear el modelo o predictor.

Clasificación

Así como la visualización, esta técnica se puede observar en gran medida en todas las otras, esto debido que para la correcta utilización de ellas se necesita un orden dentro de la base datos, esto se logra con una correcta clasificación de la información.

Permite poder crear filtros que según las características nos ayuden a definir mejor las variables en las que realmente se debe concentrar el análisis ahorrando así gran cantidad de tiempo.

Genera soluciones productivas en comparación con otras técnicas de análisis y a también facilita las predicciones automáticas si se orden de manera adecuada, como puede ser ascendente o descendente según sea necesario.

Patrones Secuenciales

Este tipo de estrategia puede no ser tan utilizada ya que dentro de una misma secuencia se busca una nueva secuencia, se asemeja a la anidación de una misma técnica dentro de otra, busca subsecuencias que sean relevantes y que puedan generar un cambio al momento de crear un modelo predictor.

Posiblemente y desde cierta perspectiva se asemeja al comportamiento de la técnica de detección de outliers, aunque en este caso no es un dato anómalo el que se busca si no al contrario, un dato o conjunto de datos que sobresalga por características muy similares entre sí.

Predicción

Relativamente esta técnica se encuentra dentro de todas las previamente mencionadas en esta sección, ya que el objetivo de todas ellas es conocer lo que ocurrirá más adelante con una observación siguiente, es de gran utilidad para el área financiera, demográfica y social en general.

Actualmente no es posible predecir con una exactitud completa un evento futuro, sin embargo con el principio coloquialmente conocido de “Quien no conoce su historia, estará obligado a repetirla” podemos dejar en claro que para poder tener aunque sea una idea de lo que a continuación puede ocurrir según el supuesto de un evento que acaba de ocurrir será necesario volver atrás para evaluar el comportamiento y las posibilidades que se tienen ante los eventos futuros.

Es indispensable la correcta evaluación del riesgo para esto, en situaciones que puedan ser de gran peligro, como ejemplo inmediato, la pandemia que vivimos, al no ser la primera que conoció el ser humano se puede crear un modelo de lo que sigue después de ella.