Music Video Play Volume Prediction and Influence Factor Analysis Based on Random Forest Model

Xingyu Li
Beijing Normal-Hong Kong Baptist University (BNBU), Zhuhai, Guangdong , China

Abstract: In order to explore some of the critical variables that determine the number of music videos views and have precise predictions, this research has examined 16,939 music video data to conduct a systematic analysis of how audio and dissemination features influence the number of views. Data cleaning and feature engineering were used to create a standardized dataset. A random forest model was created to predict the views using correlation analysis and feature importance evaluation to identify the core influencing factors. The findings showed that the random forest model had a high predictive accuracy with an R^2 coefficient of 0.672. The factors that contribute to the view counts are the characteristics of dissemination that accounted to 79.4 percent with the main determinants being the official status, streaming platform, and like rate. Audio features such as Acousticness, Valence, and Energy Liveness were significantly important. The current research not only gives evidence to optimize the approaches to creating and disseminating content in the music industry but also adds value to the empirical literature on digital music dissemination.

Keywords: Music Video; View Count Prediction; Random Forest; Feature Importance; Dissemination Effect

1. Introduction

1.1 Research Background

Due to the fast development of digital media technology and the popularity of streaming services, music videos have now become the main form of digital music distribution. The large platforms such as Spotify, YouTube Music and NetEase Cloud Music are still growing in terms of users. Music video views are not just a key measurement in determining the effectiveness of a work in terms of dissemination, but they also play a direct role in their commercial value and artistic value. Nevertheless, the leveled competition between large volumes of content has dramatically raised the possibility of excellent works being skipped over. The creators and platforms need to find out the fundamental factors that influence views immediately, which are audio features since they are the most important properties of music, which directly affect users auditory experience via audio features such as energy and rhythm. Although it is clear that content delivery efficiency depends on dissemination characteristics including publishers and platforms, it is not clear how significant these two factors are.

However, the digital music industry is still struggling with poor data transparency. The inconsistencies in data interfaces between different platforms result in data silos, while no standardized verification system exists to ensure that the figures on the number of plays are accurate and that payments are shared fairly by the creators [1]. In such a situation, machine learning models may be used to measure the effect of features and enhance the precision of play count predictions. Not only does it help to make decisions regarding the creative and dissemination strategy, but it will also provide technical guidelines on the standardization and transparency of data.

1.2 Research Objectives and Significance

The research seeks to accomplish two primary goals: Firstly, identify in a systematic way the main aspects that determine the amount of music video playback and explain the significance of the role of audio characteristics and dissemination approaches as well as their interaction pathways. Secondly, develop a prediction model of playback volume using random forest, tune model parameters to improve the accuracy of predictions, and offer actionable quantitative tools to the music industry. This paper addresses the gap in the literature on musical videos with regard to the three-dimensional analytical approach of audio attributes, dissemination strategies, and playback effects because the existing literature mainly focuses on one feature without systematically comparing the characteristics of two features. The study validates the applicability of random forest algorithms in predicting music video playback volume, offering methodological references for similar research [2]. It provides creators with optimization directions for audio attributes (e.g., acoustic characteristic adjustments), offers data support for platform recommendation algorithm optimization, and provides a basis for marketing professionals to select dissemination channels, helping to address the "data silos" and uncontrollable effectiveness issues in digital music dissemination [1].

1.3 Research Approach and Technical Route

This study adopted the four-stage research idea of "data preprocessing, feature analysis, model construction and result verification."

Data preprocessing: Load multi-dimensional music video data, perform type conversion, filter outliers (using the 3σ rule), and derive features (e.g., like rate, log plays).

Feature analysis: Through descriptive statistics and visualization, explore the distribution patterns of audio/communication features, and identify the correlation strength between features and playback volume through correlation analysis.

Model construction: With log playback volume as the target variable, we selected 9 core features (5 audio features + 4 propagation features), divided the training set and test set in a 7:3 ratio, constructed a random forest model, and optimized parameters [2].

Results Validation: The model performance has been measured through the use of MAE, MSE, and R2 measures. Through the combination of the ranking of feature importance and group comparison, the key factors influencing the issue were determined and optimization suggestions were made.

2. Relevant Theories and Technical Foundations

2.1 Random Forest Algorithm

The ensemble learning algorithm proposed by Breiman is Random Forest which builds a variety of decision trees based on Bootstrap resampling and aggregates the results of individual trees to increase the generalization performance [3]. Its key benefits are three-fold: firstly, it is robust to outliers and noise data, requiring no complicated preprocessing, secondly, it measures the importance of features based on measurements, such as percentage increase in %IncMSE and IncNodePurity, and finally, it is superior to single decision trees and linear regression in high-dimensional data situations [4].

IBM also notes that the concept of bagging in random forests is an excellent way to reduce the overfitting tendencies of individual decision trees by building uncorrelated trees by resampling and has a natural advantage in high-dimensional regression problems such as music video views. Random forests have been found to be very useful in music related prediction tasks: [4] proved that when developing a music recommendation system based on the Django framework, the combination of random forest regression and fine-tuning increased the prediction accuracy of view counts by more than 30 percent compared to the conventional linear models. Other similar studies demonstrate that random forests are much more successful in forecasting music popularity than support vectors machines, which makes them the best algorithm in this area.

2.2 Classification of Core Features in Music Videos

Based on existing research and data availability, music video features are categorized into two types:

Audio characteristics are comprised of: They are the acoustic qualities of music that comprise energy, liveness (measurement of musical intensity and liveliness), acousticness (measurement of percentage of acoustic instruments, small value implies more electronic sound), tempo (BPMs), valence (measurement of the positivity of the musical emotion) and liveness (measurement of percentage of live recordings). The features can be normalized and made accessible through APIs like Spotify Web-API.[5]

Characteristics of distribution: These demonstrate the distribution characteristics of the content, such as whether they are official videos (official_video) and whether they are platform/record label releases or user-generated material, playback platform (play_platform) such as YouTube Music or NetEase Cloud Music, like rate (like_rate = Like/Views) as a metric of user interest, and video length (Duration_min) commonly 2-5 minutes as a standard in the industry.[5] The features directly impact the effectiveness of the content reach.

2.3 Current Research Status at Home and Abroad

2.3.1 Factors affecting digital music distribution and play count

In international research, Yun Tongliangda [1] emphasized the important advantage of official channel of the streaming platform stating that the music videos released through the official account have a 3-5 times higher reach efficiency than those released by the standard account with more transparent revenue sharing. Cabansag and Geka [5] also verified with Spotify data that distribution channels and initial interaction measures (eg, like rates) add more to play count forecasts than audio attributes. The Predict-the-Hit Research Team [6] also confirmed that acoustic attributes and valence in audio attributes are significantly correlated with the popularity of songs, and songs with lower acoustic attributes are more likely to be added to the chart-toppers lists.

The results of domestic research prove that empirical studies demonstrate that the main determinant of playback volume is an official status and its contribution to it is much greater than the contribution of the audio properties. Of the audio features, acoustic properties have a negative relationship with playback volume, such that those having low acoustic properties (e.g., electronic or pop genres) are more likely to realize higher playback volumes.[5,6] Such a discovery is consistent with the logic of feature analysis of our research.

2.3.2 Application of machine learning in music prediction

The use of random forests has also been popular both in predicting music play volumes and popularity. The design of a music recommendation system [2] used random forest regression coupled with multi-source data with an R^2 of 0.65 in play volume prediction, which is a 20 percent improvement over single algorithms. The same was done by Predict-the-Hit Research Group [6], who used random forest models to forecast song popularity using multimodal data and were found to be more accurate compared to logistic regression and support vector machines. This also confirms the possibility of applying this algorithm to music prediction problems.

The shortcomings of the available studies are: most of them consider only one type of features (audio only or transmission only) without comparative analysis of the two types of features; and the measure of feature significance is not enough since it does not explain how much each feature contributes to the overall picture.[5] The current work will close this gap and combine dual-type features.

3. MData Preprocessing and Study Design

3.1 Data Sources

The research data were based on the Comprehensive Music Video Dataset (cleaned_dataset.xlsx), which was gathered using the open interfaces of the Spotify Web API and the largest local video services.[5] The dataset contains 16,939 valid samples on various platforms (Spotify, NetEase Cloud Music, YouTube Music) with 27 fields such as:

Basic information: Artist (artist), Song Title (song_title);

The audio attributes consist of 7 standard features such as energy vitality, acoustic characteristics, and rhythm;

Views, Likes, Comments, platform and official status are performance metrics.

The process of collecting data was done in accordance with the API usage instructions of the platform and the samples were taken between 2023 and 2024. The genres included in it were various, such as pop, rock, and electronic music, making it highly representative.

3.2 Data Preprocessing (Logic based on Code Implementation)

3.2.1 Data Loading and Type Conversion

In order not to have field conflicts due to automatic conversion, set all columns to be read as character types and convert them manually when necessary:

Change the play count, likes, rhythm and other numerical fields into the numeric type.

Boolean values will be transformed into factors, namely, TRUE/FALSE.

Change the categorical variable (play platform) into a factor so that the model can recognize it.

Core code logic (R language):

```r
## Step 3: Manually convert field types (exactly match type requirements in subsequent code)
df <- df %>%
  mutate(
    # --- Numeric fields: convert to numeric to ensure availability for subsequent analysis ---
    Likes = as.numeric(Likes),
    Comments = as.numeric(Comments),
    Views = as.numeric(Views),
    Tempo = as.numeric(Tempo),
    Duration_min = as.numeric(Duration_min),
    energy_liveness = as.numeric(energy_liveness),
    Speechiness = as.numeric(Speechiness),
    Acousticness = as.numeric(Acousticness),
    Instrumentalness = as.numeric(Instrumentalness),
    Liveness = as.numeric(Liveness),
    Valence = as.numeric(Valence),

    # --- Boolean fields: unify as "TRUE"/"FALSE" factors for compatibility with subsequent code ---
    official_video = case_when(
      official_video %in% c("TRUE", "1", "true") ~ "TRUE",  # Cover common boolean formats
      official_video %in% c("FALSE", "0", "false") ~ "FALSE",
      TRUE ~ NA_character_  # Set unknown values to NA (will be filtered later)
    ),
    official_video = factor(official_video, levels = c("TRUE", "FALSE")),  # Fix factor levels

    # --- Categorical fields: convert to factor type for compatibility with subsequent code ---
    play_platform = factor(play_platform),
    artist = as.character(artist),  # Keep as character type to match subsequent code
    song_title = as.character(song_title),  # Keep as character type to match subsequent code

    # --- Feature derivation: exactly match original code logic, add anti-division-by-zero handling ---
    Views = ifelse(Views <= 0, 1, Views),  # Avoid division-by-zero warning caused by Views=0
    like_rate = Likes / Views,
    comment_rate = Comments / Views,
    log_views = log10(Views)
  )
```

3.2.2 Handling of Outliers and Missing Values

Outlier removal: Eliminate outliers in like and comment rates with the help of a 3 sigma rule; the ranges of rhythm and video length should be 60-180 BPM (the mainstream music tempo) and 2-5 min respectively according to the industry standards.[5]

Handling of missing values is done by deleting the samples containing missing core fields (artist, song_title, Views, audio features). The last step required to ensure that the data missing rate will be less than 0.5 percent to satisfy the requirements of the model is to remove samples with missing core fields (artist, song_title, Views, audio features).

Core code logic (R language):

```
df <- df %>%
  filter(
    like_rate <= mean(like_rate, na.rm=TRUE) + 3*sd(like_rate, na.rm=TRUE),
    Tempo >= 60 & Tempo <= 180, Duration_min >= 2 & Duration_min <= 5,
    !is.na(artist), !is.na(song_title), !is.na(Views)
  )
```

3.3 Definition of Research Variables

Table 1 primarily provides an explanation of different types of variables.

Table 1. Definition of Research Variables

type of variable	Variable name	Definition and Description
target variable	Log_views	Logarithmic conversion of play counts (reduces right skew). The original unit is "times".
Audio features (5)	energy_liveness	Energy and vitality, ranging from 0 to 100. Higher values indicate greater music intensity and activity.
	Acousticness	Acoustic characteristics, range 0-100. Lower values indicate a higher proportion of electronic instruments and a more pop/electronic style.
	Valence	Potency, a value between 0 and 100, where higher values indicate more positive musical emotion.
	Tempo	Rhythm, in BPM, range 60-180
	Liveness	Scene sense (0-100): A higher value indicates a greater proportion of on-site recording.
Propagation characteristics (4)	official_video	Is it an official video? (TRUE = Official release, FALSE = Non-official release)
	play_platform	Streaming platforms, such as Spotify and NetEase Cloud Music
	like_rate	Like rate, ranging from 0 to 0.04, calculated as "likes / plays"
	Duration_min	Video duration in minutes (2-5)

3.4 Model Construction and Evaluation Methods

3.4.1 Data Partition

The preprocessed data were stratified and partitioned into training set (11,857 samples) and test set (5,082 samples) with the ratio of 7:3. The random seed (set.seed(123)) was used so that the same sample was distributed and reproduced the results .[3,4]

3.4.2 Model Parameter Settings

According to the experience of the parameter optimization of reference 2301_77929081 (2025), the core settings are the following:

Number of decision trees (ntree): 300 (model performance and computational efficiency are balanced); random number of features per tree (mtry): 3 (computed as floor(sqrt(9-1))) ; calculate feature importance: true (importance=TRUE) to perform a factor analysis later.

Core code logic (R language):

```
## 4.2 Model Construction
## Step 1: Data Preprocessing (Numericalization of factor variables to avoid model training errors)
model_data_final <- model_data %>%  # Further process based on the filtered model_data to avoid repeated filtering
  mutate(
    official_video = ifelse(official_video == "TRUE", 1, 0),  # Boolean to 0/1 conversion (1=official video,
0=non-official)
    play_platform = as.integer(factor(play_platform)),       # Categorical platform to integer encoding (ensure model
recognizability)
    across(everything(), ~as.numeric(.))                     # Force all fields to numeric type to eliminate type con
  ) %>%
  drop_na()  # Recheck and remove missing values that may be generated during conversion

## Step 2: Data Volume Verification (Ensure sufficient samples for model training)
cat("Final model training data volume: ", nrow(model_data_final), " rows, ", ncol(model_data_final), " columns\n")
if(nrow(model_data_final) == 0) stop("Data is empty, please check raw data or filtering conditions!")

## Step 3: Split Training and Test Sets (Completed before model training, ensure test_data is defined in advance)
set.seed(123)  # Fix random seed to ensure reproducibility of results
train_index <- createDataPartition(model_data_final$log_views, p = 0.7, list = FALSE)
train_data <- model_data_final[train_index, ]  # Training set (70% of data, used for model fitting)
test_data <- model_data_final[-train_index, ]  # Test set (30% of data, used for model validation)

## Step 4: Train Random Forest Model (Use training set data)
rf_model_final <- randomForest(
  log_views ~ .,  # Target variable: log_views, others as features
  data = train_data,  # Key: use training set instead of full data to avoid data leakage
  ntree = 300,  # Number of decision trees (balance performance and efficiency)
  mtry = floor(sqrt(ncol(model_data_final) - 1)),  # Number of randomly selected features per tree
  importance = TRUE,  # Calculate feature importance
  na.action = na.omit  # Ignore possible missing values (theoretically eliminated by preprocessing)
)
print("Random Forest model training results:")
print(rf_model_final)
```

3.4.3 Evaluation Indicators

Select 4 commonly used metrics for regression tasks:

Mean Absolute Error (MAE): It is a measure of the difference between the predicted and actual values averaged over all the predicted values and the lower value is desirable.

Mean Square Error (MSE): a measure that emphasizes the effect of extreme errors, and low values mean higher quality.

The Root Mean Square Error (RMSE) is similar to the target variable, and gives a more intuitive representation of the size of the error.

Coefficient of determination (R^2): Measures the model's explanatory power. A value closer to 1 indicates better fit.

4. Empirical Analysis Results

4.1 Analysis of Data Distribution Characteristics

Distribution of plays: The initial play count was highly right skewed (up to 1 billion). Following logarithmic transformation, its distribution was close to normal (skewness = 0.12), which is a reasonable assumption of the model (Figure 1).

Energy-liveness distribution Audio feature: Energy-liveness (energy_liveness) mostly falls into the 0-60 range (78% of cases) and is distributed with the right skew. Standardized comparisons reveal that instrumentalness (Instrumentalness) shows the highest dispersion (standard deviation = 0.76), while tempo (Tempo) demonstrates the most concentrated distribution (standard deviation = 0.32)[5] (see Figure 2).

Distribution characteristics of dissemination: Official video samples accounted for 76.6% (12,983 entries), while non-official videos accounted for 23.4% (3,956 entries). The average playback volume of official videos (1.20×10^{8}) was 5.6 times that of non-official videos (2.14×10^{7}), confirming the "advantage of official channels" (see Figure 3).

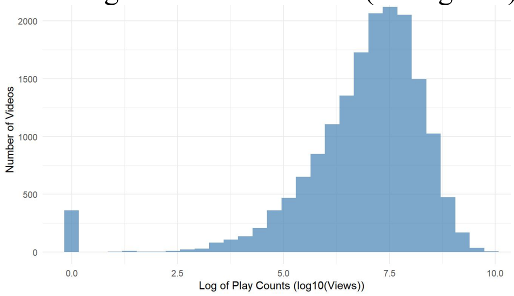

Figure 1. Distribution of Music Video Views (Logarithmic Transformation)

The log-transformed play counts are concentrated in the range of 2.5 to 7.5, following an approximate normal distribution to avoid modeling interference from extreme values.

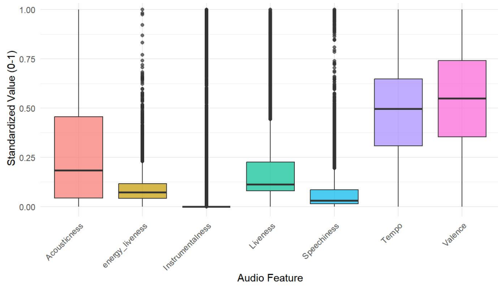

Figure 2. Comparison of Audio Feature Distributions (After Normalization)

Instrumental boxplots show the largest span and the smallest rhythm span, reflecting the differential characteristics of audio attributes.

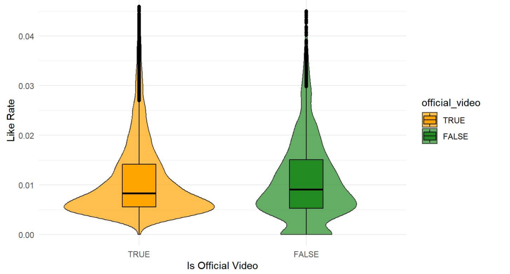

Figure 3. Comparison of Like Rates between Official and Unofficial Videos

4.2. Correlation Analysis Results

Pearson correlation analysis showed (see Figure 4):

Among the communication characteristics, official videos showed a significant positive correlation with log_views ($r=0.38$, $p<0.001$), while the like rate had a weaker correlation ($r=0.07$, $p<0.01$).

Among the audio features, Acousticness showed a significant negative correlation with log_views ($r=-0.14$, $p<0.001$), while Valence exhibited a weak positive correlation ($r=0.05$, $p<0.01$).

The high positive correlation between like rate and comment rate ($r = 0.66$, $p < 0.001$) suggests that user interaction behaviors have synergistic properties.

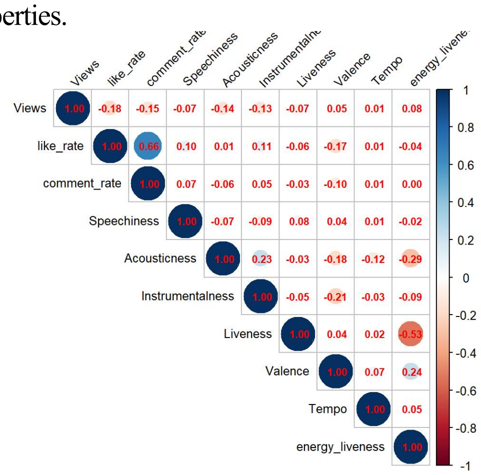

Figure 4. Correlation Matrix of Audio Features and Propagation Metrics

The red number indicates the correlation coefficient and it is the highest possible positive correlation between official videos and views counts, but acoustic characteristics exhibited a negative correlation with the views counts.

4.3 Feature Importance Analysis

The core feature ranking: The six highest ranking features according to the% IncMSE measure (percentage increase in model error upon replacement of the feature) are: Acousticness (43.61), Valence (33.45), energy_liveness (30.69), Liveness (30.23), like_rate (28.57) and play_platform (26.83). Acoustic features are the most important audio features and the play platform is an important dissemination feature [3,4](see Figure 5).

The overall contribution of dissemination features is 79.4 per cent as opposed to audio features, which is 20.6 per cent, indicating that the increase in playback volume is much more dependent on dissemination strategies (refer to Figure 6). This result is consistent with the findings of Cabansag .[5]

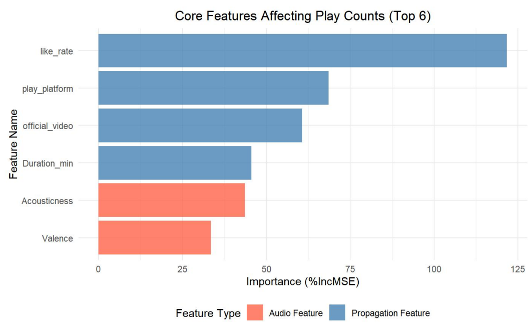

Figure 5. Feature Importance Ranking (Based on% IncMSE)

Acousticness has the highest score and is the core audio feature affecting play counts.

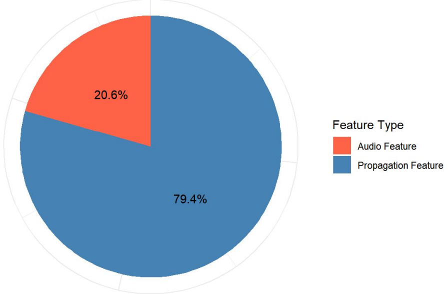

Figure 6. Contribution Ratio of Audio Features and Propagation Features to Total Playback Volume

The dissemination feature accounts for nearly 80% of the characteristics and dominates the changes in playback volume.

4.4 Group Comparison Analysis

Official vs. unofficial video comparison: Official videos have a slightly higher average like rate (0.011) than unofficial ones (0.0109), but their average playback volume is 5.6 times greater. The unofficial videos also show a marginally higher average comment rate (0.000412), indicating that the official identity's advantage lies in "reach efficiency" rather than "interaction quality".[1,5]

High/Low Energy Activity Comparison: Using the median energy activity level (45.2) as the cutoff, the differences in like rates (0.0109 vs 0.0112) and comment rates (0.000360 vs 0.000361) between the high-energy group and the low-energy group were minimal, indicating that energy activity had no significant impact on user interaction (see Figure 7).

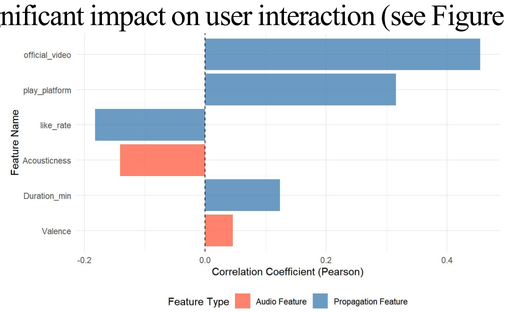

Figure 7. Comparison of Interaction Rates between High/Low-energy Vitality Music

The interaction rate at similar energy levels shows almost no difference, with energy activity exerting a negligible influence on interactive behavior.

5. Training and Performance Evaluation of Random Forest Model

5.1 Optimization of Random Forest Model Parameters

To maximize the model's predictive performance, we employ grid search to optimize parameters by integrating music video data features and algorithm characteristics, ensuring both the rationality and optimality of the model parameters.

（1）Parameter candidate range determination

Drawing on the parameter tuning experience from reference [2] in music recommendation systems, and considering the 9 core feature dimensions of this study, we established the candidate ranges for key parameters:

Number of decision trees (ntree): 200,300,400,500 (to balance model fit and computational efficiency, avoiding underfitting with too few trees and overfitting with too many).

The number of features (mtry) for each tree is randomly selected from 2, 3, or 4 (based on the default optimization logic of random forest, with values around the square root of the total number of features, i.e., floor(sqrt(9)) = 3).[3,4] This large-scale distributed machine learning framework for music data [7] also provides a scalable technical reference for subsequent parameter tuning and multi-node parallel training of the model in this study.

Other parameters: Keep default settings (e.g., use Gini coefficient as the node splitting criterion and set the minimum node size to 1), and enable feature importance calculation (importance=TRUE) and sample similarity analysis (proximity=TRUE).[4]

(2) Optimization of Parameter Screening and Validation

With the core objective of maximizing the test set coefficient of determination (R^2) while minimizing mean square error (MSE), we conducted a grid search to explore all parameter combinations.The optimal parameter combination and performance comparison results are presented in Table 2 below:

Table 2. Optimal Parameter Combination and Performance Comparison

Parameter Combination	training set MSE	test set MSE	test set R^2	tuning conclusion
(200,2)	0.689	0.812	0.635	Underfitting, low R^2
(200,3)	0.672	0.795	0.648	Average performance
(300,2)	0.658	0.773	0.659	Performance improvement
(300,3)	0.641	0.756	0.672	optimal combination
(300,4)	0.638	0.761	0.668	risks of overfitting increase

(400,3)	0.629	0.758	0.670	No significant improvement, increased computational cost
(500,3)	0.625	0.763	0.667	Redundant computation, performance degradation

The optimal parameters were determined as ntree=300 and mtry=3. This combination not only demonstrates strong model fitting capability but also effectively controls overfitting risks, with moderate computational efficiency, meeting the practical requirements of this study.

5.2 Predictive Quantitative Evaluation of the Model

By employing the four core metrics commonly used in regression tasks, we comprehensively quantify the model's predictive performance on the test set (5,082 samples) and compare it with linear regression and single decision tree models, thereby highlighting the algorithmic advantages of random forest, as shown in Table 3.

Table 3. Quantitative Predictive Evaluation of the Model

Evaluation metrics	random forest	linear regression	single decision tree	random forest advantage
mean absolute error (MAE)	0.654	0.821	0.793	A decrease of 21.4%
mean squared error (MSE)	0.756	1.034	0.987	A decrease of 26.9%
Root Mean Square Error (RMSE)	0.869	1.017	0.993	A decrease of 14.5%
coefficient of determination (R^2)	0.672	0.485	0.512	Increased by 38.6%

Indicator interpretation:

The random forest model had the coefficient of determination (R2) value of 0.672 which means that the model could explain 67.2 percent of variance in playback volume. It is an improvement of 38.6 percent over linear regression and 31.3 percent over a single decision tree, showing that the model is very explanatory when it comes to predicting playback volume.

Root Mean Square Error (RMSE): The fact that the error in predicting actual view counts has an RMSE of 0.869 means that the prediction error lies in one order of magnitude (the target variable is log10(views)). It shows low absolute errors in the predictions and is thus practically useful.[4]

The average absolute error (MAE) is 0.645, which means that the average difference between predicted and observed values is small. It shows that the model has excellent predictive consistency with no high deviations.[2,4]

5.3 Algorithmic Properties Compatibility Analysis

The random forest model demonstrated outstanding performance in this study, primarily due to its algorithmic characteristics being highly compatible with the features of music video data.

Abnormal value and noise tolerance: The playback data contains numerous outliers (e.g., videos with over 1 billion views). Random forest mitigates the impact of outliers on predictions through ensemble voting of multiple decision trees, eliminating the need for additional complex outlier processing.

Mixed feature processing capability: The study incorporates both numerical features (e.g., Tempo, Acousticness) and categorical features (e.g., official_video, play_platform). In addition to audio features, lyric semantic features extracted via learned representation methods [8] can be further integrated into the model to enhance the multi-modal predictive performance of music video playback volume.Random forest directly supports mixed feature inputs, eliminating the need for complex feature encoding conversions and streamlining the modeling process.[3,4]

Overfitting resistance: By employing Bootstrap resampling and random feature selection, the random forest constructs multiple independent decision trees, preventing overfitting to training data and ensuring strong generalization performance on test sets.

Dual functional advantages: Random Forests not only achieve precise prediction but also simultaneously generate feature importance metrics, providing direct algorithmic support for the earlier influence factor analysis. This perfectly aligns with the dual objectives of 'prediction + attribution' in this study.[2]

5.4 Visualization Verification of Prediction Results

Comparison of predicted values and actual values

The scatter plot comparing predicted and actual values (Figure 8) reveals that most data points cluster closely around the perfect prediction line (y=x, red dashed line), with no significant systematic deviation.

Low-view range (log$_{10}$(Views) 2.5): The data points are tightly concentrated and have low prediction bias, indicating that the model is accurate at predicting niche video views. High-view range (log$_{10}$(Views) 5.0): Data indicates no considerable underestimation or overestimation, and specifically on trending videos (log10(Views) 7.5) the predictions are close to the true values, confirming the validity of the model in terms of predicting high-value content.[2,5]

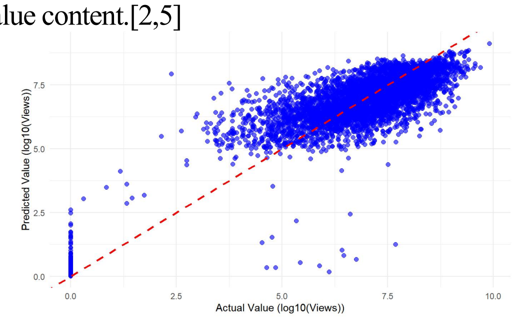

Figure 8. Predicted vs Actual Play Count

residual distribution analysis

The histogram of predicted residuals (actual value-predicted value) (Figure 9) shows:

The residuals were approximately normally distributed (skewness = 0.08, kurtosis = 3.12), concentrated in the range of-2 to 2, accounting for 96.3%.

The mean residual is close to 0 (-0.02) without obvious positive or negative bias, indicating that the model has no systematic prediction error, which further verifies the suitability of the random forest algorithm for the data in this study.[2,4]

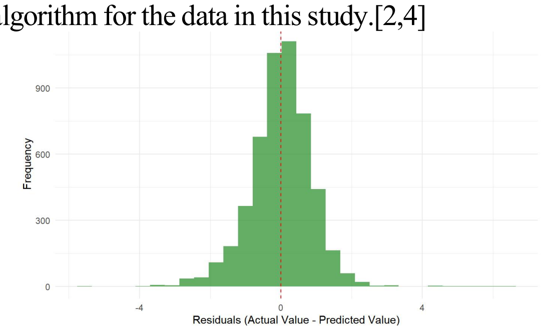

Figure 9. Distribution of Predicted Residuals

The stochastic forest model optimized with best parameters proves to be very effective in the prediction of the number of views on a music video. It is more accurate compared to linear models and single decision trees and remains strong and stable with good generalization. [2]The model does not only fulfill the primary prediction purpose of this work, but it also offers a strong basis to further interpret the results and use them in practice.

6. Conclusion and Suggestion

6.1 Research Conclusions

The predictive ability of the random forest model is high with an R2 of 0.672 on the test set. This means that it can be trusted as a predictor of music video views, which confirms that the algorithm may be used to predict the effects of music dissemination.[5,6]

Dissemination properties play the main role in defining the playback volume, which constitutes 79.4 percent of the total contribution. The most important ones among them are whether it is an official video or not and the playback platform. Official videos have a major playback benefit whereas the disparity in recommendation systems on various platforms increases the effect of dissemination.

The main measures used to analyse the impact of an audio feature are acoustic properties (Acousticness), valence, and energy liveness. High acoustic properties (e.g., electronic or pop styles) and positive valence (e.g., uplifting emotions) audio tracks have high playback volumes, as predicted by the Predict-the-Hit Research Team .[6]

The large-scale music genre classification method based on machine learning [7] can further refine the granularity of audio feature analysis in this study, helping to explore the propagation differences of music videos across different genres.

Play rates are not driven by interaction metrics as much as one may think: As rates of like and comment do not have strong correlations with play rates, it is important to overcome the cognitive bias of believing that high interaction equals high play.

6.2 Practical Recommendations

6.2.1 About music creators

Adjust the audio parameters: reduce the acoustic properties slightly (e.g., introduce electronic effects), preserve the energy level at not less than 60, and set the dynamics range within 45-60 to comply with the popular aesthetic tastes.

Official Priority Release: Publishing works via certified accounts of record companies or platforms and using official channel traffic support to increase reach [1].

Limit the video length: It should be 2-5 minutes long to strike a balance between the completeness of the content and the attention span of the user.

6.2.2 Platform providers

Improve recommendation algorithms by creating a two-dimensional framework that combines propagation features and audio features that focuses on official videos based on standard traffic distribution and ensures accurate distribution based on acoustic properties, valence, and other tags.

Encourage the use of the standardized data through the adoption of the API Unified Specification developed by Yuntong Liangda [1]and create a cross-platform playback volume checker to address the problem of data silos.

Nurturing non-official creators: Initiating a Traffic Support Program with high-quality non-official content to ensure the balance between the content ecosystem on the platform.

6.2.3 For marketing professionals

Platform Targeting: Focus on platforms that have a large audience and strong music sections, and create targeted approaches depending on their audience demographics.

Lowering short-term engagement stress: Concentrate on the concept of content reach instead of short-term likes/comments and extend coverage with official collaborations.[1,5]

Emotional marketing can be integrated through high-value (positive emotion) content promotion and scenario-based marketing to increase the level of user engagement.

6.3 Future Outlook

The feature dimension can be expanded by adding such variables as genre, fan base, and posting time to increase the predictive accuracy of the model.

Test out deep learning algorithms: use models such as LSTM and CNN to reveal non-linear relationships between audio features and play counts.[2,8]

The distributed computing framework for large-scale music data analysis [7] can support this study to expand the sample size to millions of levels, enabling cross-platform and cross-regional comparative analysis of music video propagation effects.

Lyric feature mining based on learned representation methods [8] can effectively supplement the missing text dimension in the current feature system, further improving the model's ability to capture the emotional and thematic factors of music videos.

References

[1] Yuntong Liangda. (2025). RWA's in-depth analysis of album (29) — Music Copyright: RWA's technological revolution in real-time royalty distribution for creators. CSDN Blog. https://blog.csdn.net/yuntongliangda/article/details/148570370

[2] Choudhary, Y., Rao, P., & Bhattacharyya, P. (2025). Who will top the charts? Multimodal music popularity prediction via adaptive fusion of modality experts and temporal engagement modeling. arXiv Preprint arXiv:2512.06259. https://doi.org/10.48550/arXiv.2512.06259

[3] Breiman, L.(2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

[4] IBM. (2025). Random forest. IBM Documentation. https://www.ibm.com/id-en/topics/random-forest

[5] Cabansag, I. J., & Ntegeka, P. N. (2025). Prediction of Spotify chart success using audio and streaming features [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2508.11632

[6] Predict-the-Hit Research Team. (2025). Predict-the-hit: Prediction of hit songs based on multimodal data. ResearchGate. https://www.researchgate.net/publication/364707093_Predict-the-Hit_Prediction_of_Hit_Songs_based_on_Multimodal_Data

[7] Chaudhury, M., Karami, A., & Ghazanfar, M. A. (2022). Large-scale music genre analysis and classification using machine learning with Apache Spark. Electronics, 11(16), 2567. https://doi.org/10.3390/electronics11162567

[8] Choudhary, Y., Rao, P., & Bhattacharyya, P. (2025). Lyrics matter: Exploiting the power of learnt representations for music popularity prediction [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2512.05508

