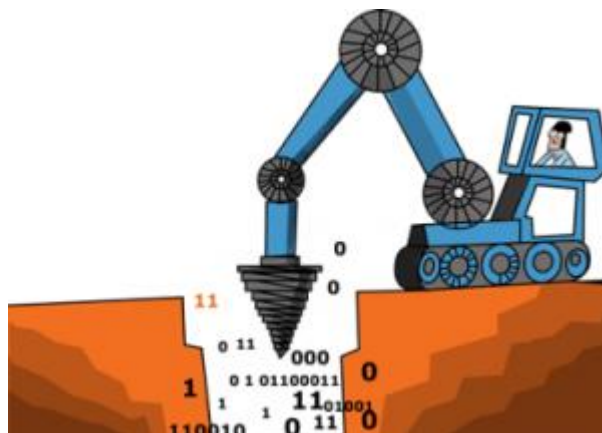


Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas



Selene Jazmín Rodríguez González

1834309

Minería de datos

Grupo 012

Maestra Mayra Berrones



Google play store

Objetivo:

Analizar la base de datos con el fin de encontrar coincidencias entre una aplicación y otra.

Problema planteado:

Revisando la base de datos de manera superficial, se puede observar que hay una gran diferencia entre el número de instalaciones y el número de reviews. Esto probablemente querría decir que la aplicación resulto interesante para el usuario, pero no fue de su agrado una vez que la instalo.

Solución:

Se podría utilizar algún procedimiento para crear grupos entre los datos (clustering) y una vez que se tengan los grupos, evaluar por medio de reglas de asociación cual es el motivo por el que estas aplicaciones son desinstaladas. También se puede hacer un análisis de los comentarios de los usuarios para así identificar las áreas de oportunidad.

Coronavirus

Objetivo:

Encontrar similitudes en los datos para intentar encontrar cuales son los principales factores que hacen que la enfermedad incremente.

Problema planteado:

Evaluando las gráficas de los datos se puede identificar que hay más infectados masculinos que femeninos. Y si bien, hay infectados jóvenes, se puede ver que la mayoría se ubica entre personas de más de treinta años.

Solución:

Evaluar las comorbilidades de todos los pacientes contemplados en la base de datos para intentar encontrar similitudes entre una persona y otra por medio de clustering. Si se identifica el grupo que posea mayor riesgo, este podría aislarse para evitar la propagación de la enfermedad.



Criticas de vinos

Objetivo:

Identificar cual es el país que posee mejores críticas en cuanto a vino.

Problema planteado:

Se ha corroborado que US posea una mayor cantidad de vinos, pero no implica que posea los vinos con mejor opinión respecto a los catadores.

Solución:

Se pretende efectuar un análisis de las opiniones de los vinos haciendo uso de patrones secuenciales, y también de las calificaciones de cada uno de ellos para determinar cuál es el país con mejores críticas.

Clasificación de plantas

Objetivo:

Identificar alguna anomalía dentro de las especies estudiadas.

Problema planteado:

Pese a que solo se estudian tres especies, se puede observar que existen variaciones entre una y otra.

Solución:

Analizar todas las medidas de las flores haciendo uso del procedimiento de Outliers para así identificar si hay algún espécimen que no figure entre el rango normal de medidas. Si se identificara alguno que tenga una variación muy grande, serviría para poder investigarlo mejor y determinar si es un error.



Shows en Netflix

Objetivo:

Identificar los directores de cine que más aparecen para agregar contenido del director en cuestión.

Problema planteado:

Se requiere analizar la base de datos para establecer una predicción de que artistas o directores podrían ser del agrado del público y así incluir más contenido en la plataforma.

Solución:

Se podría hacer uso de regresión lineal para analizar tanto los directores como los artistas y así establecer una predicción de que contenido deberían agregar a la plataforma.