

Minería de Datos

Ejercicio Bases de Datos

Omar Alejandro Garza Espinosa 1931548

Profesor: Mayra Cristina Berrones Reyes
Grupo: #03

Viernes 16 de octubre del 2020

Ciudad Universitaria, San Nicolás de los Garza, N.L.

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



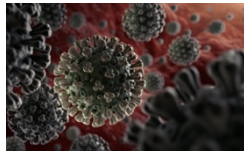
Google Play Store Apps

Objetivo: Facilitar la búsqueda de aplicaciones a través de redes para lograr una mejor clasificación de las aplicaciones según su género (familiar, educativo, etcétera) y brindar un mejor servicio.

Problema Planteado: Los clientes pierden mucho tiempo al buscar aplicaciones por la gran cantidad que hay de ellas en Google Play Store. Una buena clasificación garantizaría una rápida búsqueda de alguna aplicación para el gusto del cliente según al genero que desee.

Solución: Hacer una mejor clasificación de género de las aplicaciones en base a redes para facilitar y acelerar la búsqueda con base a su rating.

b) Coronavirus (<https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>)



Novel Corona Virus 2019 Dataset

Objetivo: Proyectar la letalidad del Virus Covid19 en la población de China a través de una regresión involucrando tres variables (edad, letalidad y ciudad).

Problema Planteado: No existen registros de la letalidad del virus Covid19 en China por lo que se dificulta hacer una proyección de sus efectos en la población según su edad. También es importante localizar las ciudades donde existe más infectados ya que el grado de contagios es muy alto.

Solución: Correlacionar las variables edad, letalidad y locación para determinar si existe alguna regresión en la que sea significativa la edad de la persona contagiada con la incidencia de muerte así como la ciudad donde se contagió.



Wines Reviews

Objetivo: Diseñar una aplicación donde el usuario tenga un catálogo de los mejores vinos por viñedo y por país.

Problema Planteado: Un gran número de personas que no son expertas en el maridaje de vinos, desconocen la procedencia de un vino así como la uva que se utilizó por lo que no reconocen un buen vino.

Solución: Clasificar las regiones que producen los mejores vinos por país para que así el cliente le sea más sencillo buscar algún vino en especial con la certeza de que este es un producto bueno y de calidad.



Iris Species

Objetivo: Diseñar un modelo matemático que permita identificar los diferentes tipos de Plantas Iris.

Problema Planteado: Para los científicos, biólogos y taxónomos en ocasiones es complicado identificar el tipo de Planta Iris a simple vista ya que son muy parecidas. La importancia de esta identificación reside en que cada tipo de Iris posee distintos tipos de sustancias que pueden usarse en áreas como medicina y cosméticos.

Solución: Encontrar esos criterios importantes y particulares de la Planta Iris que logren diferenciar cada tipo de planta a través de un modelo matemático que integre todas las variables involucradas (largo y ancho de tallo así como el largo y ancho del pétalo).



Netflix Movies and TV Shows

Objetivo: Clasificar las películas y series contenidas en Netflix tomando en cuenta el año que se subieron a la plataforma y el rating que obtuvieron en ese mismo año para optimizar la búsqueda.

Problema Planteado: No se tiene una clasificación en Netflix de la mejores películas y series por lo que se plantea la necesidad de saber que películas y series agregadas en un determinado año fueron las más vistas durante ese periodo para optimizar la elección del cliente.

Solución: Agregar un apartado en la plataforma de Netflix que permita hacer una nueva búsqueda de las mejores películas y series por año en el que se subieron.