



Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico - Matemáticas



MINERIA DE DATOS
Ejercicio Base de Datos

Nombre: Pilar Abigail Mendoza Alvarez

Matricula: 1815973

Maestra: Mayra Cristina Berrones Reyes

Grupo: 002

Miércoles 14 de octubre de 2020

Google Play Store Apps



Objetivo: Incrementar el número de descargas en aplicaciones relacionadas con la cultura (como aplicaciones acerca de trivias de cultura general, de arte, entre otras) y la lectura para mejorar el nivel de cultura a cierto sector de una población.

Problema planteado: Entre tanta variedad que existe en las tiendas de aplicaciones (específicamente Google Play Store en este caso por cuestiones de la base de datos) podrían llegar a perderse las aplicaciones relacionadas con la lectura y la cultura, teniendo una menor difusión de estas *apps*.

Solución: Buscar si existe alguna relación en cuanto al número/porcentaje de usuarios que realizan una reseña detallada de este tipo de aplicaciones, el hecho de que sea gratis, el dispositivo que posee el usuario, y/ o alguna otra variable para centrar la atención sobre estas y por consecuencias aumentar también el número de descargas.

Novel Corona virus 2019



Objetivo: Determinar las distintas probabilidades que tienen los habitantes de una región en específico de contraer el virus dependiendo de sus características individuales para proporcionar una mayor conciencia en los sectores geográficos más afectados.

Problema planteado: Con esta contingencia, la población debería, en mi opinión, estar consiente de a lo que se expone, y aun y cuando existen ciertos indicadores generales creo que también se podrían encontrar características en común más específicas para que el individuo en cuestión logre sentirse más identificado y tomar las medidas preventivas que considere necesarias basándose en eso.

Solución: Con ayuda de técnicas de agrupación, regresión lineal y múltiple , con herramientas como los arboles de decisión y clustering podemos determinar si existe además de una relación entre estas variables ya mencionadas, las probabilidades dependiendo del lugar en donde radique determinada persona, junto con sus características personales.

Wine Reviews



Objetivo: Encontrar la compatibilidad que muestran los clientes de un restaurante nuevo hacia ciertos vinos dependiendo las características que posean y el tipo de segmento de población al cual se quiera enfocar el restaurante en cuestión apoyándonos en la opiniones proporcionadas en la base de datos (considerando sus características y calificaciones).

Problema planteado: Suponiendo que un nuevo restaurante de lujo ubicado en la ciudad requiere tener en su menú una variedad de buenos vinos. Para esto, se requiere investigar en distintos lugares y con ayuda de encuestas para identificar a los mejores vinos conforme a las características que busca el restaurante.

Solución: Con ayuda de técnicas de agrupación complementándolo también con modelos predictivos, podemos desarrollar modelos que nos ayuden a determinar si los clientes de este restaurante estarían dispuesto a consumir alguno de los vinos proporcionados en la base de datos para elegir el o los que mejor se adecuen al público del restaurante.

Iris Species



Objetivo: Encontrar el lugar y esparcimiento adecuado para cada especie de la planta iris en el invernadero dependiendo de las características que muestran las especies en general por separado para lograr una distribución eficaz.

Problema planteado: Considerando que un invernadero podría presentarse la disyuntiva de la distribución de la planta de iris por sectores diferentes clasificándolos por especies por cuestiones de espacio, terreno, exposición al sol, entre otras.

Solución: Basándose en la información de la base de datos presentada, usar el machine learning para identificar el promedio en las medidas sobre el sépalo y el pétalo de cada especie y así poder clasificar las plantas correctamente y encontrar el espacio que sea más conveniente tanto para la planta como para temas de una buena distribución del terreno.

[illegible]

Problema planteado: La misma empresa Netflix requiere hacer contenido nuevo para satisfacer a sus clientes en su servicio de streaming. Para eso requiere hacer un estudio y recabar una gran cantidad de datos y decidir cuál es el contenido que consumen más sus usuarios en la región de Latinoamérica.

Solución: Se necesita desarrollar una máquina de aprendizaje para identificar cuál es el tipo de contenido más popular para satisfacer a los usuarios en Latinoamérica.