TRABAJO FIN DE MÁSTER

MÁSTER EN BIG DATA Y BUSINESS ANALYTICS

IMF Smart Educación

Análisis y desarrollo de modelos de aprendizaje automático para predecir el éxito de canciones en Spotify

Autor: D. Sergio Martínez Martínez

Tutor: D. Juan Manuel Moreno Lamparero

Murcia, mayo de 2023

Índice

Índice de ilustraciones

Índice de tablas

RESUMEN

La finalidad de este proyecto consiste en la utilización de una serie de técnicas de aprendizaje automático relacionadas con Data Science sobre un conjunto de datos de Spotify. En concreto se realizará en primer lugar un proceso de preparación, procesado y análisis de datos, seguido del uso de diferentes modelos, tanto supervisados como no supervisados, sobre nuestro conjunto de datos para más tarde finalizar con una parte visual destinada al usuario final con una serie de gráficas que ayuden a comprender nuestros datos.

Para ello, contaremos con un archivo CSV que contiene las 200 canciones más escuchadas de 53 países en los años 2017 y comienzos del 2018. Sobre este archivo se realizarán numerosas actualizaciones, las cuales iremos guardando conforme sea necesario en una base de datos que tendremos a nivel local.

Se comenzará por hacer una preparación y limpieza del CSV, seguida de su almacenamiento en una base de datos. Todo esto con el objetivo de usar dichos datos para la elaboración de varios modelos de predicción, tanto supervisados como no supervisados.

Para los modelos supervisados su finalidad será predecir el número de Streams por canción, desarrollando para ello varios modelos con el fin de contrastarlos y elegir los que mejores resultados arrojen. En el caso de los modelos no supervisados su objetivo será poder agrupar y clasificar la información de nuestro CSV de forma que podamos entender mejor el comportamiento de los datos.

Finalmente, también habrá una parte completamente visual en el cual se mostrarán diferentes gráficas con las distintas modificaciones que se han ido realizando sobre nuestros datos base. Cabe resaltar que durante todo este proceso se guardará una copia en la nube la cual se puede encontrar en (Martínez, 2023)

ABSTRACT

The purpose of this project is to use a series of machine learning techniques related to Data Science on a Spotify dataset. Specifically, it will begin with a data preparation, processing, and analysis process, followed by the use of different supervised and unsupervised models on the dataset. The project will conclude with a visual component aimed at the end user, presenting a series of graphs to help understand the data.

To achieve this, we will have a CSV file with the top 200 songs from 53 countries in the years 2017 and early 2018. This file will undergo numerous updates across our work, which will be saved as necessary in a local database.

The project will start with the preparation and cleaning of the CSV file, followed by its storage in our database. The purpose of these steps is to use the data for the development of various prediction models of both types, supervised and unsupervised.

For the supervised models, the objective will be to predict the number of streams per song, developing multiple models to compare and select the best-performing one. In the case of unsupervised models, the goal will be to cluster and classify the information from our CSV in order to understand better the data's behavior.

Finally, there will also be a visual part, highlighting different graphs with the various modifications made to our base data. It should be noted that during this entire process a copy of our project will be saved in the cloud which can be found in (Martínez, 2023).

PARA MÁS INFORMACIÓN PUEDE ENCONTRARME EN:

FOR MORE INFORMATION, YOU CAN FIND ME AT:

LinkedIn:

https://www.linkedin.com/in/sergio-mart%C3%ADnez-a94255269/

Gmail:

sersergio65@gmail.com