Encontramos un dataset con mas de 50mil filas relacionando datos de un sensor accelerometro y giroscopio <https://www.kaggle.com/datasets/aaryapandya/earthquake-detection>

Con las columnas ax, ay, az, gx, gy, gmz y Result

Se hizo una limpieza al dataset, eliminando las columnas y datos vacíos.

Se contabilizaron el numero de muestras de la clase Result

Texto

El contenido generado por IA puede ser incorrecto.Texto

El contenido generado por IA puede ser incorrecto.

Se limpiaron los valores nulos por columna

Imagen que contiene Texto

El contenido generado por IA puede ser incorrecto.

Se normalizaron los datos com StandarScaler para poder hacer comparaciones

Texto

El contenido generado por IA puede ser incorrecto.

Se utilizo un mapa de correlación(heatmap) de las variables para relacionarlas entre si y verificar que los datos si están relacionadas con la variable Result

Texto

El contenido generado por IA puede ser incorrecto.  
Imagen que contiene Tabla

El contenido generado por IA puede ser incorrecto.

Podemos observar una fuerte relación entre la variable gX y Result que no es la mejor pero mas adelante agregaremos nuevas variables.

Tambien se genero la distribución de variables en base a Result   
Gráfico

El contenido generado por IA puede ser incorrecto.Tabla

El contenido generado por IA puede ser incorrecto.

Gráfico, Gráfico de líneas, Histograma

El contenido generado por IA puede ser incorrecto.Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Imagen que contiene Tabla

El contenido generado por IA puede ser incorrecto.Tabla

El contenido generado por IA puede ser incorrecto.

En la distribución de las variables en base a Result(0,1) podemos observar la gran diferencia que existen entre 0 y 1.

Tambien podemos se hizo un análisis de componentes principales (PCA) en 2d

Para visualizar patrones

Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.

Con base a lo anterior se denomino que el dataset puede ser útil para la creación del modelo pero decidimos agregar algunas columnas de transformaciones útiles al dataset como Magnitud del vector y Valor absoluto.

Pantalla de computadora con letras

El contenido generado por IA puede ser incorrecto.

Luego se normalizaron y se hizo otro mapa de correlación Calendario

El contenido generado por IA puede ser incorrecto.

Donde podemos observar ahora un total de 4 correlaciones significativas con Result

Que son gX, a\_magnitude, g\_magnitude y abs\_gX

También con su distribución  
Gráfico

El contenido generado por IA puede ser incorrecto.Gráfico

El contenido generado por IA puede ser incorrecto.

Gráfico, Gráfico de líneas, Histograma

El contenido generado por IA puede ser incorrecto.Gráfico

El contenido generado por IA puede ser incorrecto.

Siendo similar a las anteriores con una clara diferencia de distribución por variable

Podemos observar en el siguiente pca el cambio significativo que hubo de la versión anterior teniendo en este nuevo una clara separación de grupos entre Result 1 y 0

Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.