



Spotify Hit Songs Classification

Prepared By:
Taef al khaless
Sara Hawi



Tools



About The Data

- **Spotify API**

25k+ songs

- **Billboard Top 100 Kaggle + API**

30k+ Rows (1958 - 2021)

- **Target**

1 (Hit) , 0 (Not Hit)

25K+ Row and 16 Feature



About The Features



Data Cleaning

Artist

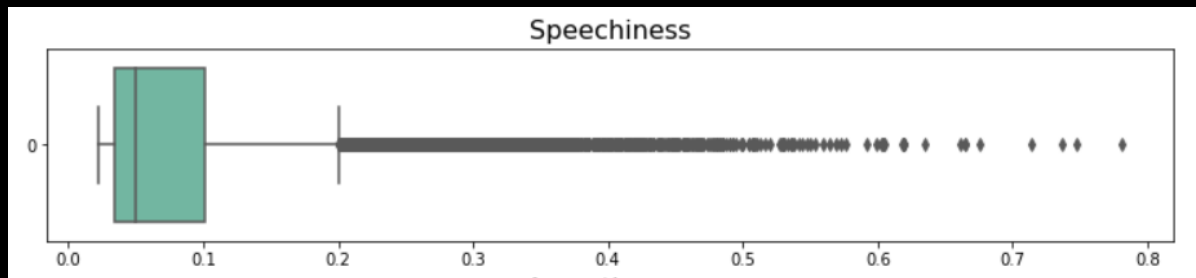
Unify the format

Duplicates

By Song's ID

Outliers

Box plot



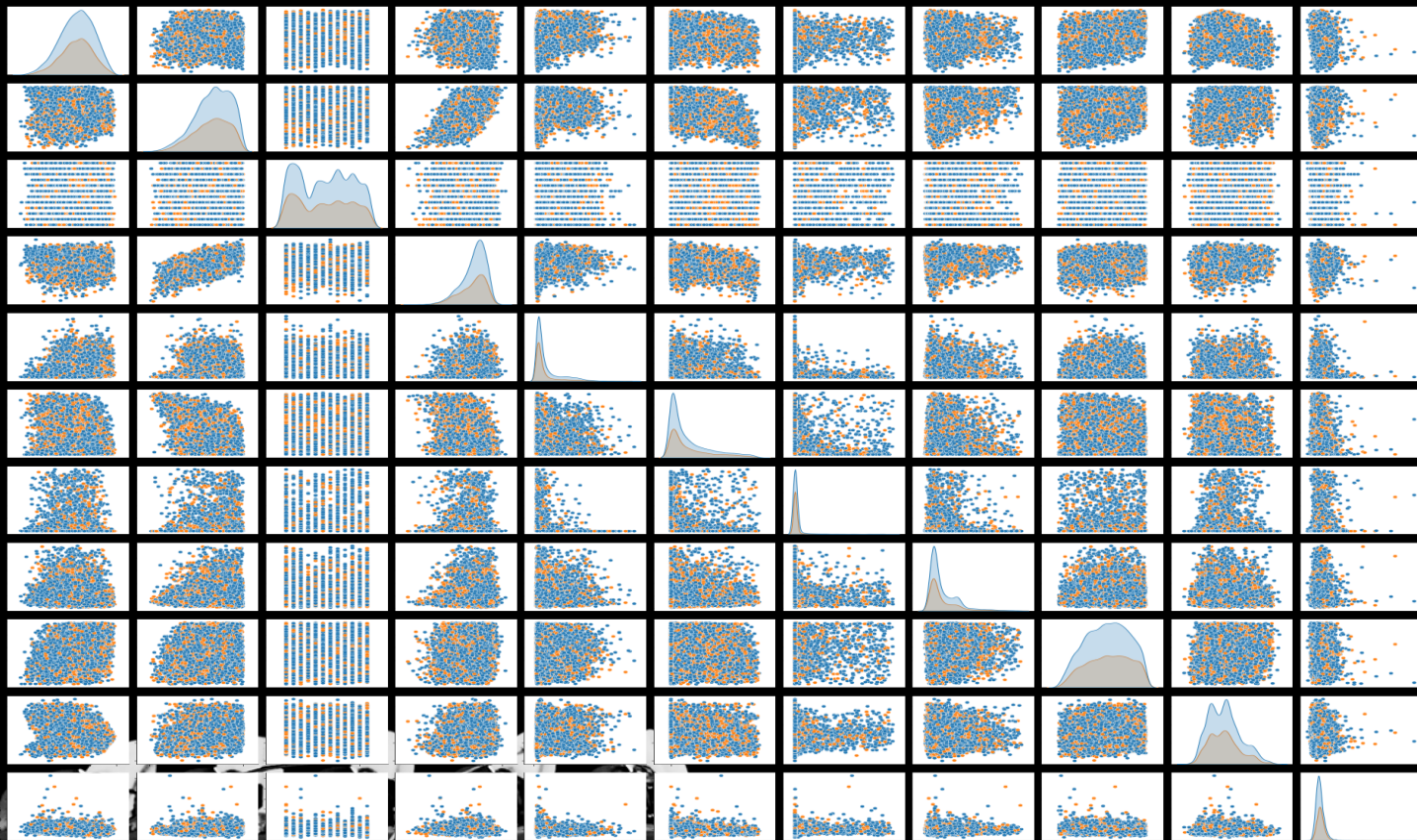
Feature Engineering



Encode and Add the Artists feature

- What if the Song has more than one artist ?
Each artist will be encoded separately

Pair Plot



Handling Imbalanced Data

Hits (+ve class)	40%
Non-Hits (-ve class)	60%

01

Over Sampling

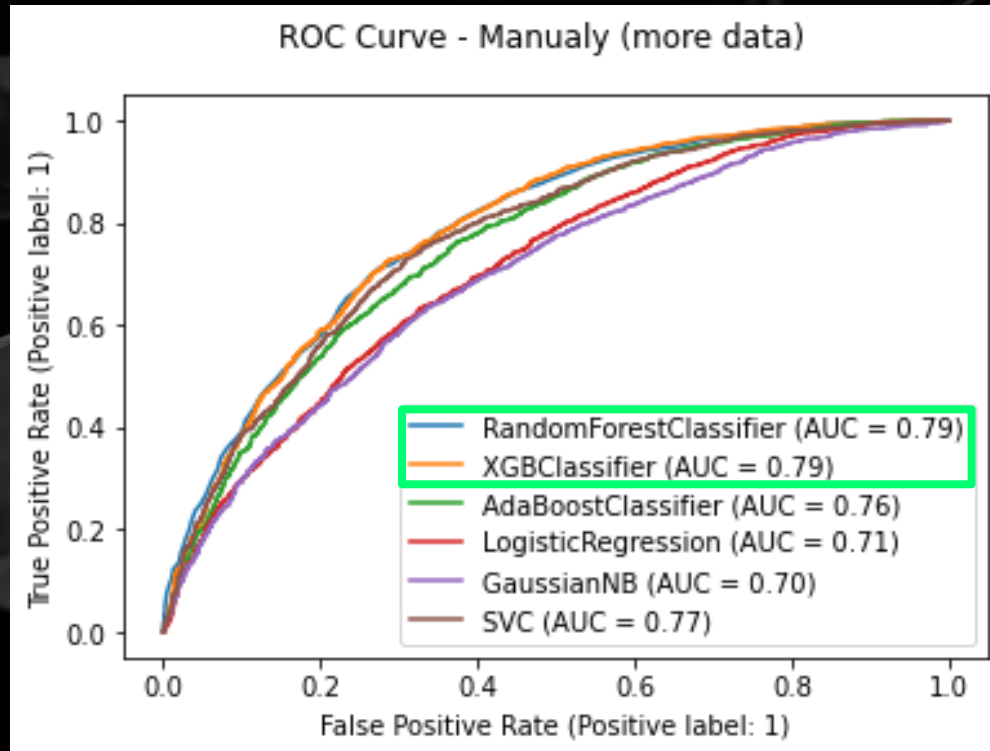
- Random Over Sampler
- SMOTE

02

Under Sampling

- Random Under Sampler
- Manually

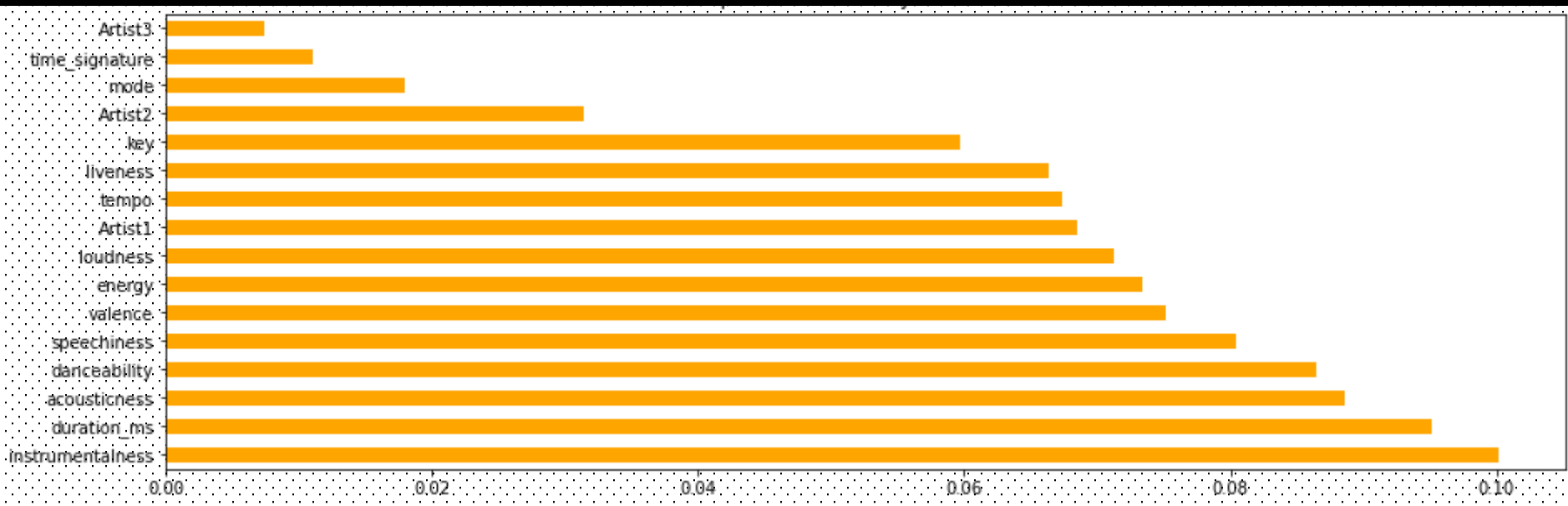
Under Sampling



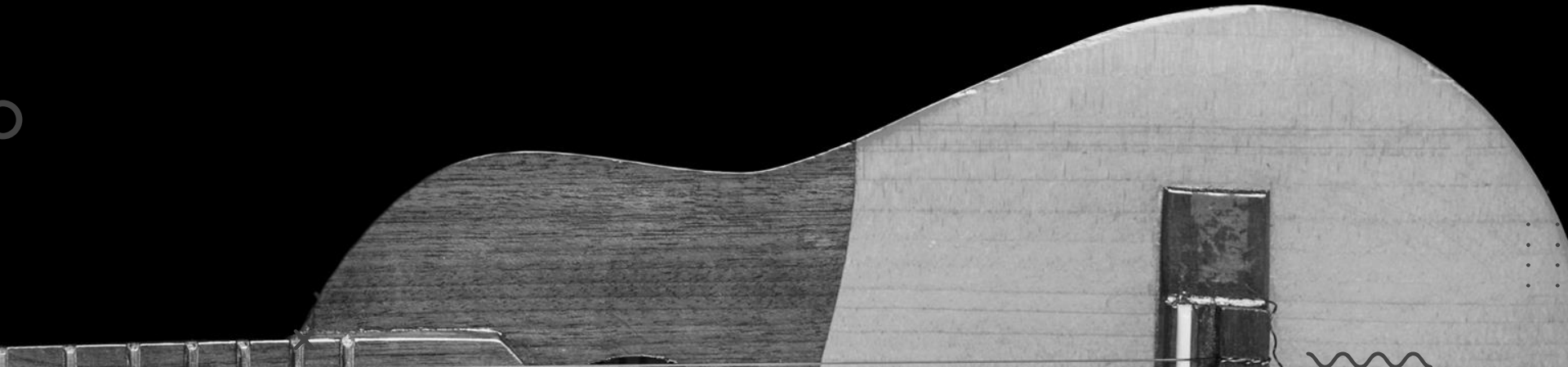
Performance Metrics

	Random Forest	XGBoost Classifier
Accuracy	72%	70%
Recall	74%	68%
Precision	72%	75%
Average F1 Score	72%	70%

Feature Importance



Conclusions & Recommendations





THANKS!