

PROJECT 4

LEVERAGING MACHINE LEARNING FOR MUSIC ANALYTICS

MINERVA FRANCO
NOVEMBER 5, 2024

AGENDA

- **Problem**
- **Data Cleaning**
- **Feature Engineering & Selection**
- **Model Implementation**
- **Predictions and Evaluations**
- **Importance Features Analysis**



THE CHALLENGE OF PREDICTING SONG POPULARITY IN THE PRE-RELEASE STAGE

- Music industry faces uncertainty in identifying potential hits.
- Traditional methods rely on subjective opinions and limited data.
- Pre-release prediction can optimize marketing and resource allocation



PROBLEM & DATASET CONDITIONS

- The aim of Project 4 is to develop a predictive model for estimating the potential popularity of new songs prior to their release on Spotify.
- This is a supervised regression problem because I trained the model using identified pre-release features.
- Dataframe: 23 columns, 1556 rows, no nulls, categorical & numerical data types, memory usage: 279.7+ KB
- Dropped 11 Columns & 10 Rows for pre-release irrelevance and empty string contents.
- Converted 11 Columns from Object to Float64

FEATURE ENGINEERING & MODEL IMPLEMENTATION

APPROACH 1= Fail

Determine most frequently occurring genres to then test their relationships to other features.

APPROACH 2= Warm

Decision Tree Regressor Model

Single Train-Test Split. Without Scaling

Target: Popularity

Predictors: Danceability, Energy, Loudness, Speechiness, Acousticness, Liveness, Duration (ms), Valence, Artist Followers (Pre-Release)

RMSE: 10.622110991431144

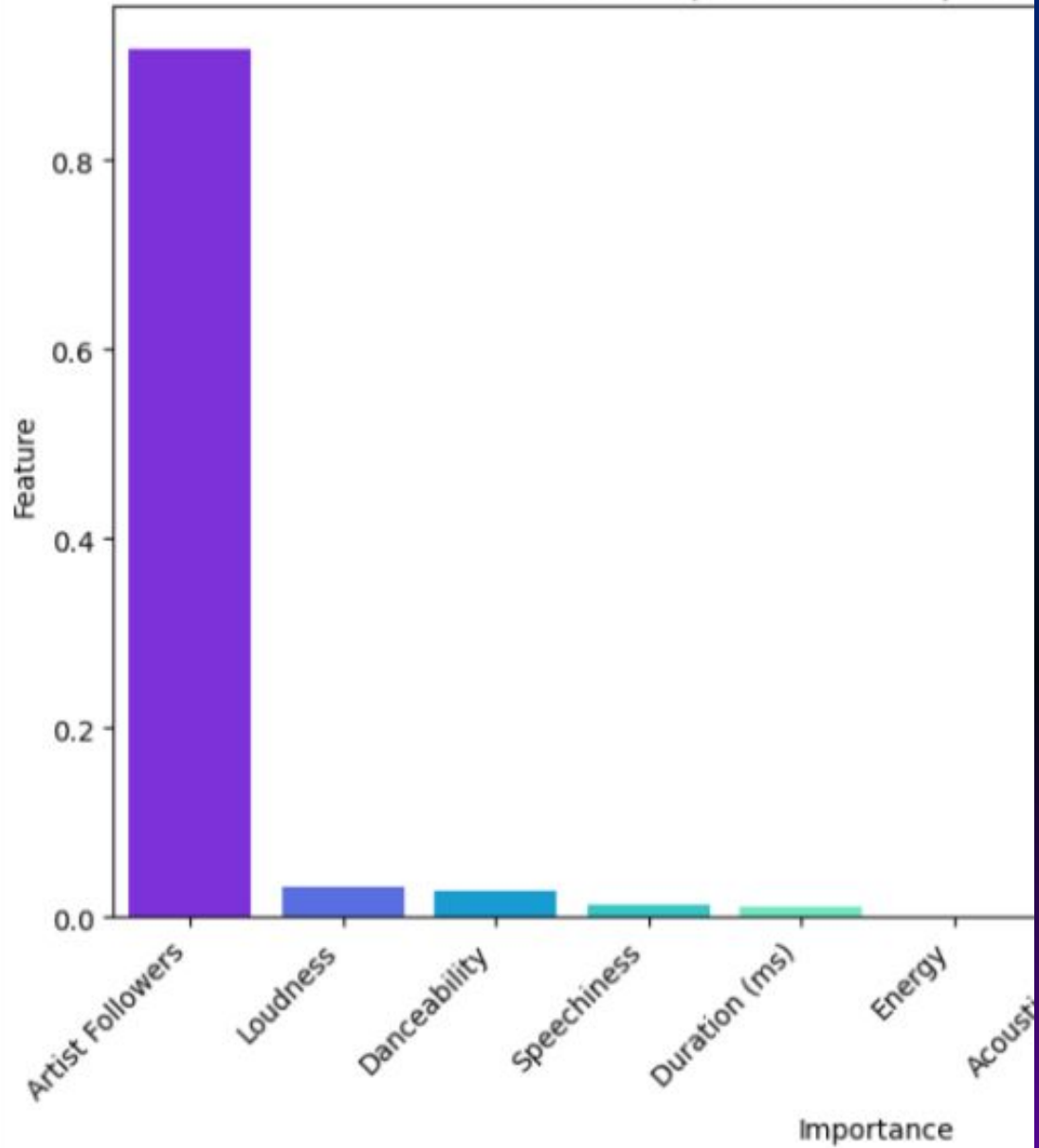
APPROACH 3= BETTER
CROSS-VALIDATION. WITHOUT
SCALING
POPULARITY & PRE-RELEASE
FEATURES

CV RMSE: 10.86

PREDICTIONS,
EVALUATIONS &
FEATURE IMPORTANCE

	Feature	Importance
9	Artist Followers	0.928143
0	Danceability	0.033606
2	Loudness	0.028500
3	Speechiness	0.009751
1	Energy	0.000000
4	Acousticness	0.000000
5	Liveness	0.000000
6	Tempo	0.000000
7	Duration (ms)	0.000000
8	Valence	0.000000

Feature Importance for Popularity



ARTIST FOLLOWERS
WAS BEST
PREDICTOR



THE POWER OF PERCEPTION