

A dark blue vertical bar is positioned on the left side of the page. A blue arrow-shaped banner points to the right from this bar, containing the date. Below the banner, several thin, curved lines in shades of blue and grey sweep upwards from the bottom left corner.

14 de octubre del 2020

# Ejercicio práctico base de datos

**Materia:** Minería de Datos

**Profesor:** Mayra Berrones

**Alumno:** Aranza Sifuentes Soto

**Matrícula:** 1887979

**Grupo:** 002

A

#### A)

- **Nombre de la base de datos:** Google Play Store Apps
- **Objetivo:** Realizar un ranking semanal, mensual y anual de las aplicaciones de Play Store más descargadas, ya sea por categoría y en general.
- **Problema planteado:** Es muy tardado clasificar y ordenar todas las aplicaciones manualmente, además de que es conveniente para Play Store realizar dichos rankings, ya que es llamativo para los usuarios ver qué aplicaciones tienen más descargas por cierto periodo de tiempo, y con ello poder descargar las aplicaciones más usadas actualmente, y con ello Google Play Store obtiene más descargas de las planteadas.
- **Solución:** Desarrollar una herramienta automatizada que ordene las aplicaciones por mayor número de descargas en cierto periodo de tiempo, obteniendo así el ranking semanal, mensual o anual; Y además de eso, que dicha herramienta nos permita clasificar cada aplicación, ya sea en juegos, social, etc. Obteniendo con ello que cada usuario vea la aplicación favorita actualmente dependiendo de la categoría que él busque o le interese.

#### B)

- **Nombre de la base de datos:** Novel Coronavirus 2019 Dataset
- **Objetivo:** Crear una curva epidemiológica del comportamiento de los casos nuevos en países donde ya se tiene controlado el coronavirus, y con ello predecir el comportamiento de casos nuevos en países con condiciones similares (número de población, cultura, extensión territorial, etc.).
- **Problema planteado:** Existen países donde ya se tiene controlado el COVID, y algunos otros donde apareció recientemente y por lo tanto apenas va iniciando la transmisión, por lo que no se puede predecir el tiempo que durará el virus, cuándo será el punto más alto de la curva epidemiológica o cuántos casos de contagios se esperan y cómo abordarlo de la mejor manera.
- **Solución:** Desarrollar una herramienta que nos permita predecir el número de contagios aproximados en un país, la duración, entre otras cosas importantes basándonos en los datos obtenidos de otros países con características similares de población, y con ello prevenir una catástrofe o un número masivo de contagios.

C)

- **Nombre de la base de datos:** Wine Reviews
- **Objetivo:** Predecir la puntuación de los vinos dependiendo de sus características, por ejemplo, lugar de origen, región o viñedo; una vez obtenido eso predecir el precio de acuerdo con su puntuación
- **Problema planteado:** Es muy tardado asignar o revisar todas las puntuaciones y precios de cada vino, es por ello que es necesario predecirlo de alguna manera y realizarlo automático. Además de que, puede ser una gran alternativa para productores de algún vino nuevo, poder saber su posible calificación y poder asignar un precio adecuado para el mercado basándose en las características de su vino.
- **Solución:** Analizar la base de datos y realizar una regresión para ver cómo se comportan los datos y que tanto se relacionan las variables mencionadas. Con ello, realizamos una herramienta que nos permita predecir la calificación y posteriormente el precio de cualquier vino.

D)

- **Nombre de la base de datos:** Iris Species
- **Objetivo:** Poder predecir el tipo de planta de acuerdo con las características y medidas dadas.
- **Problema planteado:** Al momento de seguir registrando más plantas (ya que la base de datos sólo es una muestra de 50) es más fácil capturar las medidas y características de la planta, como lo son las medidas de su pétalo o de su tallo, y con ello rellenar automáticamente el tipo de planta que es. O en otro caso, no saber de que tipo de planta se habla, pero con las características y medidas obtenidas saberlo.
- **Solución:** Realizar una regresión con los datos que tenemos para saber cómo se relacionan las variables dadas en nuestra base de datos, posteriormente realizar un modelo que nos permita predecir el tipo de planta que es, y así realizar una herramienta que prediga el tipo de planta, sólo dándole las medidas que tenemos de cada planta.

E)

- **Nombre de la base de datos:** Netflix movies and tv shows.
- **Objetivo:** Encontrar las películas que más llamen la atención de la audiencia, y que sean de su preferencia; obteniendo los resultados por zona (Latinoamérica, EEUU, etc)
- **Problema planteado:** Es muy difícil identificar manualmente qué tipo de películas o, series en su defecto son de preferencia al público, y con ello existe el dilema de qué películas nuevas agregar, cuáles quitar, de qué tipo de género agregar a la plataforma.
- **Solución:** Desarrollar una plataforma que clasifique todas las películas y series dependiendo de su género, y a su vez ordenarlos por mayor a menor número de reproducciones y calificación; con ello se podrá identificar fácilmente qué género es el preferido para los usuarios de cada zona, y qué tipo de películas y series ; así se podrá decidir qué películas agregar, cuáles dejar más tiempo y cuáles quitar por su baja aceptación.