



# Harmonizing Hit Predictions: Leveraging Audio Features to Forecast **Spotify** Top Songs.

Team A05

Dylan Kakkanad

Priyanka Chaudhari

Saachi Dholakia

Sahasra Konkala



Introduction



About Data



Pre Processing



Model



Limitations



Upgrade



Sahasra ▼

## Importance of This Analysis:

1. Strategic Release Tactics
2. Optimized Investment
3. Global Promotion
4. Enhanced User Recommendations

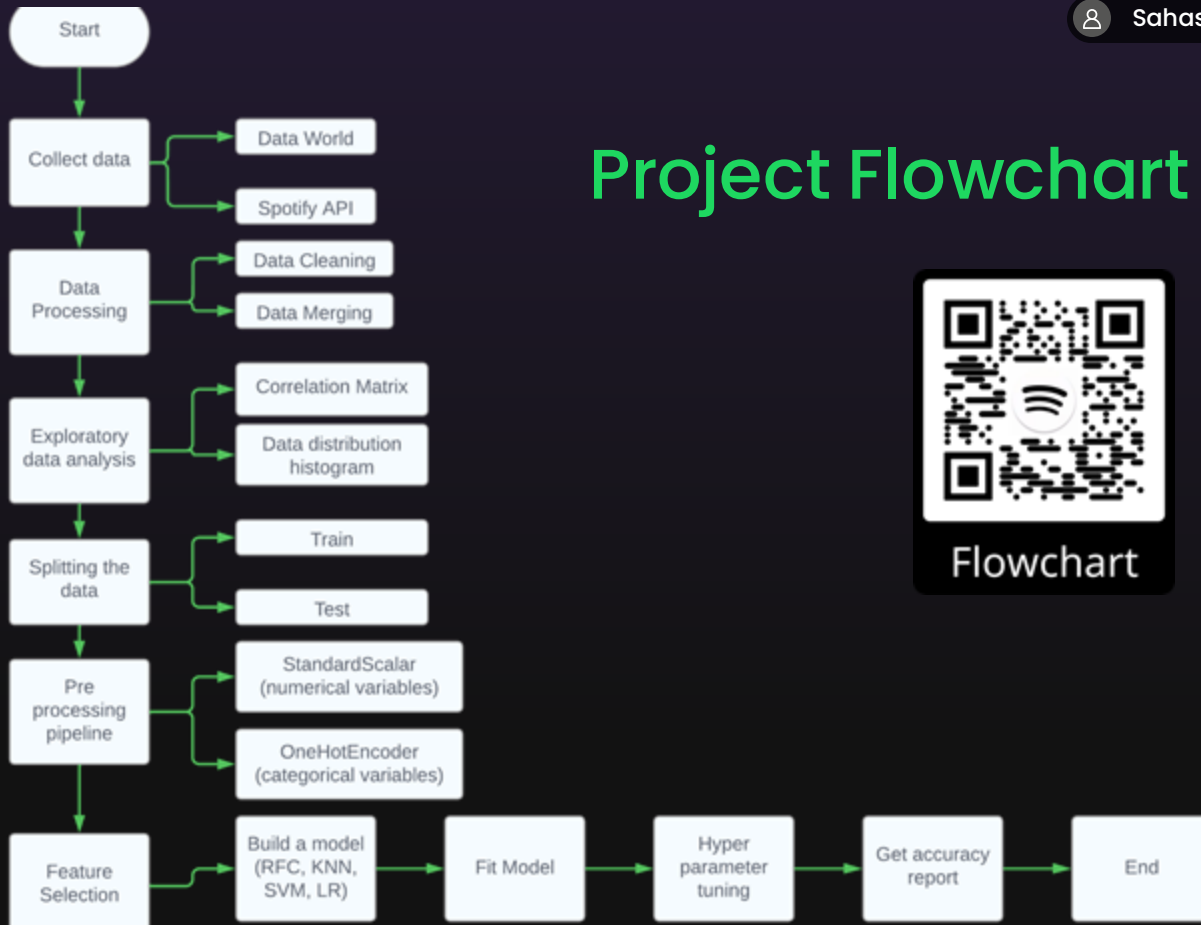
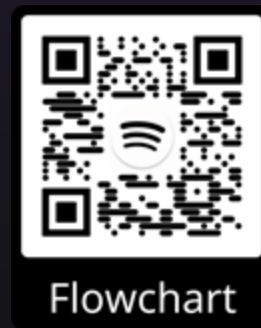
Play

Follow





# Project Flowchart





Introduction



About Data



Pre Processing




Model



Limitations



 Dylan K ▼

# Data Description



Shape: 92233 Rows x 20 Columns

## Overview of the Data

- **Highly** Imbalanced on the target variable
- **Low** Correlation between target variables and predictor variables.

## Balance the Data

- Using the Spotify API we pulled various playlists of **popular** songs.
- We then extracted audio features of these songs by a second API call.
- By adding more **popular** songs into our dataset we hoped to balance the data.



Introduction



About Data



Pre Processing




Model



Limitations

# Data Cleaning

 Dylan K ▼

- Popularity was a continuous score ranging from 0-100. We chose a certain cutoff threshold to classify a song as popular or not popular.
- Removed tracks with really low popularity score (0-10).
- From the artists columns we create 2 new columns: Multi-artists and No of artists.
- Drop Duplicate rows based on track\_ids.
- Set a cut-off threshold of 55 to classify songs equal to and above as popular and below as not popular

# Distribution of Popularity

Dylan K ▼



Introduction



About Data



Pre Processing

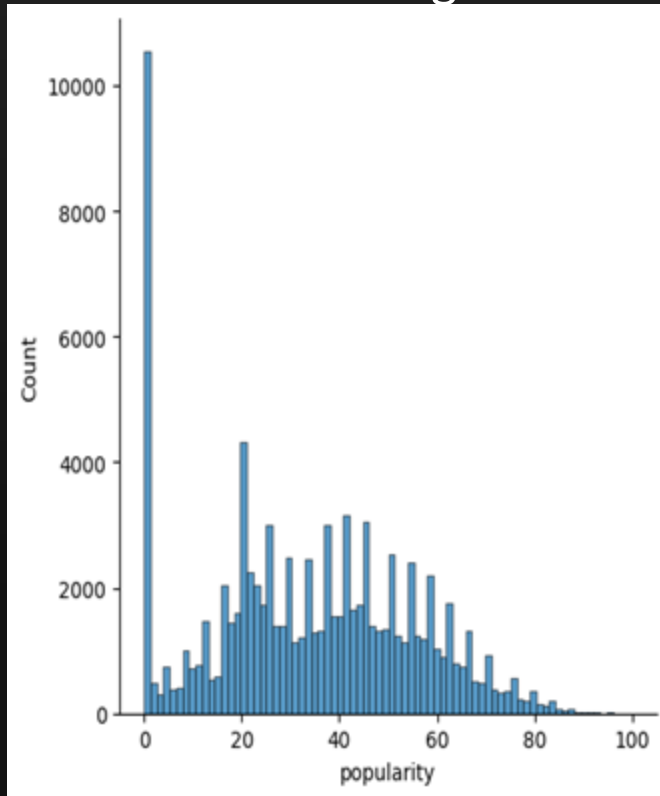


Model

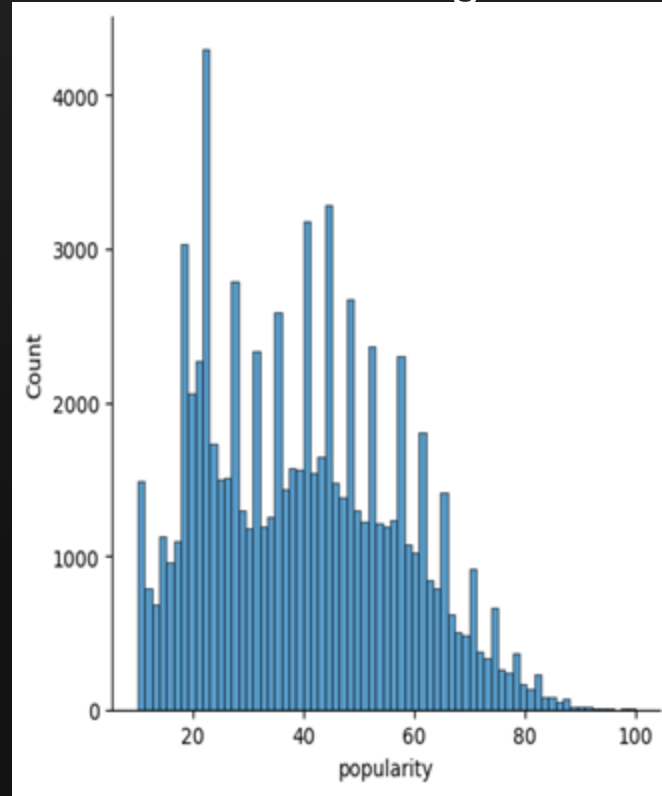


Limitations

## Before Cleaning



## After Cleaning





Introduction



About Data



Pre Processing



Model

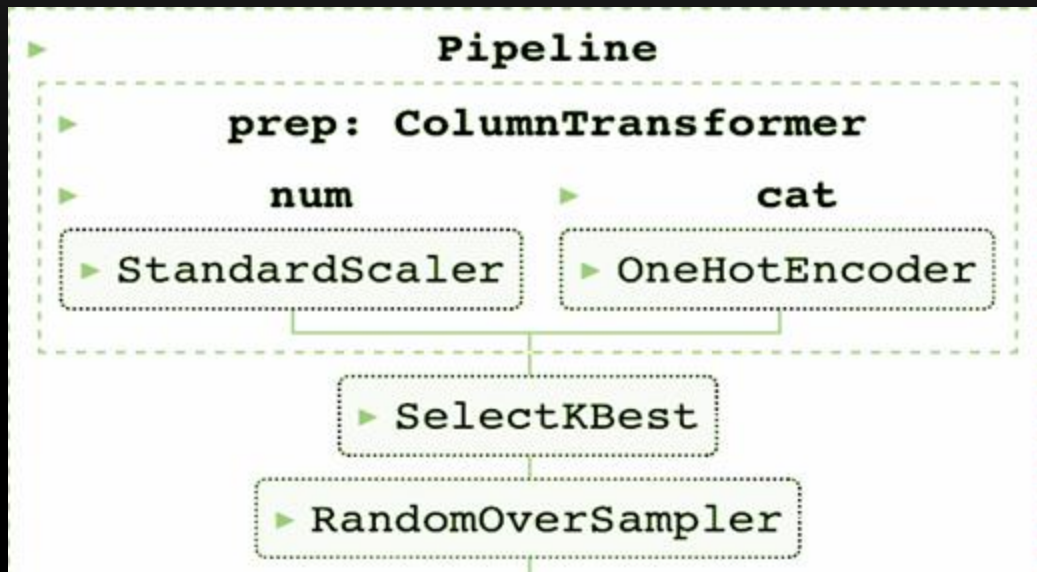


Limitations

[Upgrade](#)

Priyanka ▼

# Pipeline





Introduction



About Data



Pre Processing




Model



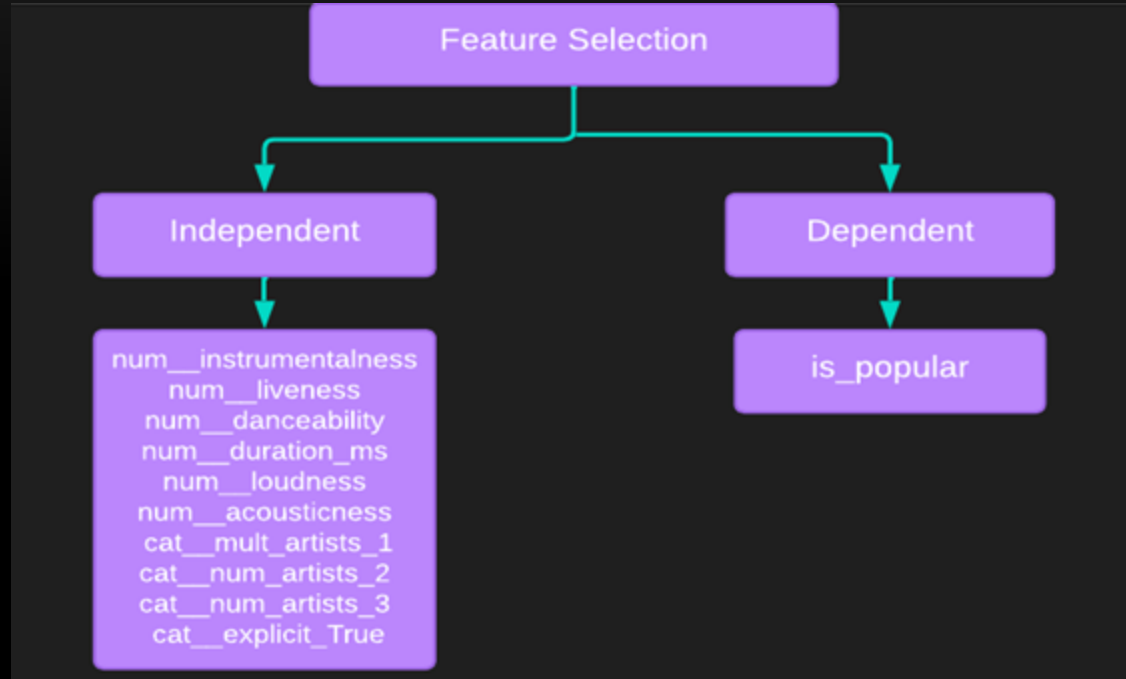
Limitations



Upgrade

 Priyanka ▼

# Feature Selection







Introduction



About Data



Pre Processing



Model



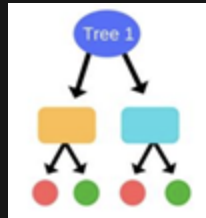
Limitations



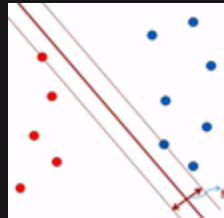
Upgrade

Priyanka ▼

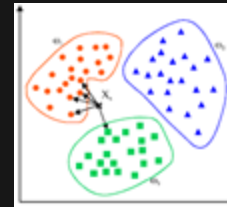
# Models



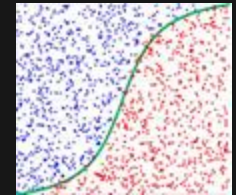
RFC



SVC



KNN



Log Reg



Introduction



About Data



Pre Processing



Model



Limitations

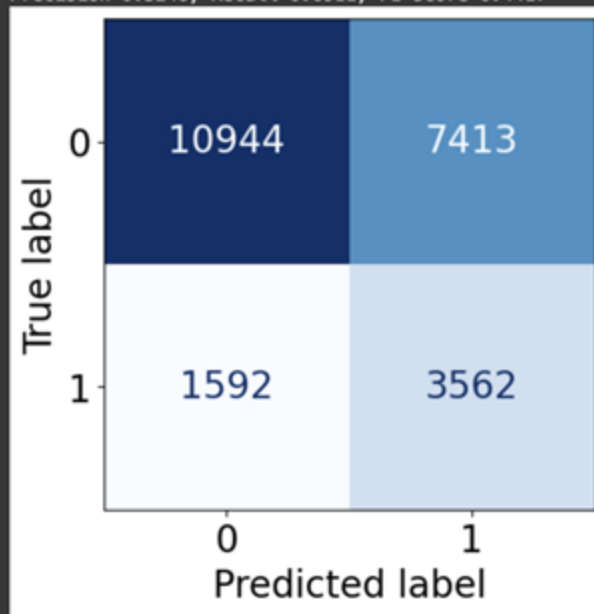


Upgrade

Priyanka ▼

# Random Forest Classifier Model

Accuracy=0.6170, Balanced Accuracy=0.6436  
Precision=0.3246, Recall=0.6911, F1-score=0.4417



- Balanced Accuracy after Hyperparameter tuning
- Instances of model correctly identifying songs and misclassification
- Challenges



Introduction



About Data



Pre Processing



Model



Limitations



Upgrade



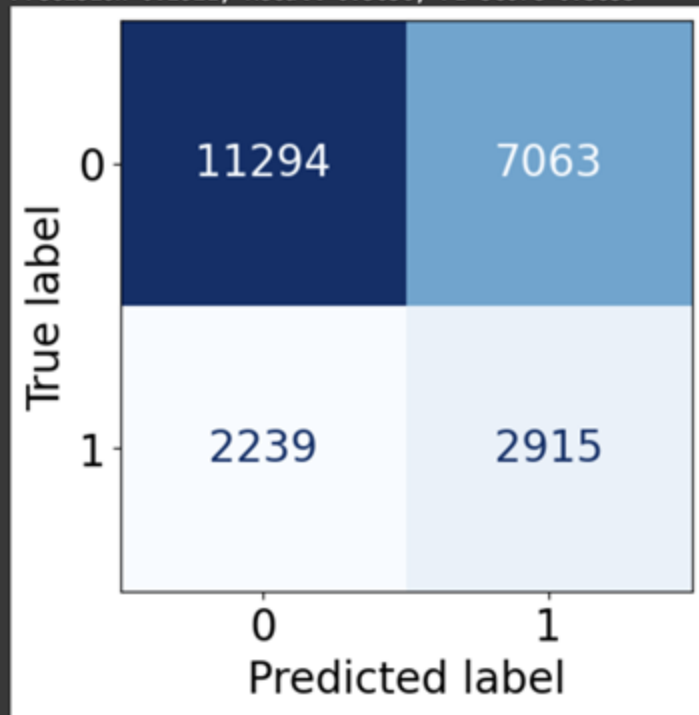
Saachi



# KNN Model

Accuracy=0.6044, Balanced Accuracy=0.5904

Precision=0.2921, Recall=0.5656, F1-score=0.3853



- Principle of Proximity
- Before using Hyper Parameter Tuning
- 3 types of Hyper parameters used:
  - I. Grid Search
  - II. Halving Grid Search
  - III. Randomized Search



Introduction



About Data



Pre Processing



Model



Limitations



Upgrade

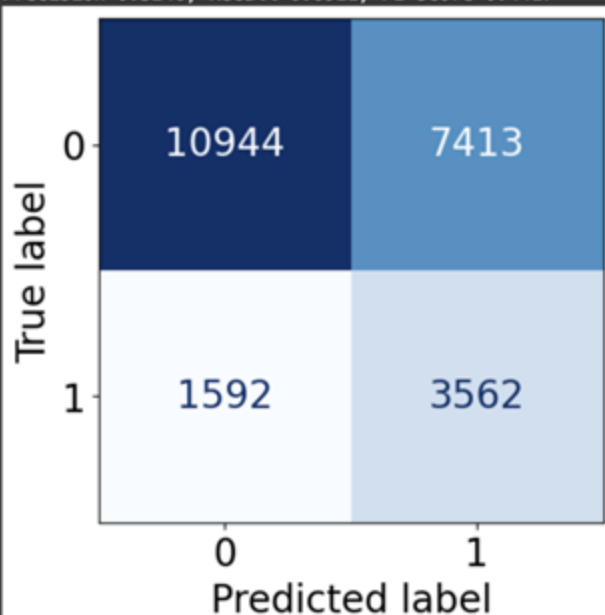
Saachi



# Hyper Parameter Tuning

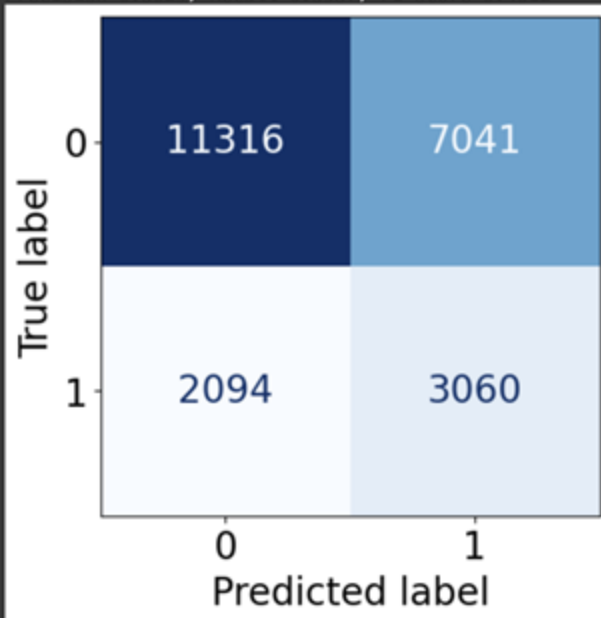
## RFC – Halving Grid Search

Accuracy=0.6170, Balanced Accuracy=0.6436  
Precision=0.3246, Recall=0.6911, F1-score=0.4417



## KNN – Halving Grid Search

Accuracy=0.6115, Balanced Accuracy=0.6051  
Precision=0.3029, Recall=0.5937, F1-score=0.4012





Introduction



About Data



Pre Processing



Model



Limitations



Upgrade

Saachi



# Model Scores



#	Models	Accuracy	Balanced Accuracy	Hyper Tuned Balanced Accuracy
1	Random Forest Classifier	77.32%	58.56%	64.36%
2	K-Nearest Neighbour	60.44%	59.04%	60.51%
3	Logistic Regression	57.22%	60%	60%
4	Support Vector Classifier	55.89%	63.45%	63.45%



Welcome



About Us



Our Teams




Agenda



Limitations



Upgrade

 John Doe ▼

# Limitations

01

Weak correlation of variables.

02

The absence of genre data (Herremans, 2014) limits insights.

03

Exclusion of external factors.

# Thank You

Link to google colab

[https://colab.research.google.com/drive/17shSZtfZwrFMzsSofDNQWH9G\\_cW\\_xOcv?usp=sharing](https://colab.research.google.com/drive/17shSZtfZwrFMzsSofDNQWH9G_cW_xOcv?usp=sharing)



Team 05

Harmonizing Hit Predictions



0:23

-3:25