



Universidad Autónoma de Nuevo León
Licenciatura en Actuaría

Minería de datos

Bases de datos

Suarez Martinez Tadeo Alejandro
7to Semestre
1806069

Mayra Cristina Berones Reyes

14 de octubre del 2020

Google Play Store

Bases de datos

Google Play Store Apps

Objetivo

Que los usuarios de la Play Store tengan recomendaciones en función a sus gustos y/o preferencias.

Problema Planteado

Muchas veces los usuarios tienen recomendaciones de aplicaciones que no son de su agrado o interés, perdiendo la oportunidad de explorar las aplicaciones de acuerdo con sus gustos.

Solución

Con algún algoritmo de las reglas de asociación y patrones secuenciales descubrir que descargas suelen hacerse en conjunto acorde a la edad, si la app es de paga o gratuita y la categoría del juego para posteriormente recomendar aplicaciones similares.

Coronavirus

Bases de datos

Novel Corona Virus 2019 Dataset

Objetivo

Encontrar algo fuera de lo ordinario que tengan en común los pacientes que no se recuperaron de la enfermedad para ver si estos tienen menos posibilidades de sobrevivir.

Problema Planteado

Mucha gente esta muriendo por esta enfermedad que por desgracia es desconocida, por lo que queremos encontrar si los pacientes que no se recuperan tienen características fuera de lo común para tener mayor precaución con estas personas.

Solución

Con algún algoritmo de detección de outliers descubrir que personas están más propensas a morir de acuerdo con sus diferencias con los otros pacientes.

Critica de Vinos

Bases de datos

Wine Reviews

Objetivo

Detectar qué relación tiene la calificación en función del precio y el lugar para saber en qué sitios puede afectar la calificación si aumentamos el precio.

Problema Planteado

Subir el precio de ciertos vinos puede generar exigencias más altas y afectar en las criticas, sin embargo, no sabemos en qué países o estados son más exigentes con el precio y calidad del vino

Solución

Comprobar si tienen alguna relación lineal con variables independientes el precio y el lugar(Entre 2 lugares), y como variable dependiente la calificación, en caso de tener esta relación hacer predicciones con algoritmos de regresión de cómo afectaría el cambio de lugares y de precio a la calificación de los vinos.

Clasificación de plantas

Bases de datos

Iris Species

Objetivo

Detectar cual es el tipo de planta teniendo en cuenta las medidas de esta como longitud del sépalo, ancho del sépalo, longitud del pétalo y ancho del pétalo.

Problema Planteado

Reducir el tiempo de saber si la planta es setosa, versicolor o virginica o lograr saber esto sin necesidad de ser experto en el tema si no solo con las medidas.

Solución

En este problema también se me ocurre una regresión lineal con variables independientes las medidas y la variable dependiente el tipo de planta, para esto igual que en el punto anterior tendríamos que comprobar los supuestos de la regresión y posteriormente hacer las predicciones de acuerdo a las medidas.

Shows de Netflix**Bases de datos**

Netflix Movies and TV shows

Objetivo

Detectar que cast, géneros o directores generan mas rating en cierto país.

Problema Planteado

Cada país tiene gustos y preferencias diferentes, tanto en directores, géneros o repartos, por eso es importante decidir qué películas pondrán en el catálogo de cierto lugar, ya que no funcionara de la misma manera en todos los países.

Solución

Construir unas representaciones graficas de los elementos más visualizados con algoritmos de visualización en cada país, ya sea directores, géneros y actores para con esto guiarnos sobre los gustos que tiene cada región y hacer las recomendaciones más optimas y adecuadas en el sector.