Regression Analysis of Song Popularity based on Ridge, K-Nearest Neighbors and Multiple-Layers Neural Networks

Aoran Dong [1, †], Ruizhe Qiu [2, †, *] and Zhen Ye [3, †]

[1] Southwest Jiaotong University, Chengdu Sichuan 610031, China

[2] Northeastern University, Boston MA 02115, USA

[3] Macao University of Science and Technology, Macao 999078, China

* Corresponding Author Email: qiu.ru@northeastern.edu

†These authors contributed equally

Abstract. Contemporarily, human beings are working on the implementation of artificial intelligence technology in the arts fields, where the music is one of the directions. Before humans can create a song with artificial intelligence, it is necessary to understand the song first. This research tries to find out the relationship between the song's popularity and several selected songs' physical parameters based on statistics and machine learning. According to the analysis, this research proves that there is no significant relationship between selected physical parameters and the song's popularity. In addition, machine learning algorithms also do not find the potential relationships between them. In this case, it is safe to conclude that creating the song by considering these selected physical parameters is meaningless. On this basis, scholars should try to find out what factors make the song popular in terms of analyzing songs differently. These results shed light on guiding further exploration of future music analysis and artificial intelligence in music fields.

Keywords: Song Popularity; Machine Learning; Physical Parameters.

1. Introduction

Music plays a vital role in human life and is inseparable for almost every individual in the world. It has already become one of the most popular types of online information. There are now hundreds of music streaming and download services operating on the World-Wide Web [1]. Few songs can be granted ‘popular’, though tons of songs are published every year across the world. Being capable of predicting what makes a song popular can help us better understand humans’ emotions, guide music companies’ development, and improve their profits. Lots of research has carried on to investigate whether machine learning can predict the contents’ popularity [2-4].

There is a wide range of factors that could affect a song's popularity. Dhar and Chang collected albums' online comments from related blogs and social networking sites to investigate whether this variable can be used to predict sales in the music industry [5]. They found that future sales are positively correlated with the volume of blog posts about an album. A similar result is also presented by Shulman et al. [6], who collected 437,000 user interaction data from the Last.fm platform. Specifically, they used the data to develop a logistic regression model to predict a song's hotness in the Last.fm, with approximately 81% accuracy. Zangerle et al. evaluated the potential success of songs based on their acoustic characteristics and used a wide and deep neural network architecture to do the hit song prediction [7]. They suggested that the features of mood and voice were salient features in the model, and incorporating these two features can improve the accuracy of prediction. Interiano et al. analyzed the acoustic features of more than 500, 000 songs released in the UK from 1985-2015, revealing the time scale of fashion dynamics in popular music [8]. It is found that those successful songs tended to be 'happier', more 'party-like', less 'relaxed', and more 'female' than most. In addition, they found that considering the feature of 'superstar' can significantly improve the accuracy of prediction. Lee et al. defined 8 popularity metrics and predicted them using acoustic features. It is found that the features 'Chroma' and 'Arousal' performed well [9]. Araujo et al. created a model to see whether a song can be listed in Spotify's Top 50 Global ranking [10]. A total of 9 fields of information

(e.g., the entry's rank, the artist's names, duration) were collected. The results revealed that incorporating the acoustic feature contributed little to the accuracy (only 5.23% higher).

While AI painting becomes hot recently, it is easy for people to associate the song with it, which is another popular and common art in human life. Although there are some AI music generators now (e.g., Boomy), they are still not mature enough to generate a popular song or music. In this case, it will be interesting and necessary to analyze what are factors that affect the song's popularity. For this research, the physical parameters of the song will be the main target. The data will be introduced first, and then how data will be engineered for machine learning use is presented. Before going to the actual work, the machine learning model and skills used in this research will be introduced. Afterwards, feature analysis with visualization will be shown to better present the features and the relationship between the song's popularity. The machine learning model and result will be presented and explained after feature analysis. Finally, the limitations of this research will be pointed out, and the outlook for this research will be explained based on limitations.

2. Data & Method

2.1 Data

The dataset is retrieved from Kaggle and was extracted from Spotify API [11]. It has 18, 835 records and 18 attributes. The original dataset requires operations of data preprocessing and predictive modeling. In fact, one of the attributes of the dataset which is the name of the songs does not play a role in the machine learning case of the study since they do not affect the prediction results of the popularity. The 14 attributes left are important and will define the model predictions.

In this study, the popularity of the songs is regarded as the target variable and the other 13 attributes of the songs are regarded as the feature variable in the machine learning model. The following is a brief description of the 14 attributes:

- Popularity. The popularity of songs is a numerical variable that ranges from 1-100 to show how popular the song is.
- Duration. The duration of the song defines how long the song will continue playing from the beginning to the end under the unit of seconds. It is a numerical variable that ranges from 100,000 to 300,000.
- Acousticness. The acousticness of songs ranging from 0.0 to 1.0 define how acoustic a song is. It is a numerical variable in which the mean value and the variance value of all the data are around 0.26 and 0.29.
- Danceability. The danceability of songs is a dancing-based feature that is a combination of physical effects presented by the music. Its value is from 0.0 to 1.0 which shows how danceable the song is. It is a numerical variable.
- Energy. The energy of songs defines the energy of the sounds in the music which is a numerical variable ranging from 0.0 to 1.0. It is a numerical variable.
- Instrumentalness. Instrumentalness is the sound like 'Ooh' and 'Aah' treated as the instrumental feature in the song. It is a numerical variable whose mean value is around 0.08, and its variance value is around 0.22.
- Key. The key of a song is the note or chord the music is centered around. It is a categorical variable that translates the key of C, C#, D, D#... into 1, 2, 3, 4, and so on.
- Liveness. The Liveness of a song shows how the audiences participate in the song. It is a numerical variable that ranges from 0.0 to 1.0. The mean value of it is around 0.18, and its variance is around 0.14.
- Loudness: the loudness of a song is related to the dB produced when playing the music. It is a numerical variable in which all its values are negative.
- Audio Mode. The audio mode is about the modality characteristic of a song. It shows whether the mode is major or minor. So, it is a categorical variable that only has the value of 0 and 1.

- Speechiness. Speechiness indicates how likely a particular audio file is speech-like. It is a numerical variable that ranges from 0.0 to 1.0 with the mean value and variance value both around 0.1.
- Tempo. The tempo of a song is related to the beats per minute in the song. It is a numerical variable that ranges from 0 to 250 with a mean value around 121 and the variance value around 29.
- Time Signature. The time signature of a song is usually written at the beginning of a song represented by a set of numbers, one on top of the other, resembling a fraction. All the values of time signature in the dataset are which mean that it does not have any contribution to the prediction of song popularity.
- Audio Valence. Audio valence is the positiveness described by the valence measures which range from 0.0 to 1.0. It is also a numerical variable.

2.2 Feature Engineering

In the part of feature engineering, the goal of the work is to find out which of the features influence the popularity of the songs by measurably showing their importance and remove all the feature variables that don't have much effect on the target variable. First, some basic preprocessing methods are used to operate on the dataset to do the feature selection in the later stage. The dataset is split into two groups in the beginning, which one of them is the group for training and the other is the group for testing. One-fourth of the dataset is used for training the machine learning model and the other three quarters of the dataset is used for testing the performance of the model after this step. Then, feature scaling is achieved by the MinMax Scaler to normalize the range of the features in the dataset to get prepared for feature selection.

Subsequently, in the part of feature selection, two main techniques of Filter and Wrapper are used to view the importance of the feature variables in some specific ways and then select the ones that are truly effective for the machine learning model. The filter is to remove all the low variance feature variables from the scaled training dataset according to the value of the variance of each. The wrapper is to use machine learning models like the Decision Tree Regression Model and the Random Forest Regression Model to assess the degree of importance for each. By using the Filter, three features are removed from the scaled training dataset. The three feature variables are instrumentalness, audio mode, and time signature.

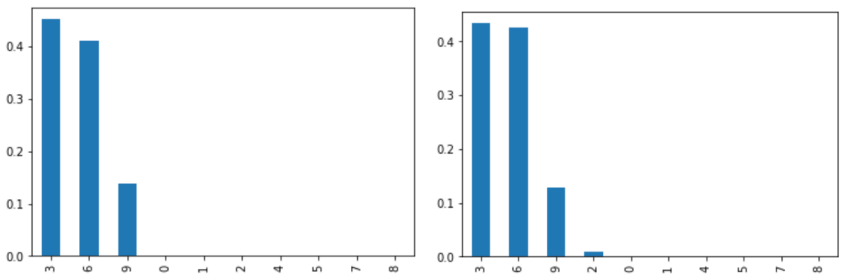


Fig 1. Feature importance.

In the stage of using the Wrapper, the process is divided into two main steps. The first step is to do the model-based ranking to score the importance of each feature variable left in the scaled training dataset. The second step is to do the Regression Feature Elimination known as RFE in short to find the optimal number of features by measuring the highest performance score given a different number of features to some effective machine learning model. In this entire process of the wrapper, two practical machine learning models including the Decision Tree Regression model and the Random Forest Regression model are chosen as the selector. In the model-based ranking, the rank of all the feature variables in order is shown in Fig. 1. According to the results, the feature variables rank in the

top 6 are energy, loudness, audio valence, song duration, acousticness, and danceability. Feature variables like key, liveness, speechiness, and tempo are obviously not that important compared to the top 6 feature variables.

In Regression Feature Elimination, the result which gives the optimal number of feature variables to do better prediction is 3. Because the optimal number of feature variables is 3, the selector then runs and returns a new feature variable dataset including the top 3 feature variables. The new feature variable dataset is the final scaled training dataset to be dealt with the machine learning model in the coming stages. Though it is not necessary to do the feature selection for a scaled training dataset with only 2 to 4 features after doing the feature extraction, the machine learning model is still tested to see how performance changes according to the number of chosen features. This also helps to clarify the importance of each feature.

2.3 Models & Evaluations

The machine learning models for first-round training include Linear Regression, Ridge, Lasso, Decision Tree, K-nearest neighbors, and Multiple Layer Neural Networks. Linear regression is a machine learning algorithm that finds a linear line to minimize the error. It will calculate the coefficient and intercept of the linear line it finds. In this research, the calculation methods of linear regression are closed-form and gradient descent. Ridge is a variant of linear regression, which works well when the predictor variables are highly correlated. This algorithm can lower the complexity of the dataset to decrease overfitting. Ridge works by limiting the sum of squares of the coefficient. Lasso is another variant of linear regression. It also works on the problem that predictor variables are highly correlated. The difference between Lasso and Ridge is that Lasso limits the sum of the coefficient. The decision tree regressor builds the decision tree to recognize the characteristic of given samples. The decision tree will find the most appropriate node of the sample by checking its features, and give the regression result based on the final node. K-nearest neighbor calculates the regression by finding the k nearest sample. It will calculate the distance between the given sample and all training samples, then find its neighbors and generate the value based on these neighbors. Multiple layer neural networks stimulate the work of human brain neural work, which includes more than one hidden layer for calculation.

For all models mentioned above, k-fold cross-validation is applied to test their performance. K-fold cross-validation works by dividing the whole dataset into k pieces. For each round, it will take one piece as a testing set, and all left as the training set. After testing all pieces, it will calculate the average performance as the final grade of the selected model. K-fold cross-validation can avoid the overfitting problem, and it works well for small datasets, so it can help to decrease the over-fitting in the song popularity regression model selection.

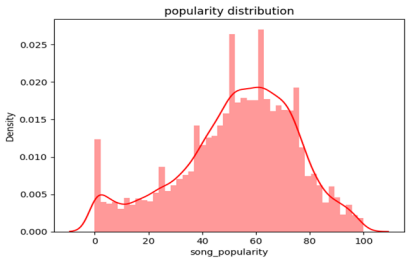


Fig 2. Song popularity distribution.

The two selected metrics are mean squared error and r squared score. The mean score error, which is calculated by equation $MSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2/n$, computes the average squared distance between the prediction and the actual number. It shows how close the regression line is to the actual data. R squared score is calculated by using equation $R^2 = 1 - RSS/TSS$, where RSS is the sum of squares of residuals, and TSS is the total sum of squares. R squared score reflects how well the

data fit the regression model. If a model performance, R squared score will be close to 1. Subsequently, three models with the best performance will be selected for the parameter tuning to get better performance. The way selected to do parameter tuning is grid search. Grid search will try all combinations of the given parameter space, which includes all parameters that need tuning and their tuning range. Finally, it will find a combination to help the model reach the best performance. Above all, these are all methods that will be used in the machine learning part.

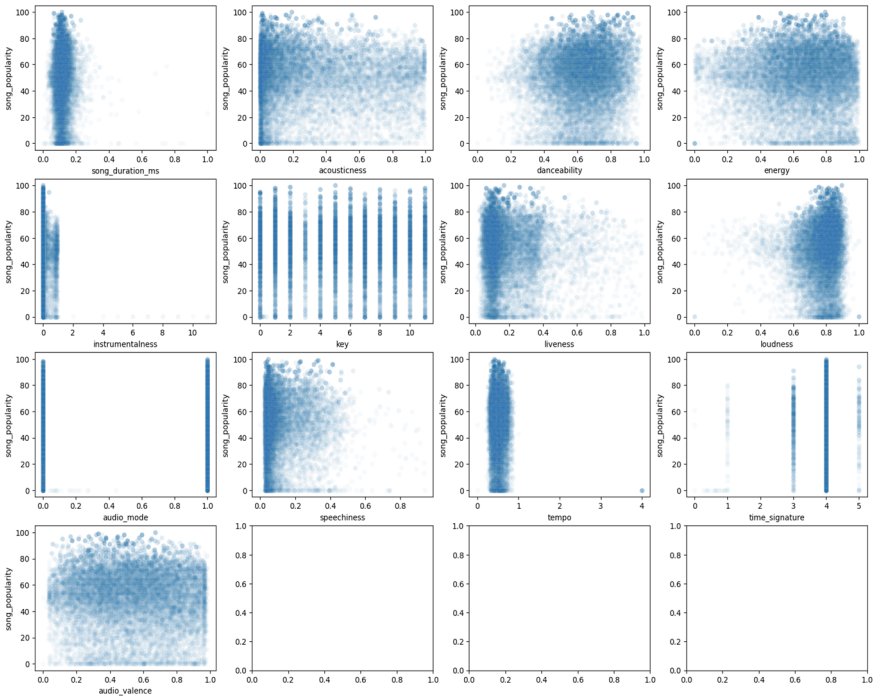


Fig 3. Song_popularity as a function of different Features.

3. Results & Discussion

3.1 Feature Analysis

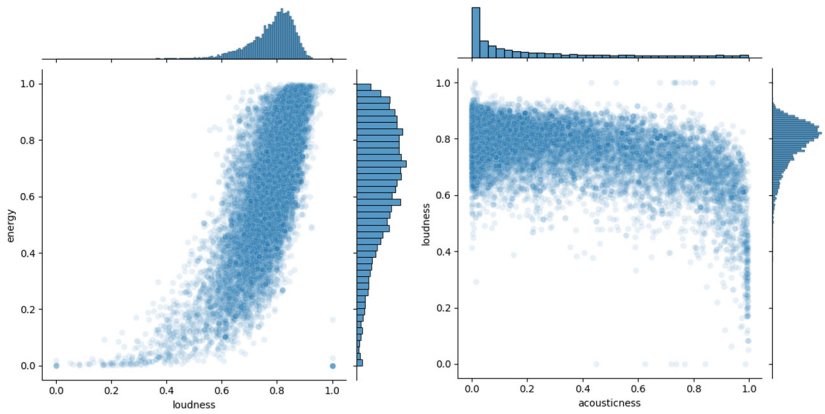


Fig 4. Energy as a function of loudness (left panel) and Loudness as a function of acousticness (right panel).

Fig. 2 shows the distribution of the popularity scores, where the scores are normally distributed, averaging around 60 units. This means most of the songs are not very popular, the score > 80 accounts for less than 5% of the total number of 18835 songs. Fig. 3 illustrates the scatter plots of the popularity vs. each feature. It is difficult to find and linear relationship between the score and any individual feature, which may cause challenges to the prediction. There are also some interesting relations that

can be found in the 'energy', 'loudness', and 'acousticness'. Seen from Fig. 4, 'energy' is positively related to 'loudness', but negatively related to 'acousticness'. This finding makes sense, due to a song with high - dB usually means noisy and high-intensity instead of acoustic.

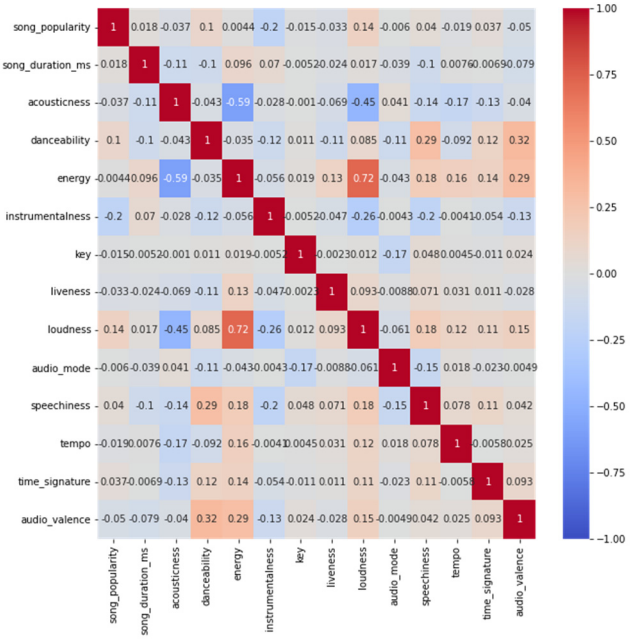


Fig 5. D heatmap with correlation coefficients

In order to see the potential multicollinearity within every individual variable, a 2D heatmap is created with Spearman's correlation coefficients as exhibited in Fig. 5. Nearly no high correlation coefficient can be found between each variable, except 'energy', 'loudness', and 'acousticness'. This matches the observation found before. It is noted that all features have very low correlation values with the song's popularity score, this may be because those previously proved salient components of music (e.g., emotion/mood and voice) have not been included in the current features, implying the difficulty in predicting the popularity.

3.2 Regression & Comparison

Because the feature selection only keeps three features and culls all other left features, its feature number might be too less for this dataset to finish the machine learning. The model performance with only three features is worse than the model with all features. Considering that some feature combinations influence the song's popularity, the models are trained with all features.

Table 1. Model performance after cross validation

Models	MSE	R^2
Linear Regression	458.294746	0.044593
Ridge	458.290828	0.44602
Lasso	479.86081	-0.000321
Neural Network	447.132039	0.067814
k-Nearest Neighbor	431.823685	0.099565
Decision Tree	538.392635	-0.122735

After k-fold cross-validation, the performance of all selected models is shown in the Table. 1. Generally, it is not a good performance for machine learning. For mean square error, its range is about 430 - 540, which is too large for popularity in the range 0 - 100. R squared score is fluctuate in the range of -0.13 - 0.1. While the R^2 score is close to zero, it means the regression is not correlated to

the actual song popularity. Given such a bad performance, parameter tuning is operated for the three best models to check if the performance of song popularity regression can be improved. The selected models are Ridge, Multiple-Layers Neural Networks, and K-Nearest Neighbor.

For Ridge, its parameter space includes max iteration number, which will be selected in list 100, 500, and 1000, and its learning rate, which will be selected in list 0.01, 0.05, 0.1, and 0.5. The result for training set is 460.26 MSE plus 0.047 R^2 score, and 428.88 MSE plus 0.09 R^2 score for testing set. While the learning rate is 0.5 and the max iteration number is 100, Ridge can reach the best performance. Compared with the Ridge without parameter tuning, its performance for the training set does not have significant change, but its performance for the testing set improves a lot.

For K-Nearest Neighbor, the parameter space only contains its neighbor's number, which has a range from 1 to 10 (inclusive). The k-nearest neighbor reaches the best performance when the neighbors' number is 10. The performance for the training set, which is 361.98 MSE and 0.25 r2 score, is much better than the 428.11 MSE and 0.09 R^2 score for testing set, which means the k-nearest neighbor model is overfitting. However, though it is overfitting, its testing performance is still higher than Ridge.

Finally, parameter space for the multiple-layer neural network includes max iteration, which will be selected in list 100, 500, 1000, activation function, which will be one of 'tanh' and 'relu', solver, which will be one of 'sgd' and 'adam', and its learning rate, which will be selected in list 0.01, 0.05, 0.1, 0.5. While the activation function is 'tanh', the learning rate is 0.05, max iteration number is 500, and the solver is 'sgd', the multiple-layer neural network has the best performance, which is 420.61 MSE plus 0.13 R^2 score for training set and 426.79 MSE plus 0.09 R^2 score for testing set. Its performance is improved. Besides, though its training set has better performance, the difference between the training set and testing set is acceptable, i.e., it is not overfitting. However, its performance is still too low for the practical prediction.

3.3 Explanation & Application

According to the metrics, all three selected models have a better performance after parameter tuning. However, their final performance is still too bad for practical application. While a good performance model could help to predict the song's popularity according to its physical parameters, these failed models are not meaningless. Since these algorithms cannot build a regression correlated well with actual data, it is safe to assume that there is no simple relationship between song popularity and its physical parameters. The correlation and importance calculation above also proved the assumption that there is no significant relationship between the physical parameters of the song and song popularity. In this case, this research points out that trying to analyze song popularity and create a popular song by only considering physical parameters is impossible. According to this conclusion, if people want to analyze the song by physical parameters or build artificial intelligence that can generate great music, it is necessary to consider extracting different physical parameters from the song. Moreover, it is also necessary to consider other factors outside of physical parameters, such as market preference, singer popularity.

4. Limitations

For this research, one of the limitations should be the miss of culture consideration. The cultural environment may have a strong influence on the song's popularity. For example, people in different countries, of different ages, and with different interests will have their own preferences. In this case, it is important to consider the target audience of the song before considering its popularity. The second limitation is the style of the song self. Different song styles should have different characteristics. Though songs with the same style will not be the same, a song will not break its style inexplicably. For instance, it will be very strange to add a rock period into new-age music. Thus, considering all songs together without classifying their styles may lead to a decrease in accuracy. Finally, current features are also a limitation of this research. As proven above, there is no significant relationship

between current features and song popularity. Therefore, it is reasonable that the performance of the machine learning model is bad. If there are more features to select, or the given features have a strong relationship with song popularity, the machine learning model can be more practical.

Based on these limitations, there are some actions that can be taken for future research. Primarily, one should collect the song's popularity in different cultures groups. This action is for building different song popularity machine learning models in different communities. As mentioned above, different cultures' preferences may have a great effect on song popularity. So, it is necessary to prove or disprove this relationship. If there is such a kind of relationship, it will be important to build a machine learning model based on this rule, even if it requires a lot of work. Secondly, the styles of the song should also be integrated into features. Moreover, a new feature that can be called "style matching degree" can also be added to new features. In other words, the features should display how well a song fits its style, and then check if these features correlated to the song's popularity. Finally, all physical parameter features used in this research should be removed, and new physical parameter features should be considered and added to the dataset. It will require analyzing the song in other ways. This will be a more profound direction, which may relate to psychology and human brain science. The new features can be if a part of a song can stimulate the brain, or continuous and smooth a song is. As there are several research directions, it may take a long time to finish it and find those features that affect the song's popularity. However, it is necessary to find out it, since it can help human better understand how song attracts and affect people. Besides, it can also help to build the base for the possible AI singer in the future, which can create a wonderful song by the computer as what AI painter doing now.

5. Conclusion

In summary, the results of importance calculation, visualization, and regression machine learning prove that there is no significant relationship between these features, which are some physical parameters of a song, and the song's popularity. In this case, it can safely state that these features cannot be used to predict the song's popularity simply. If humans want to predict the song's popularity, it is necessary to extract other features from songs. While this research is limited by culture environment, styles, and other song features, these limitations can be the direction for further study in this field. With the development of this field, human beings will finally understand how song attracts people, and the way to build artificial intelligence that can create a popular song, similar to what the current ai painting is doing now. Overall, these results offer a guideline for future music analysis and artificial intelligence in music fields.

References

[1] Casey, M. A., Veltkamp, R.: Content-based music information retrieval: Current directions and future challenges. Proceedings of the IEEE, 96(4), 668-696 (2008).

[2] Trzciński, T., Rokita, P.: Predicting popularity of online videos using support vector regression. IEEE Transactions on Multimedia 19(11), 2561-2570 (2017).

[3] Wu, J., Zhou, Y: Modeling dynamics of online video popularity. IEEE Transactions on Multimedia, 18(9), 1882-1895 (2016).

[4] Araujo, C. V. S., de Cristo, M. A. P.: A model for predicting music popularity on streaming platforms. Revista de Informática Teórica e Aplicada, 27(4), 108-117 (2020).

[5] Dhar, V., Chang, E. A.: Does chatter matter? The impact of user-generated content on music sales. Journal of Interactive marketing, 23(4), 300-307 (2009).

[6] Shulman, B., Sharma, A.: Predictability of popularity: Gaps between prediction and understanding. In Tenth International AAAI Conference on Web and social media (2016, March).

[7] Zangerle, E., Vötter, M.: Hit Song Prediction: Leveraging Low-and High-Level Audio Features. In ISMIR, pp. 319-326 (2019).

[8] Interiano, M., Kazemi, K.: Musical trends and predictability of success in contemporary songs in and out of the top charts. Royal Society open science, 5(5), 171274 (2018).

[9] Lee, J., Lee, J. S.: Music popularity: Metrics, characteristics, and audio-based prediction. IEEE Transactions on Multimedia, 20(11), 3173-3182 (2018).

[10] Araujo, C. V. S., de Cristo, M. A. P.: Predicting music popularity using music charts. In 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 859-864. IEEE (2019, December).

[11] H, M. Y. (2022, January 7). Song popularity dataset. Kaggle. Retrieved October 4, 2022, from https:// www. kaggle. com/datasets/yasserh/song-popularity-dataset, (2022, January 7).