Miniproyecto – Deep Learning

Grupo 12

Dsharlie Sanchez Rozo - Christian Beraun Chamorro - William Morales Varela

Repositorio: [**Link**](https://github.com/elwam/miad_deep_learning/blob/main/competencia_1/competencia_1_Final.ipynb)

# Exploración de los datos para su entendimiento dentro del contexto organizacional.

* El set de datos está compuesto por 77.779 filas y 92 columnas.
* Se eliminaron las columnas ID y V3 antes del análisis.
  + La columna ID se eliminó ya que es el identificador único de cada canción en el set de datos.
  + La columna V3 se eliminó ya que no contiene valores.



Se realizo el análisis sobre las 12 primeras variables del set de datos, de las cuales es posible observar que tienen una distribución normal y no hay variación en la desviación estándar significativa.

Tabla

Descripción generada automáticamente

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

# Preparación de los datos para poder utilizarlos como entrada para modelos predictivos.

Tomando como base las primeras 12 variables del conjunto de datos, hemos calculado nuevas variables que servirán como predictores durante el proceso de entrenamiento.

1. mean\_timbre (Timbre promedio)
2. median\_timbre (Timbre mediano)
3. std\_timbre (Desviación estándar del timbre)
4. sum\_timbre (Suma del timbre)
5. max\_timbre (Máximo del timbre)
6. min\_timbre (Mínimo del timbre)
7. prod\_timbre (Producto del timbre)
8. mean\_covariance (Covarianza promedio)
9. median\_covariance (Covarianza mediana)
10. std\_covariance (Desviación estándar de la covarianza)
11. sum\_covariance (Suma de la covarianza)
12. max\_covariance (Máximo de la covarianza)
13. min\_covariance (Mínimo de la covarianza)
14. prod\_covariance (Producto de la covarianza)

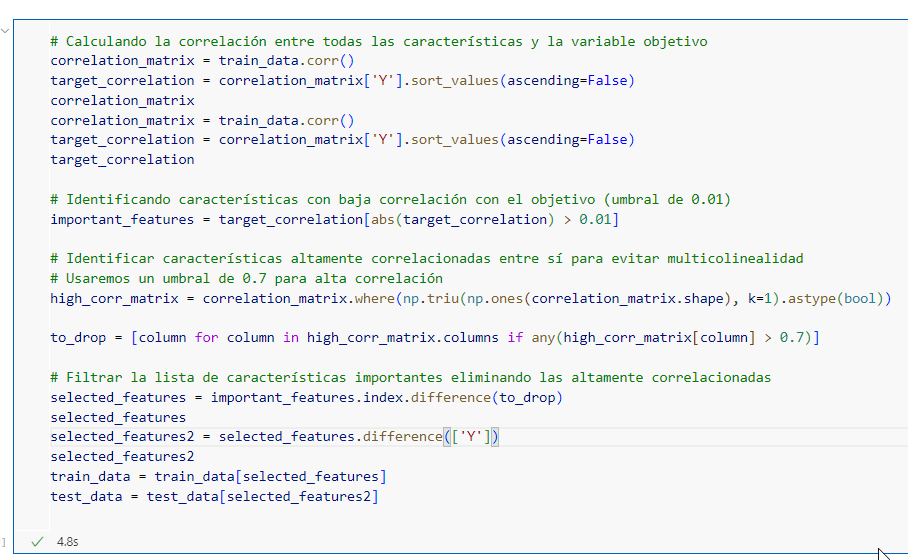


Una vez creadas las variables se hace un análisis de correlación de las variables nuevas vs la variable objetivo.

Gráfico, Gráfico en cascada

Descripción generada automáticamente

Tomando como basa estas nuevas variables, se valida la correlación de todas las variables predictoras y se seleccionan aquellas variables que tiene una correlación mayor a 0.7.



Posterior a ello se procedió a estandarizar las variables predictoras utilizando StandardScaler. Esta estandarización ajusta las variables para que tengan una media de 0 y una desviación estándar de 1, lo que facilita el proceso de entrenamiento de modelos y asegura una mejor convergencia durante el proceso de optimización.

# Análisis preliminar de selección de modelos relevantes para responder a la pregunta.

Para iniciar con el proceso de selección del modelo, se realiza la creación de una red neuronal con cuatro capas ocultas, utilizando funciones de activación Relu y Elu respectivamente y la última capa con activación linear teniendo en cuenta que es un problema de regresión.

Tabla

Descripción generada automáticamente

Tomando como base la red definida, se realizo un proceso de búsqueda de los mejores parámetros utilizando la técnica de Early Stoping para evitar el sobre ajuste, los parámetros y resultados fueron los siguientes:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente Tabla

Descripción generada automáticamente

Se continuó con el proceso y nos enfocamos en el optimizador Adam y se realizo una nueva búsqueda de parámetros con los siguientes resultados:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamenteTabla

Descripción generada automáticamente

# Desarrollo y calibración de modelos.

# Visualización de resultados.