Predicting Song Popularity

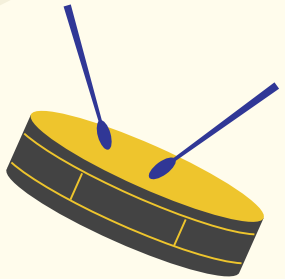
Classification Project

Presented by : **Mai & Basma**

Table of Contents

01

Introduction



About Spotify

World's largest music streaming service providers

Founded on 23 April 2006

165 million paying subscribers

About Spotify

- Sweden

Back Story

Situation

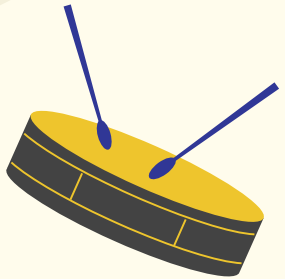
- Meeting with Spotify
- Deploy User friendly interface

Solution

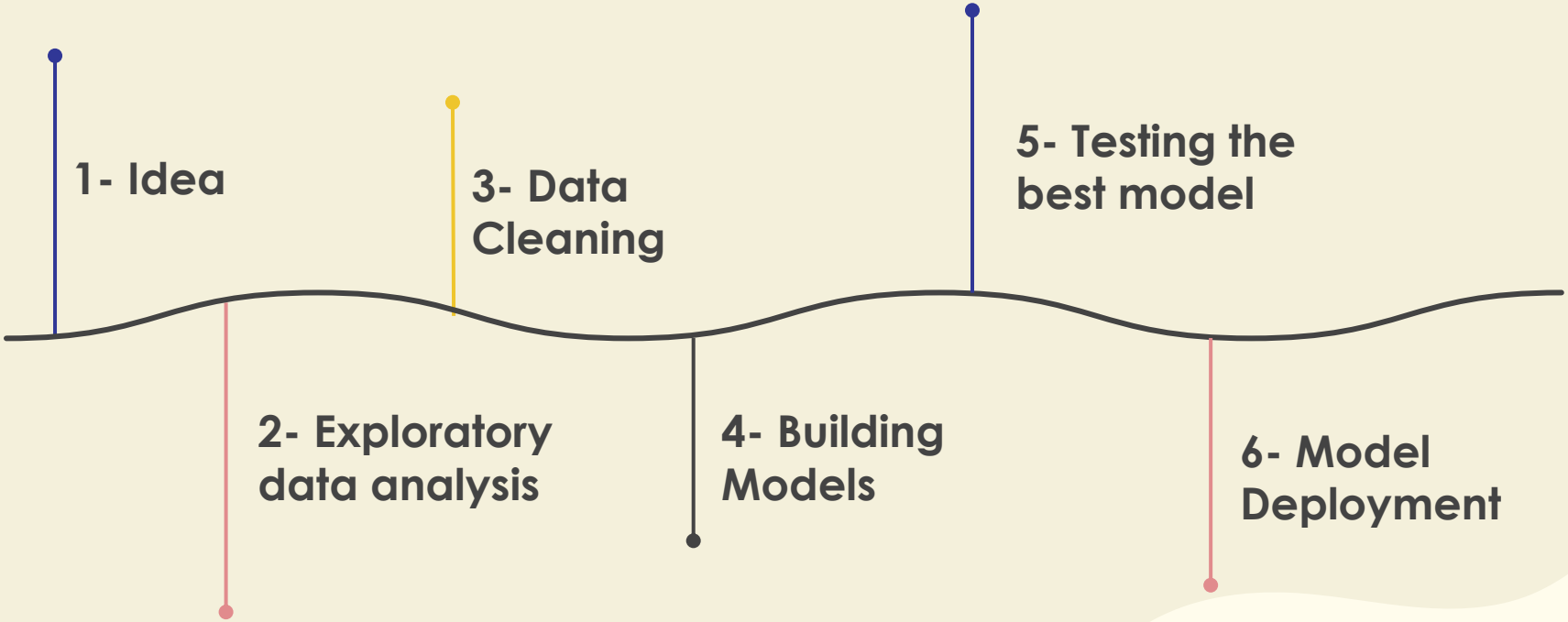
Building a classification model to predict whether a song will be a hit or not

02

Workflow

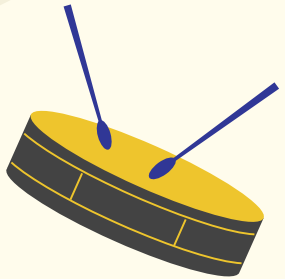


Workflow

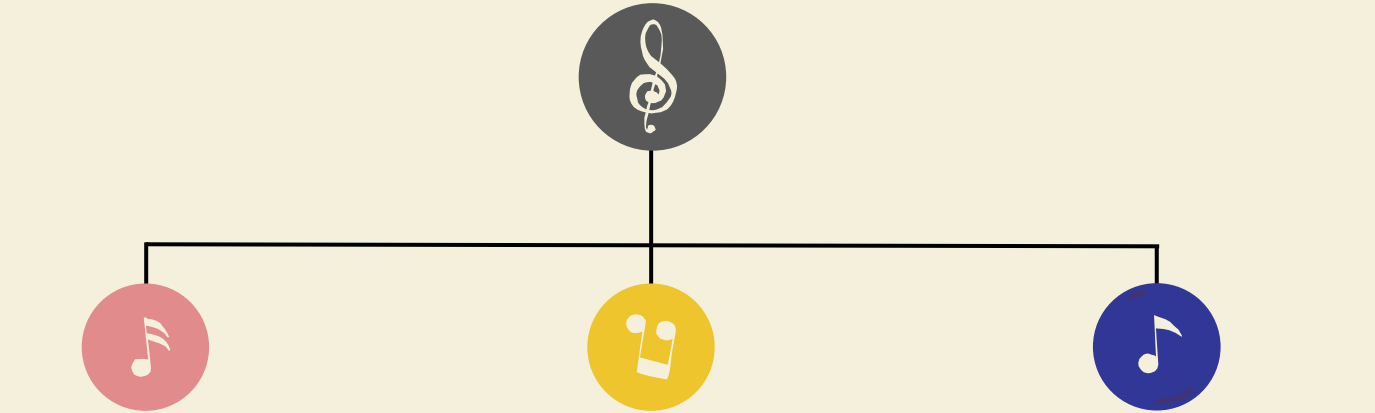

1- Idea
2- Exploratory data analysis
3- Data Cleaning
4- Building Models
5- Testing the best model
6- Model Deployment

03

Data & Design

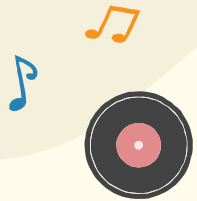


Dataset

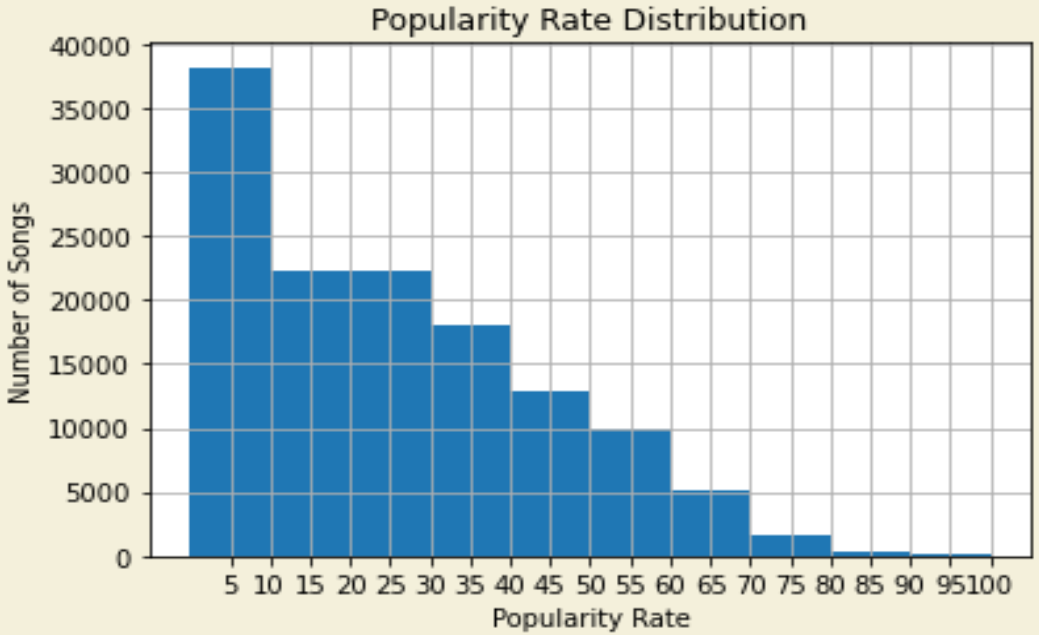

130,662
Observations
16
Features
Target
Popular, not popular

Data Cleaning

- Checking for null values
- Duplicate rows
- Renaming columns
- Creating a new column
- Drop unnecessary columns



Target Transformation

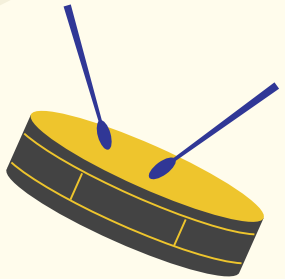


1 Popular

0 Not Popular

04

Tools



Tools

Numpy,Pandas

Data Cleaning and Manipulation

Matplotlib,Seaborn

Visualization

SKlearn

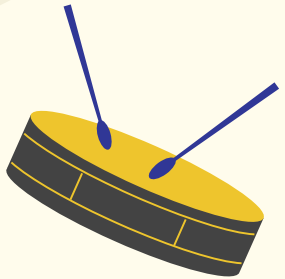
Model Bulliding

Flask

Web Deployment

05

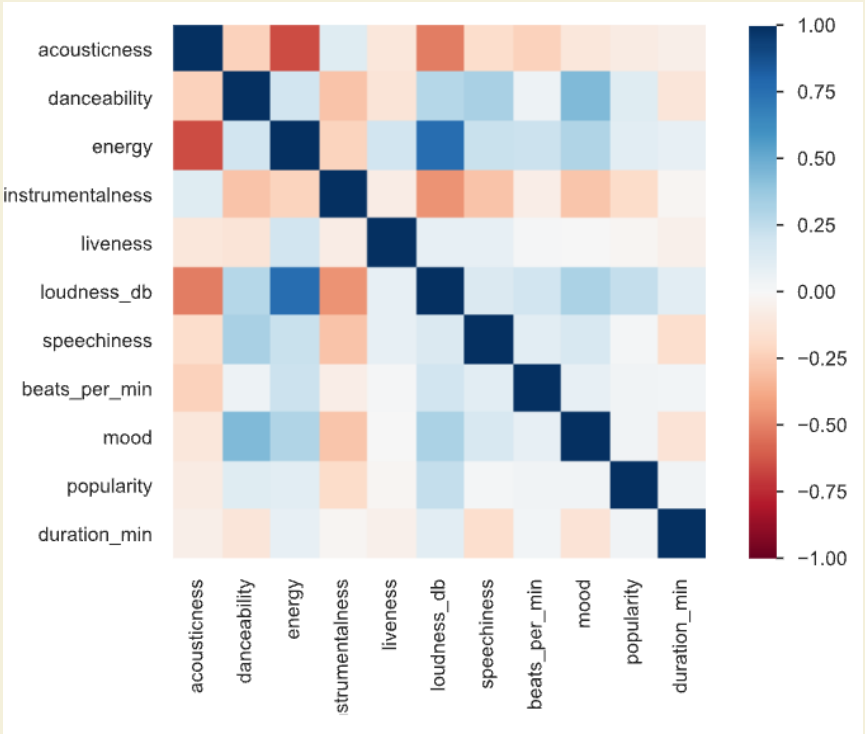
Data Visualization



Data Visualization

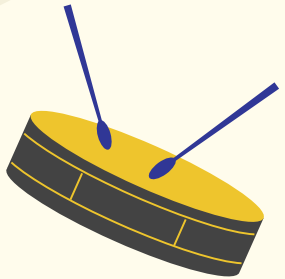
The most repeted artist names in dataset

Data Visualization

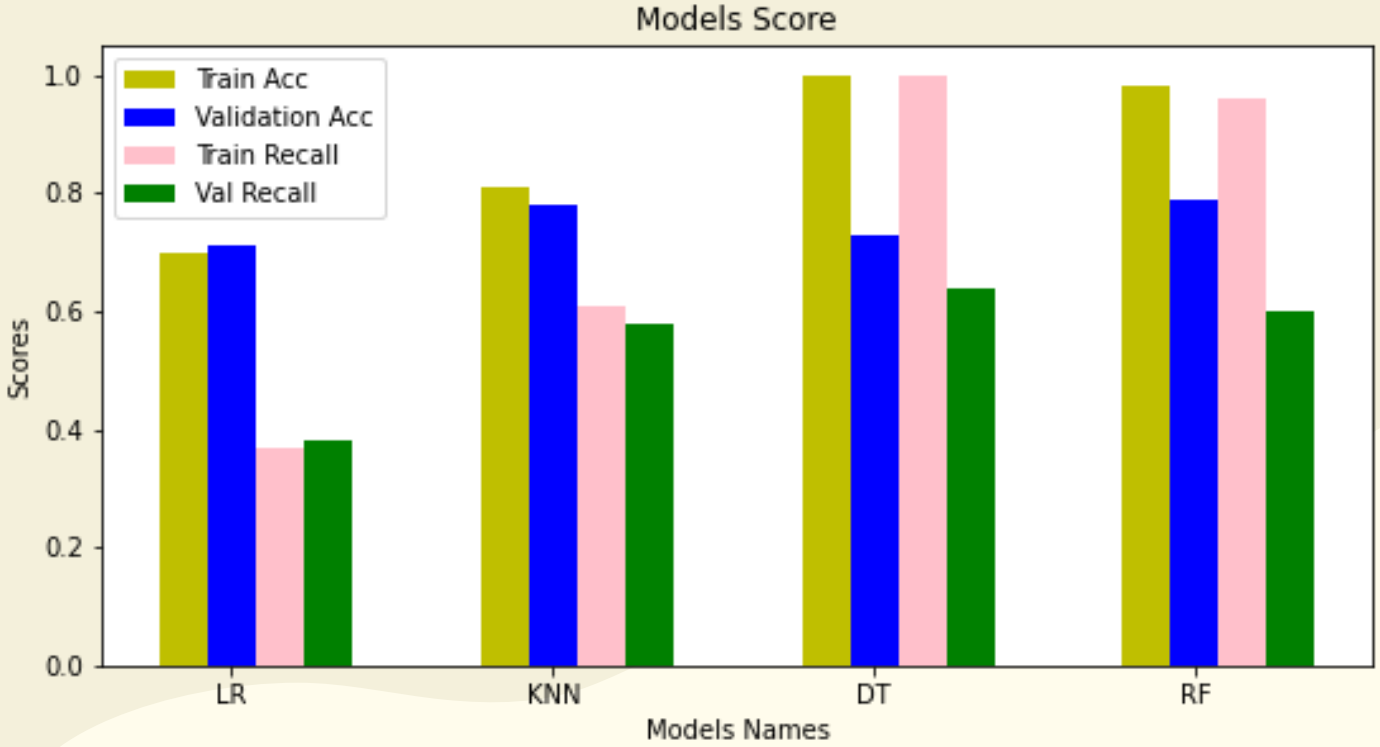


06

Models



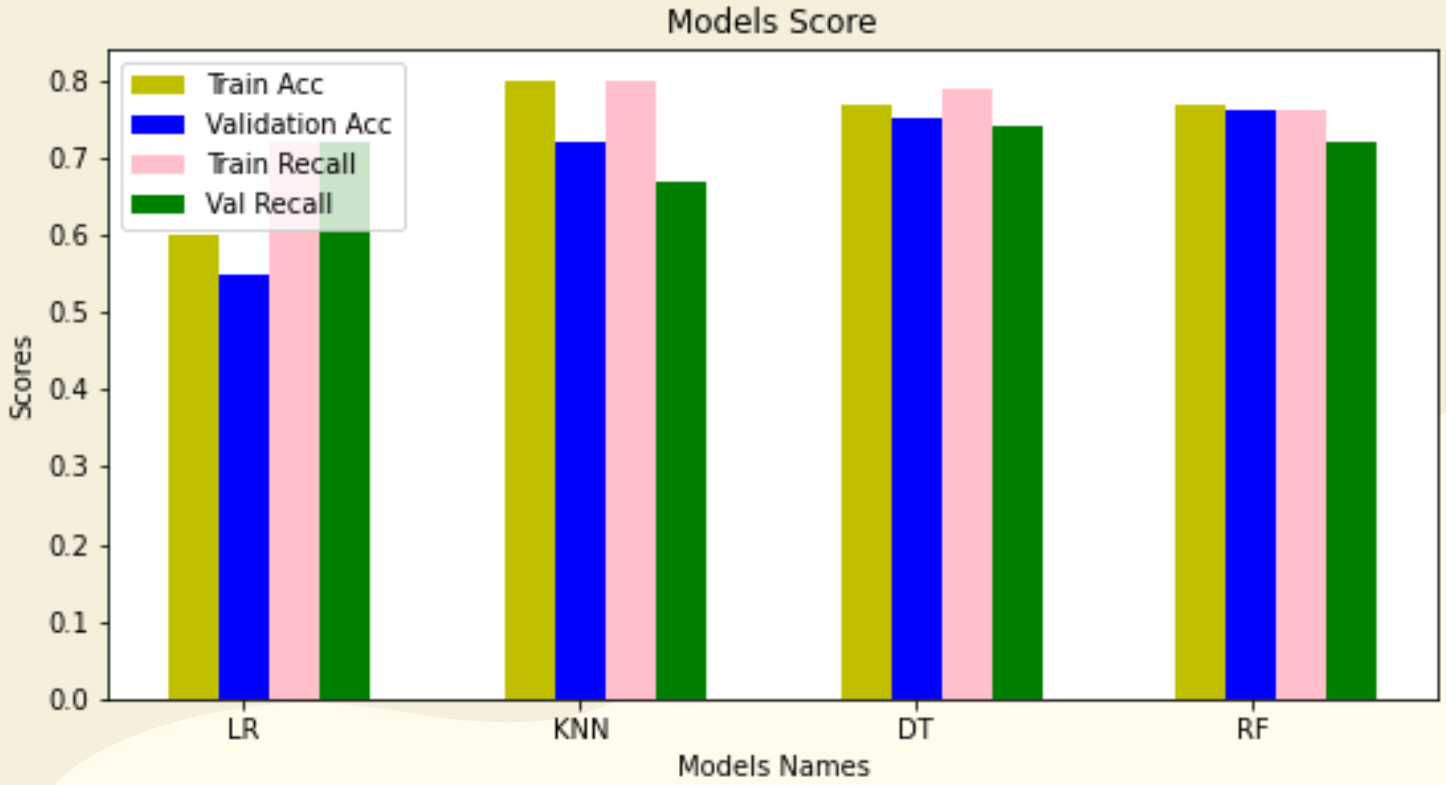
Bassline Models Metrics



Experiments

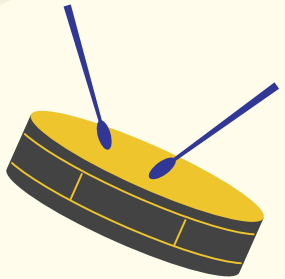
I. Class Imbalance
II. Features Importance
III. Ensembling
IV. Stacking
V. GridSearch
VI. Removing Features and Tuning Hyper Paramerters

Metrics

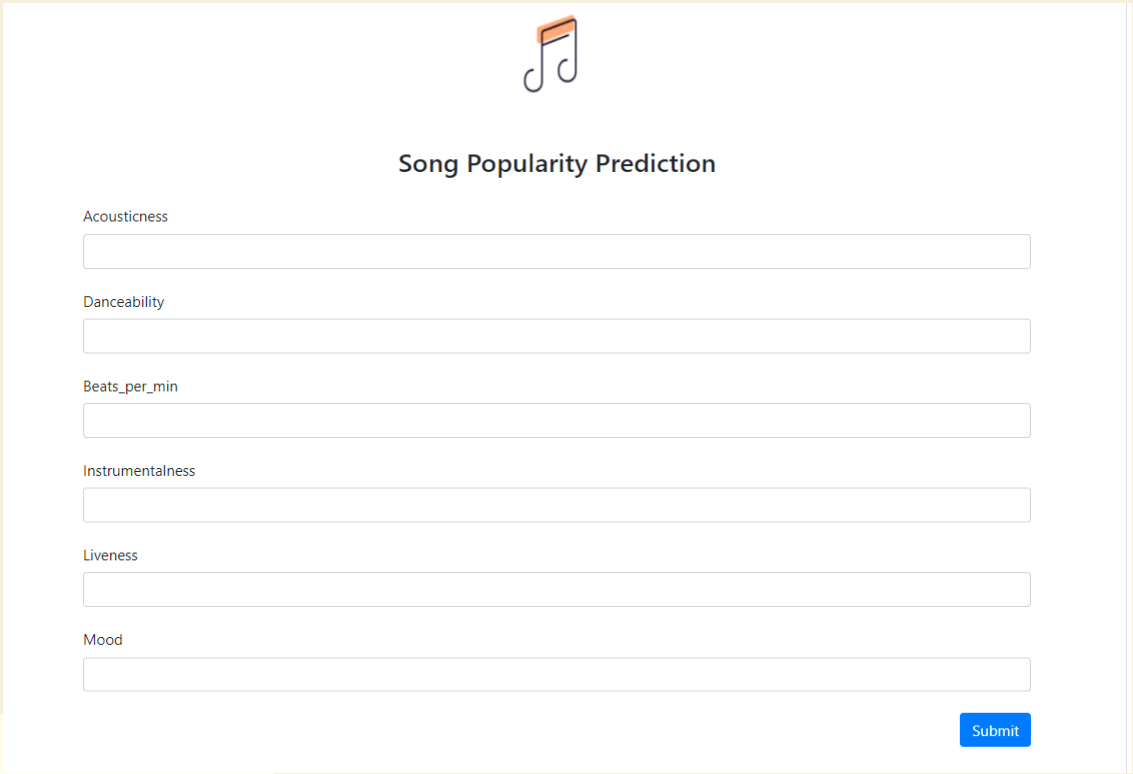

Models Score
Train Acc
Validation Acc
Train Recall
Val Recall
Scores
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
LR
KNN
DT
RF
Models Names

07

Model Deployment

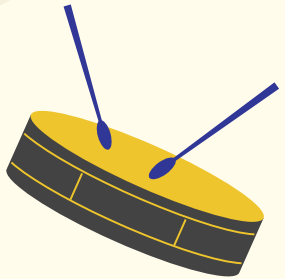


Model Deployment



08

Conclusion

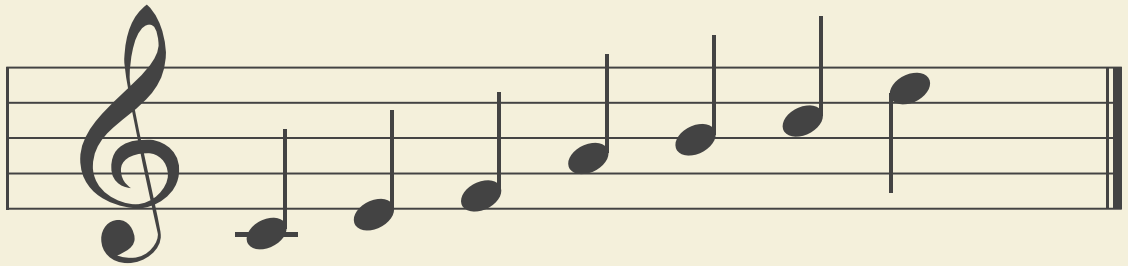


Conclusion

- This project demonstrated the possibility of predicting music popularity
- Decision Trees provided the best predictions on the validation model, with an Accuracy score of **0.75** / **0.74** Recall

Future work

- Optimizing the model
- Explore additional features



Thank you