



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

UNIVERSIDAD AUTONOMA DE NUEVO LEON

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



FACULTAD DE CIENCIAS FISICO MATEMATICAS

MATERIA: Minería de Datos

A CARGO DE: Mayra Cristina Berrones Reyes

TEMA: Ejercicios de bases de datos

POR: Mariann Adaliz Avila Rios

Siguiendo el ejemplo dado en clase, el alumno utilizará las ligas proporcionadas para analizar en equipo las bases de datos:

a)Google Play Store <https://www.kaggle.com/lava18/google-playstore-apps>

Objetivo: Crecer la descarga de apps de pago en la playstore

Problema planteado: Actualmente más del 95% de las apps descargadas dentro de la playstore son apps gratuitas. Existen funciones que permiten omitir las apps de pago y cuando se hacen recomendaciones en la store muchas son apps gratuitas.

Solución: Dentro de la playstore existen recomendaciones para que sigamos, los usuarios, descargando aplicaciones, estas recomendaciones son en base a popularidad y no en base a generar mas ingresos . Para las recomendaciones que hace play store sobre las apps podemos limitar a que en su mayoría sean apps de paga utilizando árboles de decisión según las aplicaciones que el usuario ha instalado en su dispositivo entonces recomendamos apps de la misma gama ya sea por categoría o uso pero apps de paga. Podemos incluir además que de las recomendadas, las que sean gratis incluyan compras dentro del app. Esto ayudara a incrementar los ingresos de la playstore más allá de los ingresos por publicidad

b) Coronavirus corona-virus-2019-dataset

Objetivo: Determinar si las medidas por país y por provincia son efectivas para combatir el covid, además generar bases para determinar la fuerza de recuperados y muertos por país

Problema planteado: Revisando los datos se observan diferencias entre países y en muchos casos el nivel económico del país no interfiere, véase EUA como claro ejemplo de un país con posición alta económicamente. Por lo que el problema recae en si las medidas de salud son efectivas o no y cuales se debe replicar.

Solución Se propone una regla de asociación por país y por provincia asociando medias como uso de cubrebocas y que tanta difusión y alcance se le da a la medida con el numero de infectados. Después otra asociación de poder económico vs el numero de muertos para identificar si la diferencia de muertos y recuperaciones cae en el poder económico del país y la fuerza de su sistema de salud o bien pudiera estar relacionado con temas como zona geográfica, tomando en cuenta el clima y las sospechas de que el COVID19 pudiera no sobrevivir al calor.

c)Críticas de vinos <https://www.kaggle.com/zynicide/wine-reviews>

Objetivo: Crear la compra de ciertos vinos por usuario, aumentando en variedad y agregar compras en conjunto al comprador de vinos (tablas).

Problema planteado: La base de datos obtiene clasificaciones por twitter y entrega el handle de usuario además clasifica los vinos con “variety, location, winery, price, and description”. Como se puede entregar al usuario mejor recomendaciones de vino de acuerdo a sus compras y a su descripción

Solución: Creo que deberíamos asociar el precio promedio de vinos que consumen usuarios de la variedad de vino y su descripción para poder ofrecer otras entradas de vinos del mismo promedio y descripción, sin embargo de otra variedad, es decir el usuario compra vinos tintos pero afrutados, se le pudieran recomendar vinos blancos, que igual son d sabor ligero. Con el objetivo de crecer la compra. Además se puede hacer con un árbol de decisiones para ofrecer publicidad de complementos para el vino, es decir quesos o tablas y carnes frías que vayan con la variedad determinada de vino tomando en cuenta estos datos para alguna tienda que venda vinos y otros artículos y pudiera aumentar ventas en paquetes.

d)Clasificación de plantas <https://www.kaggle.com/uciml/iris>

Objetivo: Encontrar las plantas mas grandes o con mejores características para su venta o investigacion

Problema planteado Las plantas de la familia Iris crecen de diferente forma dependiendo de su especifico, además varían en tamaño y alto. La base de datos falta información para este problema pero si se diera ubicación de la planta pudiéramos resolverlo

Solución Tomando en cuenta que estas plantas pueden ser utilizadas en decoración o investigaciones se buscan las plantas más altas y mejores flores. Considero que debemos seleccionar el grupo deseado de planta y para esto la técnica de clustering para agrupar el tipo de planta. Las plantas se pudieran agrupar por familia coma por plantas con pétalos más grandes o por plantas más altas y así poder encontrar las variables que determinan su altura o crecimiento de pétalos.Además tomando en cuenta la información ya otorgada sobre el tipo de pétalos que existen y qué 1 de ellos se excluye de los otros y los otros dos no entre si, se pudiera encontrar que relación tiene esto con el crecimiento de los pétalos.

e)Shows de Netflix <https://www.kaggle.com/shivamb/netflix-shows>

Objetivo Identificar si Netflix debe invertir en producciones propias y compra de series o si su contenido actual satisface al usuario

Problema planteado De los insights ya otorgados observamos un crecimiento exponencial en el consumo de series de Tv Vs películas. Se debe entender al usuario y la demanda para saber si Netflix deberá seguir creciendo esta industria o invertir en películas.

Solución Para entender al usuario y su demanda propongo crear perfiles de usuario y por ubicación ver cuantos de estos perfiles existen. Usando la técnica de clustering formamos estos grupos en base al tiempo que gastan viendo series o películas y si visitan series viejas o buscan nuevas series. así entonces si encontramos que en un determinado país los usuarios optan por series clásicas se puede optar por seguir comprando series de años anteriores que pudiera cambiar de país a país y entonces en otro país el usuario puede que esté tratando nuevos productos cada mes por lo que debiéramos decidir invertir en la creación de nuevas series.