



**Universidad Autónoma de Nuevo León**

**Facultad de Ciencias Físico – Matemáticas**

**Licenciatura en Actuaría**

***Minería de datos***

**Profesora:**

**Mayra Cristina Berrones Reyes**

***Ejercicio de bases de datos***

**Equipo 3**

Ana Saraí Dávila Martínez	1986209
Jesús Mauricio Rodríguez Pacheco	1583211
Milton Aldair Martínez Acosta	1870354

## Ejercicio de base de datos

- **Nombre de la base de datos:** Aplicaciones de Google Play Store. Datos extraídos de la web de aplicaciones de 10k Play Store para analizar el mercado de Android.
- **Objetivo:** Encontrar una oportunidad de inversión en una aplicación que este diseñada para los teléfonos Android.
- **Problema planteado:** En el 2021 se estima que habrá más de 3.002.406 aplicaciones en el mercado de Google Play y al ser un mercado muy cotizado, pero disparejo podría haber una oportunidad para realizar una inversión que sea redituable y de bajo riesgo.
- **Solución:** Desarrollar un algoritmo, el cual se basará en un modelo de minería de datos de puntajes donde cada aplicación tendrá su métrica a cumplir y la aplicación con mejor puntaje en base a la métrica establecida que proporcione el modelo planteado será el seleccionado para hacer dicha inversión.



- **Nombre de la base de datos:** Novel Coronavirus 2019 Dataset, de la organización mundial de la salud, del 31 de diciembre del 2019.



- **Objetivo:** Agrupar las diferentes zonas geográficas en todo mundo con un modelo de minería de datos, dicho modelo se basará en los casos confirmados y muertes por COVID-19, esto con el fin de planificar y realizar políticas para reducir el impacto del virus SARS-COV-2.
- **Problema planteado:** Los gobiernos de distintas zonas del mundo necesitan tener políticas de medidas para mitigar la propagación del COVID-19 ya que este virus podría dañar partes vitales del cuerpo humano como el pulmón, corazón y cerebro, y en casos con mayores complicaciones causar la muerte, por lo que es necesario tener una planeación para enfrentar este problema e incluso desarrollar políticas de vacunación óptimas contra el virus, por esta razón el análisis de los agrupamientos más afectados ayudaría a la toma de decisiones.
- **Solución:** Desarrollar un algoritmo, el cual se basará en un modelo de minería de datos de agrupación. Esta herramienta permitirá evaluar la incidencia de casos positivos del virus SARS-COV-2 respecto a la ubicación geográfica. En adición, se contemplarán otros indicadores como la tasa de letalidad y tasa de incremento diario, esto con el fin de agrupar las zonas geográficas más vulnerables y permitir el diseño de políticas contra el COVID-19 a través de una herramienta de aprendizaje automático.

→ **Nombre de la base de datos:** Iris Species Dataset, extraído del artículo de Fisher *The Use of Multiple Measurements in Taxonomic Problems* en 1936.



→ **Objetivo:** Optimizar la producción de la planta Iris, esto basándose en técnicas de aprendizaje automático.

→ **Problema planteado:** El cultivo y producción de plantas medicinales y ornamentales, como por ejemplo la planta de iris, la cual tiene propiedades medicinales que podrían ser relevantes para la industria farmacéutica y cosmética. En la producción puede haber problemas relacionados con la nutrición y las enfermedades que se podrían presentar en los cultivos.

Además es de vital importancia el monitoreo de la calidad en la producción, por lo que se necesitan herramientas que provean un panorama en la industria de la agricultura.

→ **Solución:** La creación de un modelo de predicción basado en técnicas de aprendizaje automático permitiría establecer estándares de calidad en cuestiones de longitud en las plantas(ancho y largo) y además podría indicar si existe alguna irregularidad en el cultivo de las plantas de la especie Iris.

- **Nombre de la base de datos:** Netflix Movies and TV Shows, extraído de **NETFLIX** Flixable (motor de búsqueda externo).
- **Objetivo:** Conocer cuáles son los géneros de contenido más populares en cada país en el cual está disponible la plataforma.
- **Problema planteado:** Se necesita aumentar la cantidad de personas suscritas para generar una mayor ganancia para la empresa, por lo cual se considera adecuada la estrategia de brindar un contenido más personalizado a cada país, esto se lograría mejorando la interacción del usuario con la plataforma y analizando patrones.
- **Solución:** Desarrollar un algoritmo, el cual se basará en un modelo de minería de datos de agrupación. Agrupando los diversos géneros según el país y la popularidad de estos, para ello se contemplarán las búsquedas, las puntuaciones, los contenidos que son buscados y la fecha en que se visualiza, de esta manera se prevé tener un aumento en la cantidad de suscriptores en la plataforma de Netflix.



- **Nombre de la base de datos:** Wine reviews, extraído de WineEnthusiast el 15 de Junio de 2017.
- **Objetivo:** Desarrollar un modelo que pueda predecir la calificación de un vino con base en sus datos de procedencia y de elaboración, para incrementar las ventas en empresas vitivinícolas
- **Problema planteado:** La calificación de un vino es una tarea subjetiva y lenta, se pretende automatizarla y en base a estas mejoras en el proceso, aumentar la calidad y los estándares de producción del vino, esto con la finalidad de aumentar las utilidades que generen las empresas vitivinícolas.
- **Solución:** La creación de un modelo de predicción basado en técnicas de aprendizaje automático y minería de datos, que permitieran establecer juicios a nuevos vinos con base en lo que todas las otras calificaciones han introducido, de esta manera se pretende mejorar el proceso de producción y comercialización de vinos.

