

ELEMENTOS DE APRENDIZAJE DE MÁQUINA Y BIG DATA



Integrantes: Grupo 20

- Lucia ALONSO
- Pastor PAREDES
- Viviana SUAREZ
- Heber ALONSO

<https://github.com/lu8dcf/UPSO-BigData-G20.git>

LABORATORIO

Clasificador de canciones de Spotify

En este laboratorio, nos enfocaremos en un conjunto de datos repleto de atributos clave de distintas canciones. La meta es establecer a través de herramientas de aprendizaje automático, un clasificador que pueda predecir las inclinaciones musicales de un individuo y predecir si disfrutará o no de una determinada canción.

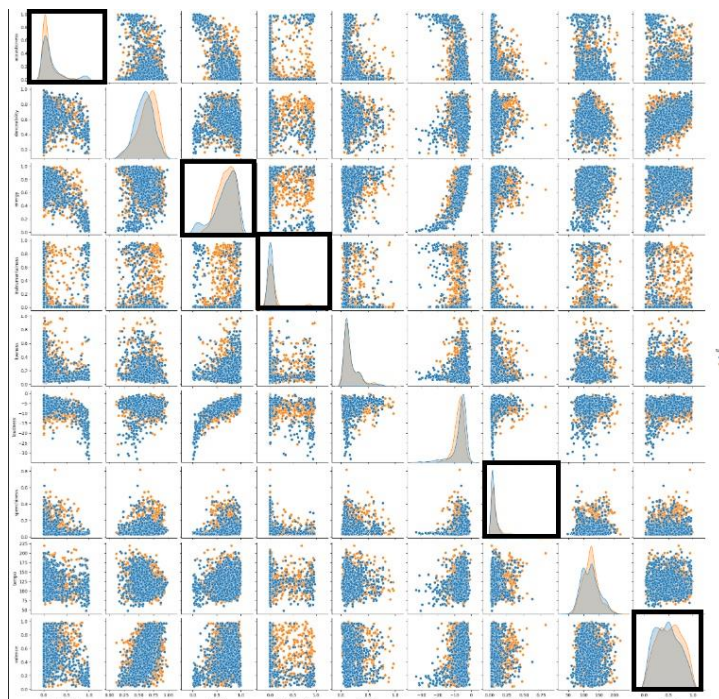
CONCLUSION

ARCHIVO clasificador_Spotify1.ipynb

Del dataframe original se separaron los parámetros que tienen información que quedaron nueve (9), y se establecieron los mejores valores de K hasta encontrar los mejores resultados.

ARCHIVO clasificador_Spotify2.ipynb

Para buscar mejores resultados se realizaron los gráficos de dispersión, se decidió seleccionar los parámetros con mayores diferencia, en el caso del segundo se utilizaron "acousticness", "energy", "instrumentalness", "speechiness" y "valence"

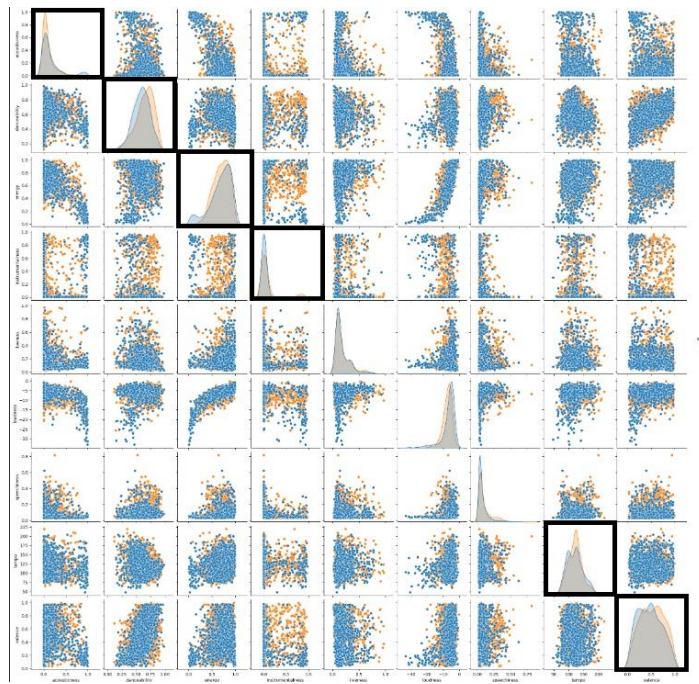


Del mismo modo se buscaron los mejores valores de K para obtener la mayor eficiencia del clasificador.

ARCHIVO clasificador_Spotify3.ipynb

Se seleccionaron 6 parámetros, "acousticness", "danceability", "energy", "instrumentalness", "tempo" y "valence"

CONCLUSION

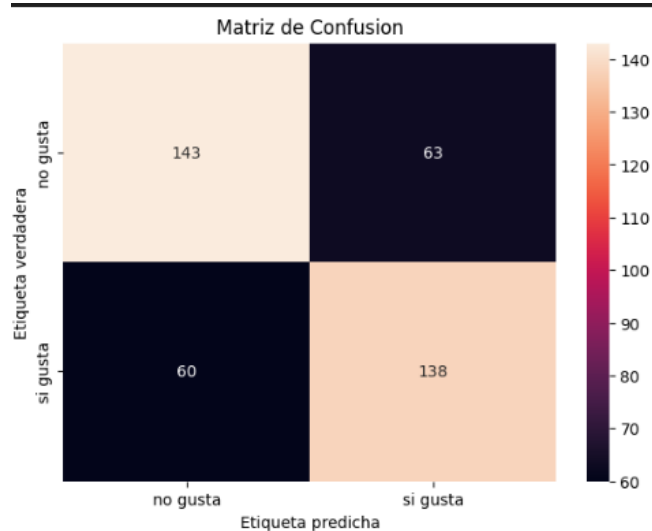


De los cuales se obtuvieron los siguientes resultados de precisión:

Tabla 1: Precisión de los diferentes modelos en función de la cantidad de parámetros utilizados.

PARAMETROS	KNN	SVN	Árbol	Bayes
9	0,63	0,67	0,69	0,64
5	0,66	0,68	0,62	0,59
6	0,59	0,61	0,65	0,59

- Mejor valor obtenido - Árbol de Decisiones con 9 parámetros : **0.69**



- Mejor valor en Bayes con 9 parámetros: **0.64**
- Mejor valor con KNN **0.66** con 5 parámetros.
- Mejor valor con SVN **0.68** con 5 parámetros

Conclusión:

- Utilizando todos los parámetros fueron más eficientes para el Árbol de decisiones, como en el caso de Bayes al utilizar menos parámetros la precisión disminuyó significativamente. Para el caso de KNN y SVN fue más efectiva la disminución de parámetros, tomando en cuenta los de mayor dispersión.
- A pesar que todos los valores están en un rango de 0,59 a 0,69, valores ya aceptables, se buscaron muchas combinaciones de distintos parámetros dando los mejores cambios resultados en la tabla 1.
- En relación el SVN es el más eficiente ya que con menos procesamiento, solo 5 parámetros, consigue casi el mismo resultado que el Árbol de decisiones con 9, a gran escala esto puede reflejarse en una mejor velocidad de respuesta o menos recursos necesarios.