

Universidad autónoma de Nuevo León

Minería de datos

Ejercicio práctico de bases de datos

Profesora: Mayra Cristina Berrones Reyes

Alumno: Miguel Eduardo Ovalle Blanco 1801990

Grupo: 002

Datos Google Play Store (Google play store apps)

**Objetivo:** Ayudar a verificar que las aplicaciones que se encuentran en la tienda sean de buena calidad o acorde a la buena reputación que se le ha otorgado

**Problema planteado:** Normalmente cuando buscamos una aplicación observamos que al momento de comparar es fácil ver la reputación, pero resulta que muchas también tienen una reputación similar, por lo que una segunda medida sería observar los comentarios que se encuentran de cada aplicación. Entonces esta es una oportunidad para que pueda entrar ese nuevo proceso de verificación ya que el checar todo ello es gastar tiempo del cliente que, aunque se debe de tomar su tiempo para elegir podemos ahorrarle eso para mejorar su experiencia al momento de averiguar cual se ajusta a sus preferencias o a lo que le da mayor importancia, así de esta manera ir empezando a tener un buscador mejorado que cumpla con dicha verificación.

**Solución:** A partir de los datos que tenemos podemos notar que tenemos la valoración, reseñas e instalaciones que nos pueden ayudar a saber de una mejor manera si una aplicación es verdaderamente buena o no, entonces con el aprendizaje de máquina podemos realizar la búsqueda de los tipos, la categoría, clasificación de edad, si es de pago y tamaño, por lo que el buscador también tendrá añadido el verificador de que tan bueno puede resultar una aplicación, con esto las personas no solo encuentran las aplicaciones de acuerdo a su necesidad sino que se le presentan las mejores recomendaciones para ello y se resuelva su problemática de la mejor manera posible.

Datos Coronavirus (Novel Corona Virus 2019)

**Objetivo:** Buscar una manera en la cual podamos encontrar un comportamiento en como se ha ido desplazando la pandemia de coronavirus en la semana del 22 de enero al 27 de enero del 2020

**Problema planteado:** Podemos observar que tenemos datos importantes como las recuperaciones, las muertes, el número de contagiados y también la fecha en que fueron reportados en diferentes lugares del mundo, con lo que podemos tener una visión global de lo ocurrido en esas fechas, por lo tanto el hacer diferentes inferencias acerca de como se comportó en ese entonces puede ser una herramienta clave para poder combatir en cuestión de predicción no solo en este caso del presente sino para poder ver en diferentes pandemias como se ha extendido de esa manera.

**Solución:** Mediante aprendizaje de máquina podemos realizar predicciones en el comportamiento similar comparando en diferentes lugares del mundo como se ha comportado el nivel de mortalidad, así que como se menciona de un análisis global, con ello podemos formar diferentes tácticas o niveles dependiendo de la región ya que como se puede observar es de una forma diferente el como es el comportamiento de la disipación de la enfermedad por la región.

## Datos Críticas de vinos (Wine Reviews)

**Objetivo:** Poder realizar una predicción de vinos dependiendo de ciertas características que se presenten con lo cual se puede ajustar recomendaciones para elegir bien un vino, ya sea para comprar los en un lugar dependiendo de como se describa o ya sea para una investigación en ello.

**Problema planteado:** Como tenemos dicho, esto va a interesar en que una persona esta interesada en un vino con ciertas características por lo tanto tenemos que a partir de como fue descrito según los expertos, se puede predecir la región de como es, el precio que tiene y su valoración. Entonces eso ayudaría en una tienda de vinos que dependiendo de cada cliente se puede ofrecer algo con los gustos del cliente, de hecho, también existen cursos para poder catar vinos con lo cual sería de mucha utilidad para fines didácticos un programa de predicción en ese aspecto, con lo cual facilitar el aprendizaje y tenerlo de una forma más dinámica. De igual forma como se presento también una forma de consulta para personas que están interesadas en ello, así que de igual forma sirve para aumentar o conocer aún más el vino.

**Solución:** La solución que se presenta es mediante el aprendizaje de máquina se usar una herramienta de relación o alguna decisión ya que presentando características puede irse averiguando que tipo de vino es el que se quiere, por lo que estas herramientas pueden ser de gran utilidad dado a que teniendo el fin de poder hallar el vino al que se trata de describir se puede encontrar de una forma eficiente

## Datos Clasificación de plantas (Iris species)

**Objetivo:** Poder encontrar el tipo correcto de la planta que se está hablando, dependiendo de las características que se tienen de una hoja muestra de dicha planta.

**Problema planteado:** Normalmente ha ocurrido que una persona puede encontrar una planta la cual le agrada mucho, o un caso de estudio en donde se pueda investigar acerca de estos tres tipos de planta que se presenta en la base de datos Iris, donde podemos decir que tenemos de datos como la longitud del sépalos, su anchura y de la misma manera contamos con esa información pero del pétalo, por lo que a partir de estos datos podemos averiguar si pertenece a una clase de tres plantas que tienen dichas características

**Solución:** Como se ha mencionado, tenemos un problema de clasificación de la planta que tenemos y mediante las características que se presentan del pétalo y sépalos podemos decir que le podemos asignar ciertas probabilidades al momento de que puedan pertenecer a cierto grupo, o como se planteó como un árbol de decisión en donde va yéndose con las características que hemos dicho.

## Datos Shows de Netflix (Netflix Movies and TV shows)

**Objetivo:** Poder saber en qué momento se puede sacar una serie/película, con qué características y, además, si ya tenemos una existente, saber si seguir dándoles continuidad dependiendo del gusto del público.

**Problema planteado:** De los datos interesantes que tiene este conjunto es la fecha en que se estuvo poniendo en el catalogo de Netflix, por lo que también con los datos de el tipo de elenco, su director, la categoría, su duración y el rating que tuvieron, ayuda en poder observar cómo le puede ir a una futura producción que tenga cierto estilo, por lo cuál el hacer programas/películas que mejor se adaptan a las necesidades o gustos que puede tener cierto público. El saber que puede ser una buena producción sería una herramienta vital para este tipo de compañía ya que ayudaría a ganar más el interés de las personas en adquirir suscripciones o el que haya mejores recomendaciones de acuerdo con los gustos de las personas que, gracias a este tipo de generación de producciones buenas, puedan encontrar muy buenos gustos en ello.

**Solución:** Una solución en donde podemos usar el aprendizaje de máquina en este sentido sería para poder encontrar esa tendencia de como podría ser un lanzamiento o el tipo de características que son demandadas por el público, de igual manera incursionar en proyectos que se puedan arriesgar ya que eso nos ayudaría en abrir un poco más el panorama de conocimiento de nuestros clientes que decidirán después si algo es bueno o malo. El escuchar lo que quiere la audiencia y lo que se planea debe de ser cosas que trabajen en conjunto.