

Ejercicio práctico base de datos.

Materia: minería de datos

Maestro: Mayra Cristina Berrones Reyes

Integrantes: Gabriel Adrian Contreras García 1752950

Semestre: 7mo

Licenciatura. Actuaría

Monterrey N.L. a 14 de Octubre del 2020

Google Play Store Apps

Objetivo: Mejorar las recomendaciones que se muestran luego de la descarga de una app, con la finalidad de que los usuarios luego de descargar la aplicación que estaban buscando puedan seguir descargando aplicaciones que podrían ser de su interés. Incluso atraer al cliente a comprar aplicaciones.

Problema planteado: Dentro de la Google play store se tienen miles de aplicaciones al alcance del usuario, uno como empresario le interesa que las aplicaciones sean descargadas y si se puede buscar también que los clientes compren las aplicaciones que son de paga, necesitamos generar una herramienta que nos ayude a recomendar aplicaciones que vayan de acuerdo a cada usuario, de esta manera se disparará el numero de descarga de aplicaciones e incluso la venta de algunas que son de paga.

Solución: Desarrollar una herramienta de aprendizaje máquina, en este caso utilizando la técnica de reglas de asociación para hacer recomendaciones personalizadas al usuario, para ello se hará uso de la base de datos de Google play store, con la cual podemos hacer asociaciones de las aplicaciones, se podría empezar por la categoría que selecciona el usuario al elegir una aplicación, también si el usuario tiene poca memoria se puede hacer un segundo filtro con aplicaciones no muy pesadas, luego el género de la aplicación y dependiendo de la edad sea la aplicación que te recomiende después de que descargues una aplicación. También podría hacer recomendaciones arriesgadas (refiriéndose a aplicaciones de paga) que, aunque es menos probable, el usuario podría terminar comprando la aplicación.

Novel Corona Virus 2019 Dataset

Objetivo: seccionar a los pacientes con covid 19 de forma que podamos tener a pacientes con alto riesgo, riesgo medio, y bajo riesgo, esto con la finalidad de realizar una mejor atención medica a la hora de ingresar el paciente al hospital, asi se puede dar una atención mas rigurosa a los que tienen alto riesgo y disminuir el número de muertes.

Problema planteado: en los hospitales se tienen pacientes los cuales presentan los mismos síntomas de covid 19 pero no a todos les afecta de la misma manera esta enfermedad, se tiene cierto numero de pacientes con características similares que presentan o les afecta mas la enfermedad y de una manera mas rápida, es importante para el hospital salvaguardar el mayor numero de infectados, esperando que con la atención medica brindada puedan recuperarse de esta enfermedad, para darle un mayor peso a los pacientes que tengan un alto riesgo o que se vean mas afectados por la enfermedad se debe buscar una solución.

Solución: Desarrollar una herramienta de aprendizaje máquina, en este caso utilizando la técnica de clustering, para poder seccionar a los pacientes de covid19 con alto riesgo, riesgo medio y bajo riesgo, esto se haría por medio de particiones basándonos en similitudes como podrían ser su edad, sexo, lugar de origen, y si en el pasado histórico de la base de datos las personas de estas características se recuperaron o no, esto podría mejorar aun estas agrupaciones y poder tener una mejor estimación para cuando llegue un nuevo paciente poder atenderlo de mejor manera ya sabiendo su información.

Wine Reviews

Objetivo: poder diseñar mejores productos que satisfagan mejor las necesidades de nuestros clientes, debemos de buscar maneras de tener mas contentos a los clientes, elaborando vinos de mejor calidad, basándonos en lo que les gusta a los clientes, donde estaría el mercado al que nos dirigiríamos y de donde conseguir la uva para producir el vino.

Problema planteado: hay muchas variantes en el área de los vinos, cada uno con características diferentes, preparaciones especiales dependiendo el lugar de origen, pero en donde se decide todo es en el puntaje que le da el catador al vino, no necesariamente el precio influye en el puntaje, es una variante a tomar en cuenta pero debemos de tomar a consideración más factores, nuestro problema es que necesitamos mejorar las puntuaciones de nuestros vinos, dándole al cliente lo que a él más le gusta.

Solución: gracias a la base de datos de las reviews de vinos podemos tomar mucha información relevante y armar una herramienta de aprendizaje máquina, podríamos crear reglas de asociación entre los diferentes vinos que se han hecho review, poder observar el lugar de origen, los viñedos de donde proviene la uva que utilizan, la descripción del vino, el puntaje que le dieron, el precio que este tiene y al final poder hacer una selección de todos estos conjuntos para elaborar un vino que se adecue a diferentes perfiles de catadores, poder agarrar las características de los vinos que mejor puntúan, y utilizando estas preferencias poder crear un buen vino que tenga una excelente relación calidad/precio.

Iris Species

Objetivo: clasificar las especies de flores en base a imágenes que los compradores nos proporcionen. En nuestra florería creemos que sería más sencillo para el usuario si este nos presenta una foto de la flor que el busca (en caso de que no sepa el nombre de ella) para así recomendarle arreglos que contengan dicha flor.

Problema planteado: a veces los usuarios se pasan horas buscando la flor que ellos quieren comprar porque no recuerdan o no saben el nombre de la flor que quieren, a veces esto puede llegar a cansar al usuario y terminar por no comprar nada de la tienda, debemos buscar una solución a esto y poder permitir al usuario tener de una manera mas sencilla acceso a esa flor que el busca, pero no sabe cómo se llama.

Solución: desarrollar una herramienta de aprendizaje de maquina que permita identificar por medio de una foto que tipo de flor es la que se encuentra en ella, dependiendo de las características que presente como podría ser largo y ancho del sépalo, del pétalo, color, etc. Esto podría ser llevado acabo por medio de un árbol de decisión, que este pueda separar las características de la flor y dependiendo de ellas le presente un resultado al usuario, de esta forma seria más sencilla la búsqueda y podría mejorar el servicio que se le brinda a los compradores, permitiéndoles no batallar a la hora de investigar el nombre de la flor que buscan y aumentando las ventas del negocio.

Netflix Movies and TV Shows

Objetivo: Brindar al usuario mejores recomendaciones que se adecuen a su perfil y a las series y películas que frecuenta ver, la zona geográfica del cliente, y demás factores, esto con el fin de mantener mas tiempo al usuario dentro de la aplicación consumiendo contenido.

Problema planteado: necesitamos mantener al usuario consumiendo contenido de la aplicación el mayor tiempo posible, necesitamos mantener la atención del usuario y este es el problema, debemos de mejorar las recomendaciones que salen al final de una película o serie, deben ser más asertivas y generar una mejor respuesta para que el usuario decida ver esta recomendación y se quede mas tiempo dentro de la aplicación.

Solución: desarrollar una herramienta de aprendizaje de máquina que permita seleccionar diversas recomendaciones utilizando datos como lugar de origen de la serie que está viendo, el género, el total de duración, el director, los actores, esto podría llevarse a cabo por medio de la técnica de reglas de asociación, dependiendo de las selecciones que hizo el usuario anteriormente le predice una nueva recomendación que se adecue a su perfil.