Data Analytics Project Report

A.Y. 2023/2024

Claudia Citera

claudia.citera@studio.unibo.it

Riccardo Evangelisti

riccardo.evangelist6@studio.unibo.it

INDICE

da aggiungere

1 Introduzione

The following project was developed as part of the **Data Analytics** course of the Master's Degree in Computer Science at the University of Bologna.

The objective of the project is to carry out a data analytics study, which involves the implementation of all the analytical pipeline phases studied during the course:

- Data Acquisition
- Data Visualization
- Data Preprocessing
- Modeling
- Evaluation

The main purpose of this study is to recognize the year in which a song was published based on the features of its audio track.

The following functionalities were developed:

- Traditional non-deep supervised Machine Learning techniques
- Supervised ML techniques based on neural networks
- Supervised ML technique with deep models for TabularData

2 Data Acquisition

The data acquisition phase involves collecting the data that needs to be analyzed. Data can be acquired in various ways, including static acquisition, which was used in this project.

The dataset used in this project consists of a single csv file with 252175 rows and 91 columns. One of the columns represents the year of publication of the song, ranging from 1956 to 2009. All other columns contain floating-point numbers related to the audio track, making the entire dataset continuous. For this reason, regression models were used to solve the problem.

Before proceeding with further operations, the dataset was analyzed to check for missing or duplicate values. It was found that there are no missing values and only 52 duplicate rows. Since they are very few and mainly related to the most represented classes, we decided to remove them.