

MUSIC POPULARITY PREDICTOR

Predictive Modelling using Ridge Regression

INTRODUCTION

- In today's music industry, streaming platforms generate massive amounts of data about songs and listener preferences.
- Predicting a song's popularity helps artists, producers, and streaming services make data-driven decisions.
- This project focuses on predicting the popularity of songs using various audio features provided by Spotify.
- By analyzing audio characteristics like tempo, valence, danceability, and more, we aim to understand what makes a song popular and build a predictive model.

OBJECTIVE

- Goal :
 - Analyze Spotify song features dataset
 - Build a regression model to predict target feature(s)
- Approach :
 - Data exploration
 - Preprocessing
 - Ridge regression with hyperparameter tuning

WHICH DATASET???

- Dataset Name: SpotifyFeatures.csv
- Features:
 - Includes numerical and categorical song features
 - Quick stats shown using `df.describe()`, `df.info()`, `df.shape()`, etc.

TOOLS AND TECHNOLOGIES

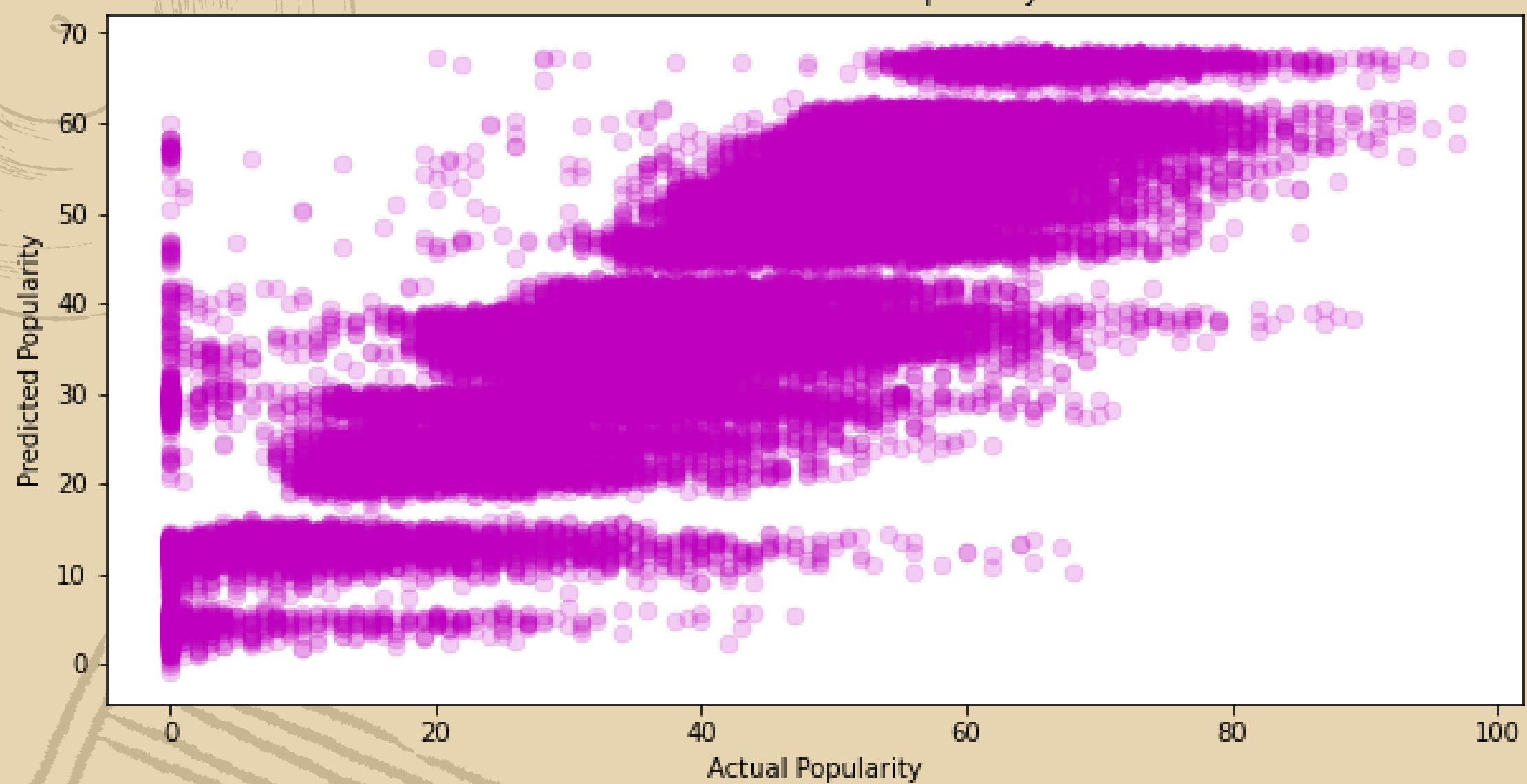


- Libraries Used :
 - pandas, numpy (data manipulation)
 - matplotlib, seaborn (visualisation)
 - scikit-learn (preprocessing, modelling, evaluation)
- Checked :
 - Shape, size, info, head, describe
- Techniques Used for Data Preprocessing :
 - StandardScaler for numerical features
 - OneHotEncoder for categorical features
 - Combined using ColumnTransformer

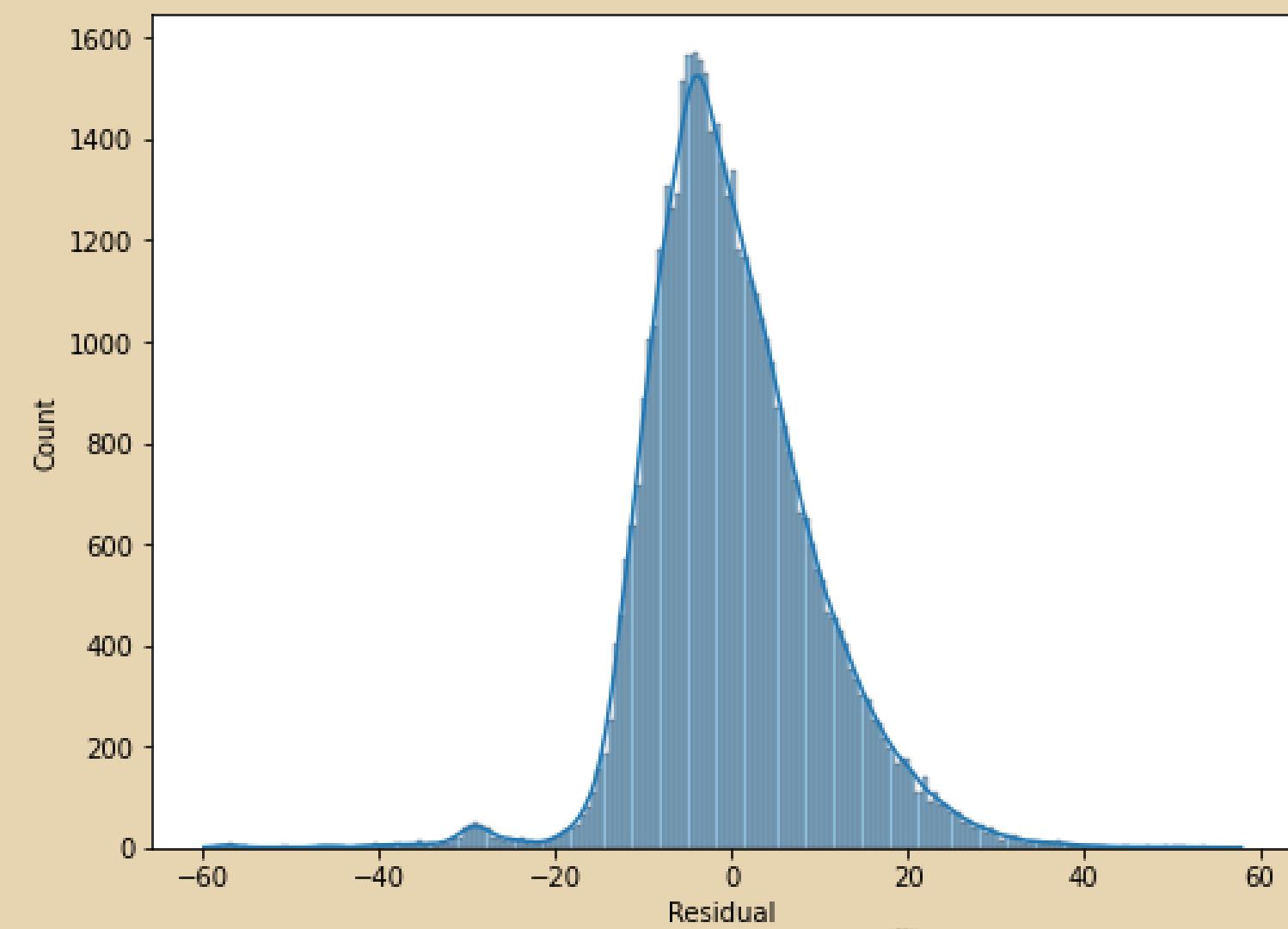
AND RESULTS???

- Finally, we got the following evaluation metrics for the model trained on our test data using Ridge Regression :
 - RMSE : 9.018948133091445
 - R^2 Score : 0.773757365034048
- More effectiveness of Ridge Regression for dataset in comparison to RandomForest and other methods, as the model results lower RMSE score and higher R^2 Score for the dataset than any other model.

Actual vs Predicted Popularity



Residuals Distribution



```
predict_popularity(  
    genre='Pop',  
    key='C',  
    mode='Major',  
    time_signature='4/4',  
    acousticness=0.3,  
    danceability=0.7,  
    duration_ms=210000,  
    energy=0.8,  
    instrumentalness=0.0,  
    liveness=0.1,  
    loudness=-5.0,  
    speechiness=0.05,  
    tempo=120.0,  
    valence=0.9  
)
```

Predicted Popularity Score: 66.42

```
simplified_predict_popularity(  
    genre='Pop',  
    mood='Sad',  
    tempo_category='Slow',  
    duration_category='Medium'  
)
```

Predicted Popularity Score for a Sad Pop song (Slow tempo, Medium duration): 67.62

Welcome to the Music Popularity Predictor

```
Enter genre ('Movie', 'R&B', 'A Capella', 'Alternative', 'Country', 'Dance', 'Electronic', 'Anime',  
Enter mood (Happy, Sad, Energetic, Calm): sad  
Enter tempo category (Slow, Medium, Fast): medium  
Enter duration category (Short, Medium, Long): medium  
Predicted Popularity Score: 40.24
```

**THANK YOU
FOR PLAYING!**

