Predicting Song Popularity Using Audio Features from Spotify:

A Comparative Analysis of Machine Learning Models

COLUMBIA UNIVERSITY

Frannie Jin, Wenling Min, Xinwen Zhang, Yongcan Tang

April 25th, 2024

1 Abstract

Our study provides a comprehensive analysis of the factors influencing song popularity on Spotify by using a dataset with 130,633 tracks and various machine learning models. By integrating logistic regression, Support Vector Machines (SVM), and advanced model blending techniques, we have identified crucial audio features such as loudness, danceability, and energy that significantly affect popularity. Our findings highlight the effectiveness of combining multiple regression approaches, including Ridge and LASSO, to address multicollinearity and enhance prediction accuracy. The implementation of non-linear models like Random Forest and Support Vector Regression (SVR) further refines our understanding of complex feature interactions. This research offers valuable insights for artists, producers, and industry stakeholders, providing a robust tool for predicting trends and making strategic marketing plans in the competitive music industry.

Contents

1 Abstract **1**

2 Introduction **3**

3 Data Description **3**

4 Exploratory Data Analysis (EDA) **4**

5 Methodology **7**

6 Analysis **7**

6.1 Multiple Linear Regression (MLR) . . . 7

6.2 Diagnostics . . . 8

6.2.1 Assumptions of MLR . . . 8

6.2.2 Multicollinearity . . . 10

6.3 Ridge Regression . . . 11

6.4 LASSO (Least Absolute Shrinkage and Selection Operator) . . . 12

6.5 Random Forest (RF) . . . 13

6.5.1 RF without applying PCA . . . 13

6.5.2 RF after PCA Application . . . 14

6.6 Support Vector Regression (SVR) . . . 15

7 Results and Discussion **16**

8 Future Recommendations **18**

2 Introduction

The continuous innovation of technology affects our way of listening to music. It is quite crucial to figure out what kind of factors determine the popularity of songs on music streaming platforms. In the past, people only focused on the lyrics and some musical factors to guess which songs would be hit songs. However, music preference for people is a complicated subject. It involves multiple different factors. Recent studies have been using machine learning models, neurophysiological insight, and comprehensive factor analysis to figure out what could make a song popular.

According to Gao (2021), the paper illuminates the key role of specific audio features in relation to the music popularity index, which inspires us to explore the modeling weight of 'loudness' from the feature data. Moreover, for a comparison of the performance of machine learning and traditional linear models in Merritt, Gauri, and Zak (2021), we are inspired to insert logistic regression and Support Vector Machines (SVM) to improve accuracy. Cross-validation techniques will also be used to prevent overfitting.

Building on these foundations, according to the paper by Ge et al. (2021), "Popularity Prediction of Music Based on Factor Extraction and Model Blending," the introduction of a sophisticated blend of PCA and machine learning techniques can help prevent large variables and high dimensions due to too many factors affecting the popularity of songs. In addition, conducting forecasts based on Model Blending make use of the advantages of each model, conduct forecast based on model blending makes use of the advantages of each model, showed great outperforms compared to the individual predictors.

Our study intends to investigate what factors most influence song popularity and employ multiple methodologies to forecast a song's level of popularity. We are creating models with features from the song tracks and the artists, which may help streaming platforms, artists, and others in the music industry predict which songs will be successful. In an effort to comprehend music on a deeper level, we are also researching genre prediction. We want to use these efforts to discover new, effective approaches to developing marketing plans and promoting musical creation.

3 Data Description

The dataset consists of 130,633 tracks with various audio features. Each row represents a song with details about 13 numerical features that describe its audio characteristics:

Acousticness: A measure from 0.0 to 1.0 indicates the confidence that the track is acoustic.

Danceability: Evaluates how suitable a track is for dancing based on elements like tempo, rhythm stability, beat strength, and overall regularity.

Duration: The length of the track in milliseconds.

Energy: A measure from 0.0 to 1.0 that represents a perceptual measure of intensity and activity.

Instrumentalness: Indicates the likelihood that a track contains no vocal content.

Key: The key the track is in, using standard pitch class notation.

Liveness: Detects the presence of an audience in the recording.

Loudness: The overall loudness of the track in decibels.

Mode: Indicates the modality (major or minor) of a track.

Speechiness: Detects the presence of spoken words in a track.

Tempo: The overall estimated tempo of a track in beats per minute.

Time Signature: An estimated time signature of the track.

Valence: A measure describing the musical positiveness conveyed by a track.

Additionally, each song has a popularity score, which reflects the current popularity based on the number of plays. This metric is dynamic and can change over time.

This set of features allows for robust analysis and model building for predicting track popularity based on its audio features. This dataset is aimed at understanding how different audio features influence listener preference and track success on the platform.

4 Exploratory Data Analysis (EDA)

Firstly, we analyzed the dataset by doing Exploratory Data Analysis (EDA).

After reviewing the dataset, we are certain that it is well-structured and does not have issues such as missing values or duplicates. It allows us to proceed directly to do EDA without the preliminary step of data cleaning.

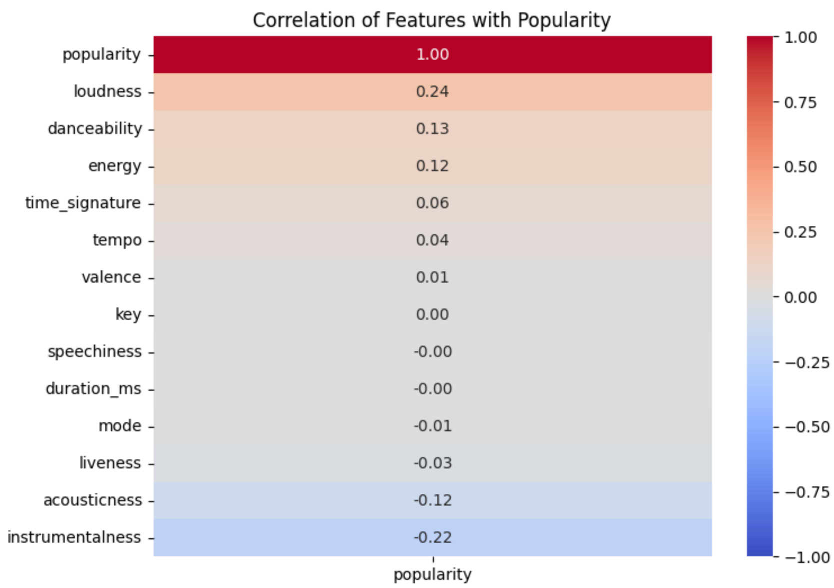

Figure 1: Correlation of Features with Popularity

Figure 1 shows that certain audio features have a low to moderate correlation with the popularity of the tracks. Specifically, loudness has the most significant positive correlation. It indicates that louder tracks tend to be more popular. Loudness is followed by danceability and energy, which shows that vibrant and rhythmic tracks are liked by listeners. On the other hand, instrumentalness has the most negative correlation, which indicates that tracks with fewer vocals and more instrumental elements are less popular.

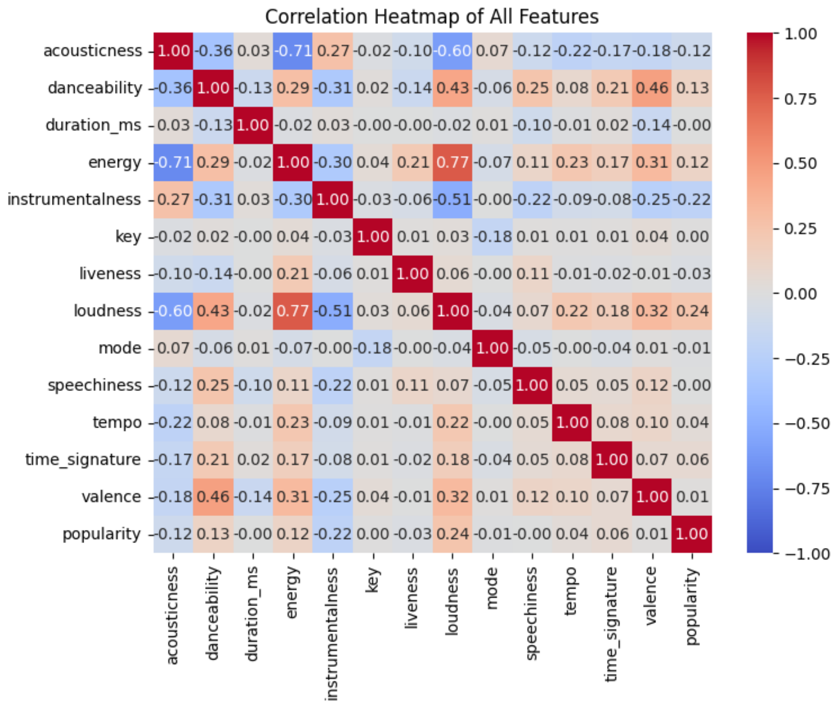

Figure 2: Correlation Heatmap of All Features

From Figure 2, we can know the interrelationships between various audio features of music tracks. The correlation matrix further explores relationships between the features. Acousticness and energy are strongly negatively correlated, indicating that acoustic tracks tend to be less energetic. Loudness and energy show a strong positive correlation, showing that louder tracks are generally more energetic. Danceability shows a positive correlation with valence, meaning that more danceable tracks are often more cheerful. Instrumentalness has a negative correlation with loudness and danceability, indicating that instrumental tracks are often quieter and less danceable.

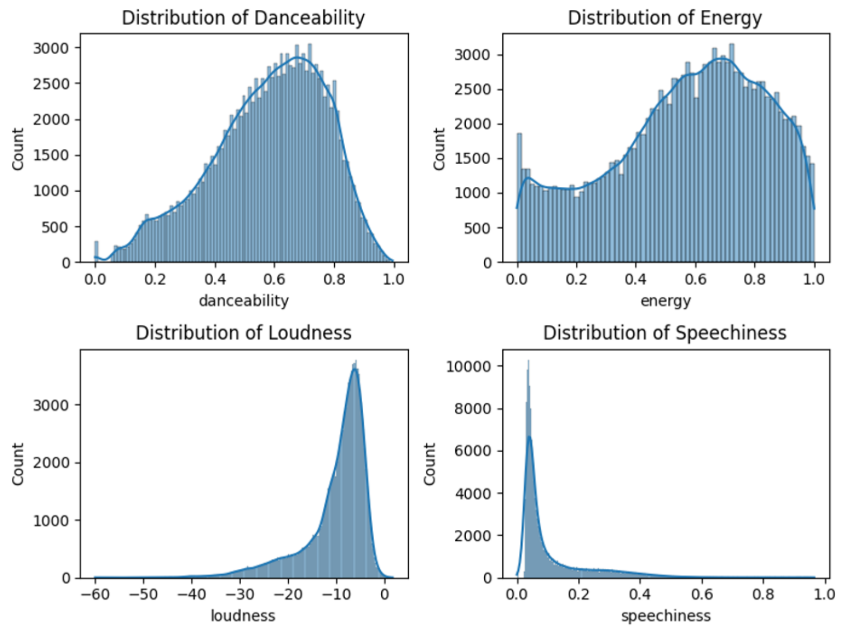

Figure 3: Histograms of Musical Feature Distributions

Figure 3 indicates the distributions of some features. Danceability is normally distributed, showing a balanced mix of tracks in terms of danceability with the dataset. The energy feature also follows a roughly normal distribution with a slight right skew, indicating that there are more tracks with higher energy. Loudness shows a left-skewed distribution with most tracks having loudness values between -10 and 0 dB. Speechiness is heavily right-skewed with a peak close to 0, implying that most tracks have few spoken words.

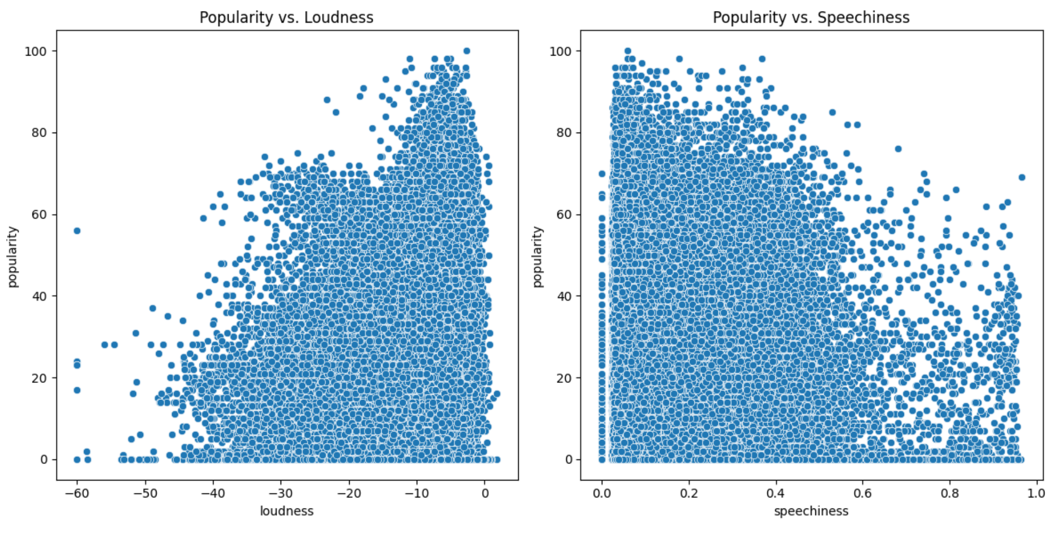

Figure 4: Scatter Plots of Track Popularity Against Loudness and Speechiness

We can conclude from Figure 4 that the relationship between the popularity of music tracks and two audio features: loudness and speechiness. The plot on the left indicates a trend where tracks with higher loudness tend to have higher popularity. There is a dense clustering of points towards the louder end of the spectrum, showing that louder tracks are often popular. The plot on the right does not show a clear trend between speechiness and popularity. Most tracks have low speechiness. The spread of points across levels of popularity is fairly even for tracks with low speechiness, showing speechiness may not be a strong predictor of a track's popularity.

5 Methodology

We aim to predict the popularity of songs based on 13 audio features extracted from Spotify's API. The study involves four distinct machine learning models, Ridge Regression, LASSO (Least Absolute Shrinkage and Selection Operator), Random Forest (RF), and Support Vector Regression (SVR). The first two are good alternatives of addressing multicollinearity, and RF and SVR can adequntely avoid the influence bround by the outliers. Each model's performance will be evaluated using metrics such as R-squared, Mean Squared Error (MSE), and cross-validation scores.

To start with, we use the Multiple Linear Regression (MLR) as a baseline, and diagnostics and VIF analysis are followed to check any abnormality that may affect the accuracy and robustness. For the MLR, Ridge, and LASSO models, feature selection and regularization techniques will be applied to enhance model performance and prevent overfitting. Ridge and LASSO will particularly focus on penalizing the regression model to maintain simplicity and robustness. The Random Forest model will be used to handle nonlinear relationships and interactions between features through an ensemble of decision trees. Lastly, SVR will be employed to model non-linear relationships that might exist between the features and the song's popularity.

6 Analysis

In this section of our study, we embark on a comprehensive analysis with a range of machine learning models. Each subsection is dedicated to a different model or diagnostic test.

6.1 Multiple Linear Regression (MLR)

The OLS regression analysis provides a quantitative framework for understanding the influence of various musical features on a track's popularity. The positive coefficient for danceability suggests that tracks with a higher danceability score are likely to be more popular. Significant coefficients for loudness reiterate findings from EDA, where louder tracks tend to be more popular. The negative association with speechiness could indicate a preference for musicality over lyrical content in popular music. The R-squared value tells us popularity is a complicated outcome not only determined by the audio features but also influenced by other factors such as cultural, social, and economic factors.

Variable	Coef.	Std Err	t	P > $\|t\|$	[0.025	0.975]
const	35.9939	0.710	50.724	0.000	34.603	37.385
danceability	5.5811	0.405	13.766	0.000	4.786	6.376
energy	−6.4946	0.434	−14.970	0.000	−7.345	−5.644
key	−0.0131	0.016	−0.794	0.427	−0.045	0.019
loudness	0.7826	0.017	47.103	0.000	0.750	0.815
mode	−0.2429	0.122	−1.989	0.047	−0.482	−0.004
speechiness	−5.9846	0.502	−11.919	0.000	−6.969	−5.000
acousticness	0.7332	0.252	2.906	0.004	0.239	1.228
instrumentalness	−7.0069	0.197	−35.528	0.000	−7.393	−6.620
liveness	−2.9983	0.370	−8.104	0.000	−3.724	−2.273
valence	−7.0113	0.268	−26.137	0.000	−7.537	−6.486
tempo	−0.0033	0.002	−1.631	0.103	−0.007	0.001
duration_ms	−1.025	4.84	−2.118	0.034	−1.97	−7.66
time_signature	0.7685	0.117	6.552	0.000	0.539	0.998

Table 1: OLS Regression Results

To ensure the robustness of the multiple regression model, verifying the key statistical assumptions to ensure model validity is necessary.

Metric	**Value**
RMSE	18.79
R^2 Score	0.09

Table 2: MLR Performance Metrics

6.2 Diagnostics

In order to ensure the model accurately captures the underlying dynamics of the data, we check these assumptions (Linearity, Independence, Homoscedasticity, Normality, and Absence of Multicollinearity) and address any violations through statistical techniques or model adjustments in the following.

6.2.1 Assumptions of MLR

A preview of the result we have here, heteroscedasticity, non-linearity, and heavy-tailed are detected. We consider trying different regression techniques that are less sensitive to these assumptions. Feature engineering techniques will be included to improve the performance of some models.

Linearity and Homoscedasticity: we use the Residuals Plot to examine the linearity and homoscedasticity. The residuals are on the y-axis, and the predicted popularity is on the x-axis. If the relationship is linear, the points should be randomly around zero, without patterns. For Homoscedasticity, we expect the spread of residuals should be roughly the same across the x-axis. Other wise, it indicates the existance of heteroscedasticity, violating the assumption.

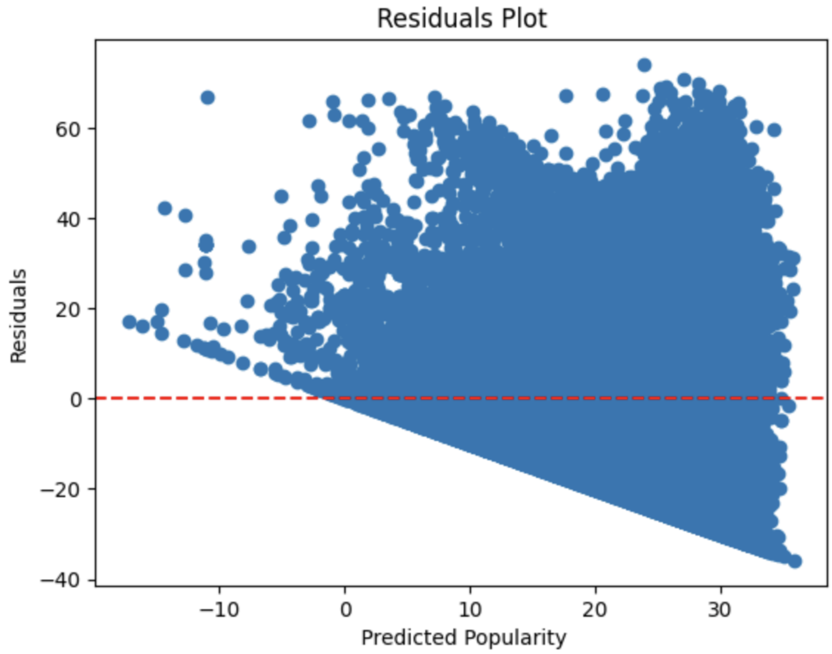

Figure 5: Residuals Plot

The residuals plot displays a clear pattern of increasing spread in residuals as the predicted popularity increases, indicating heteroscedasticity and non-linearity.

Independence: We use the Durbin-Watson statistic, ranging from 0 to 4, to detect the presence of autocorrelation in the residuals. The reported Durbin-Watson statistic is 2.002, which suggests that the independence assumption is likely met (no autocorrelation.)

Normality: Normality of residuals can be assessed with a Q-Q Plot (Quantile-Quantile plot). The x-axis displays the theoretical quantiles of a standard normal distribution, while the y-axis shows the ordered values of the model's residuals. If the residuals are normally distributed, the points will fall approximately along a straight line.

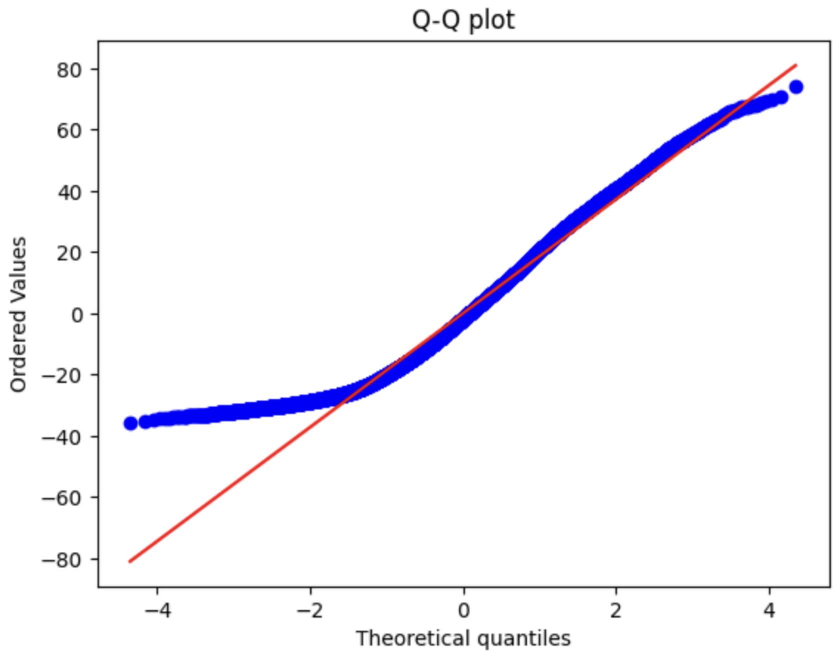

Figure 6: QQ Plot

The points at both ends of the plot deviate significantly from the red line, which is the phenomenon of heavy-tailed, indicating the normality check failed.

6.2.2 Multicollinearity

Feature	**VIF**
danceability	15.59
energy	16.19
key	3.16
loudness	8.49
mode	2.60
speechiness	2.08
acousticness	3.84
instrumentalness	2.05
liveness	2.65
valence	5.49
tempo	15.96
duration_ms	4.02
time_signature	41.25

Table 3: Variance Inflation Factor (VIF) for Model Features

Besides the Normality and Constant Variance assumption, Multicollinearity emerged as a primary concern during the assumption checks. The phenomenon, characterized by high correlations between predictor variables, can significantly impair the interpretability and predictive accuracy of a regression model. Evidence of multicollinearity was indicated by Variance Inflation Factor (VIF) scores. According to Table 3, features such as 'danceability,' 'energy,' and 'time_signature' registered values well above the threshold of 10, highlighting the potential for distorted coefficient estimates.

6.3 Ridge Regression

To address this issue, Ridge and Lasso regressions were employed—both being regularization techniques that introduce a penalty to the model coefficients. These two techniques are particularly adept at reducing multicollinearity and overfitting. Lasso regression also can shrink some coefficients entirely to zero, thus performing feature selection.

Metric	**Value**
Best Alpha	10
Test RMSE	18.80

Table 4: The Output of Ridge Regression via Cross-Validation

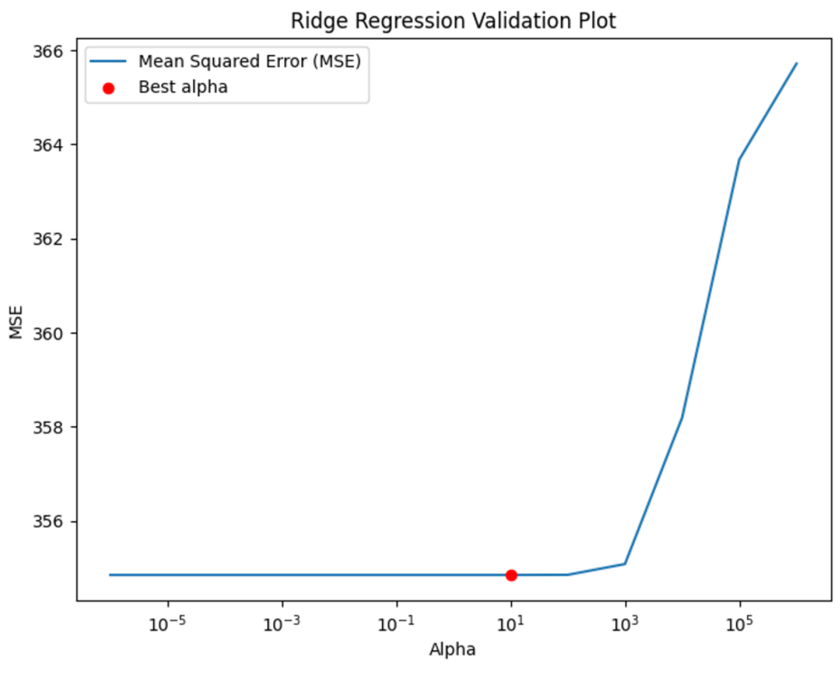

Figure 7: Ridge Regression Validation Plot

According to the above, the Ridge regression model optimized through cross-validation identified a best alpha value of 10. This level of regularization suggested a model less susceptible to multicollinearity, maintaining all features within the analysis. The Root Mean Square Error (RMSE) derived from this model stood at 18.79, which is slightly smaller than the RMSE for the Multiple Linear Regression model previously.

6.4 LASSO (Least Absolute Shrinkage and Selection Operator)

Metric	**Value**
Best Alpha	0.0001
Test RMSE	18.80

Table 5: The Output of LASSO Regression via Cross-Validation

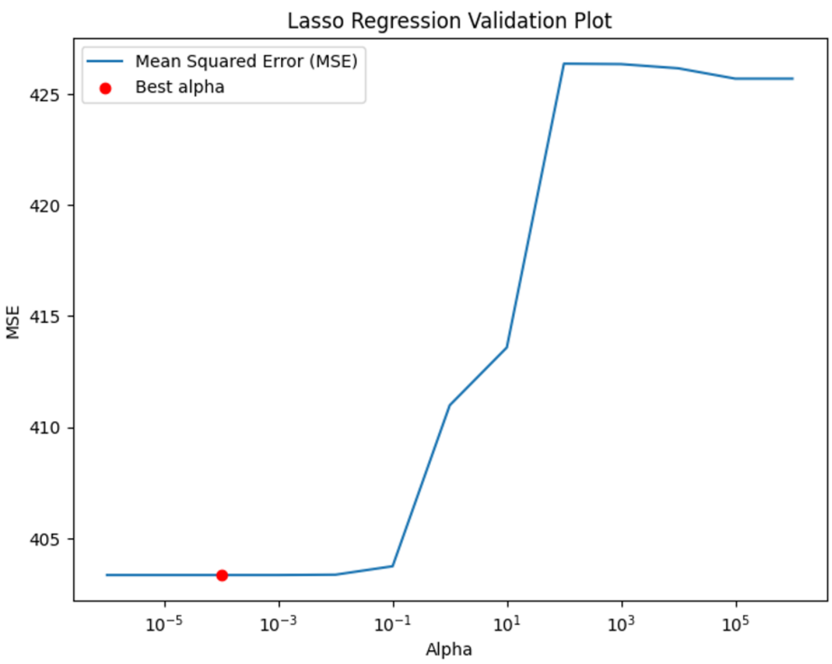

Figure 8: LASSO Regression Validation Plot

Coefficient	5.4e+00	-6.6e+00	-8.5e-03	7.9e-01	-1.7e-01	-6.0e+00
feature	danceability	energy	key	loudness	mode	speechiness
7.5e-01	-7.0e+00	2.9e+00	-7.1e+00	-2.7e-03	-8.4e-07	8.0e-01
acousticness	instrumentalness	liveness	valence	tempo	duration	time_signature

Table 6: The Coefficient of LASSO Regression via Cross-Validation

In contrast, the Lasso regression process adopted a selective approach to feature inclusion. Despite this, the optimal alpha (0.0001) via cross-validation determined through the process resulted in no predictors being excluded from the model, which means there is no coefficient to be shrunk to zero in the plot, although there are some of them are quite small. This indicated that all features had some level of contribution to the model of predicting the popularity of the track. The resulting RMSE for the Lasso model was consistent with that of the Ridge model, affirming comparable efficacy in prediction.

In this analysis, the implementation of regularization techniques is critical to improving the model in the face of multicollinearity challenges. Such analytical rigor allows for a deeper understanding of the dynamics that influence song popularity. This analysis allows us to gain a deeper understanding of the factors that influence a track's popularity, and to get a more rigorous model with fewer assumption violations than Multiple Linear Regression.

6.5 Random Forest (RF)

Since Ridge and LASSO regressions are commonly applied to address multicollinearity in linear regression models, exploring nonlinear models can offer alternative insights, as they are typically less affected by this issue. Random Forest model was adopted here to evaluate its performance and try to capture more complex relationships in the data.

6.5.1 RF without applying PCA

Parameter	**Value**
n_estimators	400
max_features	auto
max_depth	41
min_samples_split	50
min_samples_leaf	1
Test RMSE	18.38

Table 7: RF Parameters via Cross_Validation and Test RMSE

Initially, hyperparameter tuning was conducted using RandomizedSearchCV from scikit-learn, with a focus on adjusting parameters including n_estimators, max_features, max_depth, min_samples_split, and min_samples_leaf. The optimized parameters and the corresponding test RMSE are presented in Table 7. The selection of the best parameter set took 4253.30s with 10 iterations.

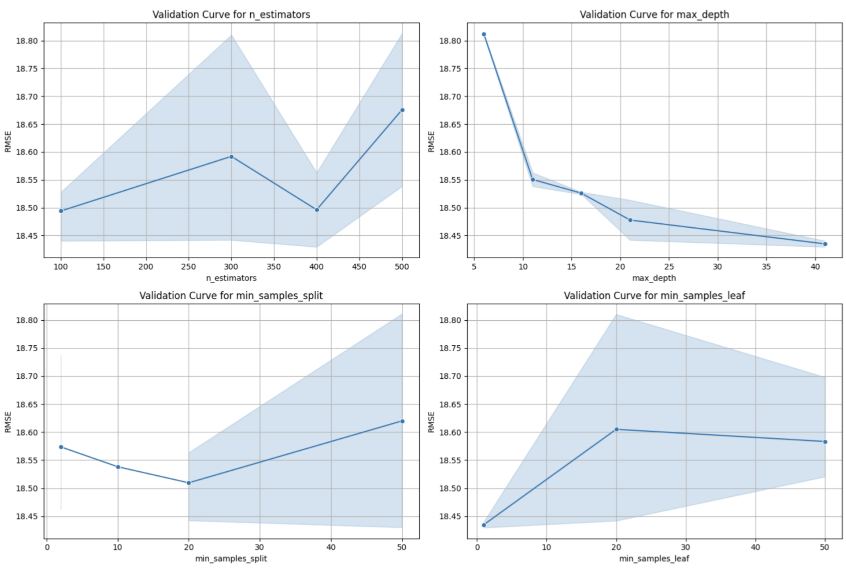

Figure 9: Validation without PCA

From the validation plots in Figure 9, we can see that these optimized parameters, corresponding to the values in the best parameters, consistently achieved the minimum cross-validation RMSE in each subplot, which highlights the reliability and stability of the optimized parameters under different data splits.

Feature	**Importance**
Danceability	0.011994
Energy	0.095576
Key	0.000000
Loudness	0.609980
Mode	0.000000
Speechiness	0.001899
Acousticness	0.040297
Instrumentalness	0.157215
Liveness	0.003103
Valence	0.022634
Tempo	0.005400
Duration_ms	0.051900
Time Signature	0.000000

Table 8: Feature Importance Scores

Moreover, the feature importances from our RandomForest model can be derived to see the significance of each feature in the RandomForest model's decision-making process and its impact on model predictions.

6.5.2 RF after PCA Application

Parameter	**Value**
n_estimators	300
max_features	0.33
max_depth	26
min_samples_split	5
min_samples_leaf	1
Test RMSE	18.56

Table 9: RF After PCA via Cross Validation and Test RMSE

Given the large scale of the dataset, Principal Component Analysis (PCA) was applied to see if the efficiency of Random Forest (RF) model training could be improved by reducing its dimensionality. Through this technique, 8 principal components were selected from the original 13 features. Simultaneously, hyperparameter tuning was still conducted through RandomizedSearchCV from scikit-learn, focusing on adjusting parameters including n_estimators, max_features, max_depth, min_samples_split, and min_samples_leaf across 20 iterations—more than previously used. These results are listed in Table 9. The selection of the best parameter set took 10914.52s this time.

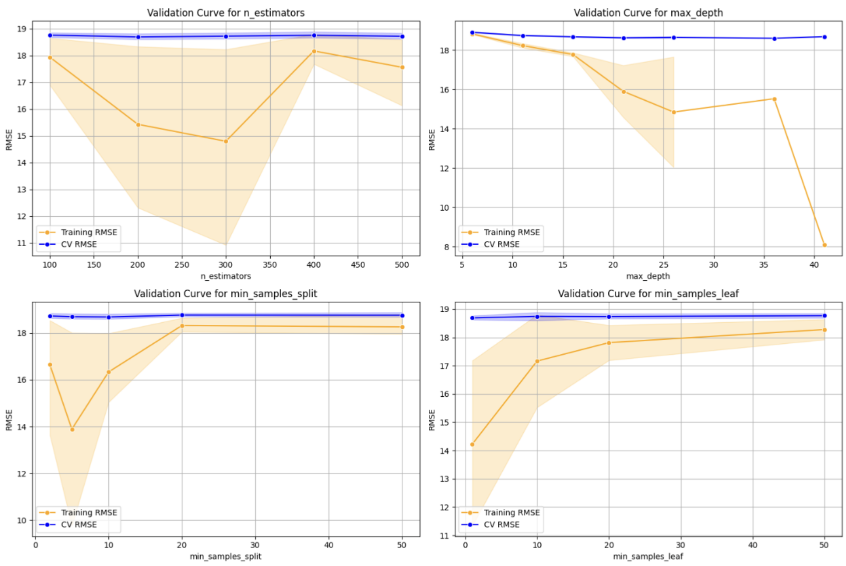

Figure 10: Validation with PCA

In each validation plot of Figure 10, the orange line represents the mean RMSE on validation data across each fold, while the blue line indicates its performance on the test data. As before, the parameters that achieve the minimum RMSE are consistent with the tuning results.

Next, the RF model was trained with these optimized parameters on the entire training dataset and tested on the test data, achieving a Root Mean Square Error (RMSE) of 18.56. Compared with the RF model with PCA, the performance deteriorated slightly, and with twice the number of iterations, the efficiency also dropped. The increase in RMSE might be attributed to the loss of variance and information after applying PCA, which explained only 80% of the variance in the original data. Regarding the undesirable impact on efficiency, it may be due to the complexity of the lower-dimensional space created by PCA. Although this space has fewer dimensions, it can be structurally more complex than the original feature space, potentially resulting in longer training times as models adjust to this new configuration.

6.6 Support Vector Regression (SVR)

RMSE	18.27

Table 10: The Output of SVR Model

We used Support Vector Regression (SVR) with an RBF kernel to predict the popularity of tracks. Firstly, we selected the numeric features. The dataset is split into training a testing sets with a ratio of 80:20 using a random state for reproducibility. Then, a pipeline is created that first scales the features and then applies the SVR.

The regularization parameter 'C' is set to 1.0 and it determines the trade-off between a smooth decision boundary and classifying training points correctly. The 'epsilon' parameter is set to 0.2. It determines the epsilon-tube within which no penalty is associated with the training loss function with points predicted within a distance epsilon from the actual value.

The model performance is evaluated using the RMSE. The RMSE is approximately 18.27, indicating that the predicted popularity scores deviate from the actual scores by an average of about 18.27 points on the popularity scale. It shows that predictions made by this model are close to the actual data points.

linear	18.52
poly	19.05
rbf	18.26

Table 11: SVR Kernel Performance

We did Feature Engineering (FE) to improve the performance of the model. We used polynomial feature expansion to predict a dependent variable. The application of polynomial features aims to capture non-linear relationships between the original features. The RMSE value resulting from using an SVR model with an RBF kernel is smaller than the RMSE we calculated before, indicating a more precise fit to the data. It also shows that RBF kernel is more successful in capturing these complexities compared to other kernels.

7 Results and Discussion

According to the feature importance score table from Random Forest, it shows that 'loudness' and 'instrumentalness' have relatively high importance scores, which suggests they are key factors in predicting song popularity within the Random Forest model. And songs with higher loudness and specific instrumental characteristics tend to be more popular. This is also supported by the correlation heatmap, where it shows a strong positive correlation for 'loudness' with popularity and strong negative correlation for 'instrumentalness' with popularity. Through LASSO regression, where 'loudness' also carries a substantial positive coefficient, showing a similar trend in a linear context.

'Danceability' and 'energy' despite having moderate feature importance and LASSO coefficients, exhibit weaker correlations, compared with the two features above. This might indicate a more nuanced relationship with popularity that Random Forest captures, potentially due to interactions between features or non-linear effects that are not as evident in a simple linear correlation. However, across the three analytical perspectives—feature importance, LASSO coefficients, and correlation, the evidence consistently points out that 'danceability' and 'energy' are favorable traits in popular tracks.

Features like 'key', 'mode', and 'time_signature' have minimal impact across all analyses, suggesting they have little linear relationship with popularity, nor do they contribute significantly to the models' predictive accuracy. Notably, 'speechiness' and 'acousticness' present a negative relationship with popularity based on their LASSO coefficients, despite their relatively lower feature importance in Random Forest. This indicates the potential complexity of their roles, possibly being overshadowed by more dominant features or interacting with other variables in complex ways not fully captured by linear models.

Overall, these analyses suggest that while some features have a direct and linear relationship with song popularity, others may influence popularity in more complex patterns or through interactions. Thus, strategies aimed at increasing a song's popularity should focus on enhancing clear positive drivers of popularity while considering the intricate part of feature interactions that may exist.

Model	**Test RMSE**
OLS	18.79
Ridge	18.80
LASSO	18.80
RF (no PCA)	18.38
RF (PCA)	18.56
SVR	18.27
SVR (FE)	18.26

Table 12: Models accuracy comparison

Table above presents a comparison of the test Root Mean Squared Error (RMSE) across five different models. Ridge and LASSO regressions both show very slight improvements in RMSE compared with the basic linear regression model OLS. Non-linear ensemble model Random Forest yields an RMSE of 18.36 and 18.56. This improvement over the linear models suggests that capturing non-linear relationships and interactions between features helps in predicting song popularity. Its better performance may also indicate that the data has a non-linear structure that Random Forest can exploit. SVR achieves the lowest RMSE of 18.26 after doing feature engineering, suggesting that this model is the best at capturing the complex relationships in this dataset.

From these results, it can be concluded that while linear models provide a baseline for prediction accuracy, exploring non-linear relationships with models like Random Forest and SVR can lead to improved performance. The SVR's superior performance indicates the possible presence of non-linear patterns that it can perform more effectively than the other methods. However, the differences in RMSE among the models are relatively small, which might suggest that the choice of model could be influenced by considerations other than accuracy alone, such as interpretability or computational efficiency.

8 Future Recommendations

The purpose of this analysis report is to reveal the factors that influence the popularity of a track and to develop a predictive model that predicts the popularity of a track based on its musical characteristics. The entire analysis process is based on multiple regression analysis, developing the application of Ridge regression and LASSO regression, as well as the implementation of Random Forest (RF) and Support Vector Regression (SVR), aiming to find a better predictive model. Each model provided unique insights, with MLR establishing a foundational relationship, while Ridge and LASSO addressed multicollinearity, and RF and SVR offered non-linear and complex decision boundaries.

For linear models, while Ridge and LASSO regression are used to mitigate these problems, the effectiveness of these techniques in substantially reducing multicollinearity remains to be demonstrated. Therefore, it is recommended to further use relevant statistical tests such as conditional exponents or variance decomposition proportions to assess the extent of post-multicollinearity regularization.

RF and SVR models are limited by current operational limitations and are unable to make extensive parameter tuning that would improve model accuracy. In future work, it would be beneficial to explore a wider range of parameter settings once computational resources allow. Additionally, advanced ensemble methods that combine multiple learning algorithms could yield a model with superior predictive capabilities. Techniques such as Gradient Boosting or stacking may leverage the strengths of individual models while compensating for their weaknesses. And given the capacity for handling large datasets and learning feature interactions, deep learning could also significantly enhance the predictive accuracy of the popularity model.

Furthermore, extending the feature set might also contribute to improving the performance of the model. Combined with additional song attributes, perhaps from external music databases or user-generated content such as reviews or ratings, can provide richer data for understanding listener preferences.

To summarize, while the current models serve as a solid starting point for predicting song popularity, there is substantial scope for refinement.

References

[1] A. Gao, *Catching the earworm: Understanding streaming music popularity using machine learning models*, E3S Web of Conferences, vol. 253, p. 03024, 2021. https://doi.org/10.1051/e3sconf/202125303024

[2] Y. Ge, J. Wu, and Y. Sun, *Popularity prediction of music based on factor extraction and model blending*, 2020 2nd International Conference on Economic Management and Model Engineering (ICEMME), 2020. https://doi.org/10.1109/icemme51517.2020.00214

[3] M. Interiano, K. Kazemi, L. Wang, J. Yang, Z. Yu, and N. L. Komarova, *Musical trends and predictability of success in contemporary songs in and out of the top charts*, Royal Society Open Science, vol. 5, no. 5, p. 171274, 2018. https://doi.org/10.1098/rsos.171274

[4] S. H. Merritt, K. Gaffuri, and P. J. Zak, *Accurately predicting hit songs using neurophysiology and Machine Learning*, Frontiers in Artificial Intelligence, vol. 6, 2023. https://doi.org/10.3389/frai.2023.1154663

[5] Interiano M, Kazemi K, Wang L, Yang J, Yu Z, Komarova NL. 2018. Musical trends and predictability of success in contemporary songs in and out of the top charts. *R. Soc. open sci.* **5**:171274. http://dx.doi.org/10.1098/rsos.171274

[6] DeWall CN, Pond RS, Jr., Campbell WK, Twenge JM. 2011. Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular U.S. song lyrics. *Psychology of Aesthetics, Creativity, and the Arts* **5**(3):200–207. https://doi.org/10.1037/a0023195

[7] Askin N, Mauskapf M. 2014. Cultural attributes and their influence on consumption patterns in popular music. In Aiello LM, McFarland D, editors. Social Informatics - 6th International Conference, SocInfo 2014, Proceedings. (pp. 508-530). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 8851). Springer Verlag. https://doi.org/10.1007/978-3-319-13734-6_36