

Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas

Minería de Datos

Bases de datos

Nombre: Valeria Nohemí Navarro Cabello

Matricula: 1820160

14 de Octubre de 2020

Aplicaciones de Google Play Store

Objetivo: Hacer recomendaciones en el inicio de la página que estén más relacionadas con tus últimas búsquedas y te muestre aplicaciones similares con la misma categoría y que se ordenen de manera que te muestren primero las que están mejor calificadas y con mantenimiento y actualizadas recientemente.

Problema planteado: En el inicio primero muestran todas las aplicaciones más populares y no las más relacionadas con tus búsquedas e intereses. En ocasiones al hacer una búsqueda aparecen aplicaciones que tal vez en el pasado fueron buenas pero no se les ha dado el mantenimiento necesario ni se han actualizado por lo que van quedando obsoletas.

Solución: Usar la base de datos de "Category", "Rating", "Reviews" e "Installs" y con análisis estadístico y haciendo uso de técnicas como la regla de asociación para encontrar patrones dentro de nuestras búsquedas y que en el inicio te muestre las aplicaciones más relacionadas y con árboles de decisiones tomando en cuenta diferentes condiciones para ordenarlas de tal manera que probablemente no van a desaparecer las aplicaciones peor calificadas o que han quedado obsoletas pero no aparecerán entre las primeras opciones.

Coronavirus (Nuevo conjunto de datos del virus Corona 2019)

Objetivo: Hacer un análisis de cómo influye el nivel socioeconómico de un país en la propagación de una pandemia, en este caso del coronavirus y hacer conciencia de las medidas que se pueden tomar para pandemias o enfermedades similares a esta en el futuro.

Problema planteado: Con todo lo que sucedió acerca de la pandemia del Coronavirus hubo una elevada cantidad de contagios y a la vez de muertes en donde en muchos países se sufrió más por no contar con suficientes recursos y la infraestructura hospitalaria era limitada.

Solución: Usar la base de datos de "país" y "ubicación" para observar que países son los que tiene mayor cantidad de contagios y de muertes relacionándolo con el estado socioeconómico del país haciendo conciencia de cómo la economía influye directamente en las medidas y en las consecuencias que trae una pandemia haciendo estragos mayores en países menos desarrollados.

Critica de vinos

Objetivo: Que consideren al restaurante como uno de los mejores proveedores de vino invirtiendo en un buen viñedo y en donde el producto este basado en lo que el cliente busca frecuentemente en un vino y encontrar lo que nos va a caracterizar y diferenciar de la competencia según las regiones en donde se encuentre el restaurante.

Problema planteado: Considerando un dueño de un restaurante que quiere implementar la venta de los mejores vinos e invertir de manera segura.

Solución: Implementar la base de datos prestándole atención a las descripciones y la calificación haciendo uso de la técnica de Clasificación y arboles de decisión para llegar a lo que quiere y pide el cliente considerando sus gustos y para invertir en el viñedo adecuado utilizando los datos que nos proporciona la base de “designación” que nos ayudara a decidir de donde vamos a obtener las uvas para elaborar el vino.

Clasificación de plantas

Objetivo: Reducir el tiempo que se tarda un experto en identificar alguna clase de anomalía en las características de la planta denominada Iris y una vez encontradas poder hacer un estudio de investigación acerca de lo que causo esta clase de diferencia entre las demás plantas.

Problema planteado: Se requiere mucho tiempo en hacer este tipo de estudio de cada una de las plantas por clasificación.

Solución: Tomar en cuenta todas las columnas que nos muestra la base de datos y desarrollar una herramienta de máquina que nos ayude a detectar de una manera más rápida cualquier anomalía valiéndose de las características que cada tipo de la planta Iris debe tener y se puede usar alguna técnica como la de Outliers que nos ayudara a encontrar los datos atípicos que se encuentren fuera de un rango que se va a determinar en la longitud y ancho del sépalos y del pétalo dentro de nuestro conjunto de datos.

Películas y shows de Netflix

Objetivo: Prolongar el tiempo que el espectador está en la aplicación proporcionándole películas o material relacionado en cuanto hace clicks o termina de ver algo y al mismo tiempo hacer que la audiencia se acerque a ver películas o series que están más escondidas.

Problema planteado: En ocasiones el usuario no se anima a buscar material nuevo si no que se espera a lo que le aparece en el inicio por lo tanto si no aparecen cosas de su interés el usuario va a decidir irse.

Solución: Usar patrones secuenciales para que le aparezca al usuario material relacionado con lo que ha visto o seleccionado y arboles de decisión en donde dependiendo del historial te muestre de diferentes formas la portada de una película o la serie. También estudiar los días en los que el usuario utiliza mas la aplicación en donde los días más concurridos mostrar shows que no son muy populares y que pueda conocer películas o series que se encuentran más escondidas pero que están relacionadas con lo que el historial muestra del usuario y en los días en donde el usuario pasa un menor tiempo en la aplicación se le presenten principalmente su material favorito para no recurrir al aburrimiento y decida irse. Haremos uso de la base de datos de "Titulo de película o programa de televisión", "Director" y datos que nos proporcione el historial de cada usuario.