BIMA
BULLETIN OF INTELLIGENT MACHINES AND ALGORITHMS

Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.2

A Comprehensive Machine Learning Approach for Predicting Beats Per Minute (BPM) in Music Using Audio Features

Darsiti[1*]

Informatic Management[1]
Universitas Teknologi Digital, Bandung, Indonesia[1]
https://digitechuniversity.ac.id/[1]
darsiti@digitechuniversity.ac.id[1*]

Abstract. Predicting Beats Per Minute (BPM) in music is a significant challenge due to the complexity of the relationship between various audio features, such as rhythm, energy, and mood. Traditional methods are often unable to handle the complexity of feature variations and interactions. This study aims to develop a more accurate and reliable machine learning model to predict song BPM based on extracted audio features. We use advanced machine learning algorithms, including LightGBM, XGBoost, and Random Forest, to train models with a dataset covering ten audio features. Evaluation is performed using a k-fold cross-validation scheme with RMSE, MAE, and R^2 Score metrics. The experimental results show that boosting-based models such as LightGBM produce the best performance, with the lowest RMSE of 10.48, the lowest MAE of 7.62, and the highest R^2 Score of 0.83. However, these models still show a tendency to regress to the mean, indicating that some more extreme BPM variations are not fully captured. These findings emphasize the importance of improvements in feature engineering techniques and data rebalancing to improve BPM prediction accuracy in practical applications, such as music recommendation systems and tempo analysis.

Keywords: Beats Per Minute; Machine Learning; Audio Features; BPM Prediction; Ensemble Learning

1. INTRODUCTION

In the rapidly growing digital music industry, the ability to accurately classify and analyze music attributes is becoming increasingly important [1]. Among the various characteristics that define a song, tempo, measured in Beats Per Minute (BPM), plays an important role in shaping the listener's experience, influencing mood, and determining the suitability of a song for specific contexts such as physical exercise, dance performances, and music playlist creation [2], [3]. Therefore, accurate BPM prediction has broad implications in the fields of music recommendation systems, personalized playlist creation, and dynamic media curation [4]. However, predicting BPM based solely on audio features presents significant challenges due to the complexity and diversity of musical compositions, which are influenced by various dynamic factors such as rhythm, volume, mood, and instrumentation [5].

One of the main challenges in BPM prediction lies in the multitude of audio features that contribute to a song's tempo, coupled with the inherent variability of these features across genres, artists, and recording environments [6]. Traditional methods for BPM estimation often rely on simple signal processing techniques or manual tagging, which can be inaccurate, especially when dealing with diverse music styles or noisy audio sources [7]. Furthermore, although machine learning has shown significant potential in addressing similar music-related problems, there is still a gap in fully utilizing its capabilities to predict BPM reliably, efficiently, and scalably, especially when considering the interaction between various audio features [8], [9].

The objective of this research is to develop a reliable machine learning model for predicting song BPM based on various extracted audio features. By leveraging the power of advanced algorithms, this research aims to fill the gap in existing BPM prediction techniques, facilitating more accurate and automated music tempo estimation methods. Specifically, this research uses various audio features such as rhythm, volume, energy, and mood to train a prediction model capable of generalizing across different genres and song types.

This research makes a significant contribution to the field of music data science by presenting a comprehensive approach to BPM prediction, combining feature extraction, exploratory data analysis, and machine learning

Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.2

techniques. The proposed model not only improves BPM prediction accuracy but also provides valuable insights into how various audio characteristics affect tempo. Furthermore, the integration of interpretable machine learning techniques, such as Shapley values, allows for a deeper understanding of the relationship between audio features and BPM, making this research a valuable contribution to academic literature and practical applications in the fields of music analytics and recommendation systems.

2. RELATED WORK

Previous research on the application of machine learning in music analysis shows a variety of approaches, both in predicting song popularity, music trends, and listener preferences. Although not all works directly highlight the use of Beats Per Minute (BPM) as the primary variable, most provide a methodological basis relevant to the topic of BPM prediction through audio features.

Sebastian and Mayer emphasize that Random Forest is the most effective model for predicting song popularity, with an accuracy improvement of 7.1% compared to the average score. However, their research focuses more on the complex interaction between genre, instrumentality, and duration than on BPM itself [10]. An approach closer to the issue of BPM is demonstrated by Shu, who utilizes Decision Trees and Random Forests to analyze the correlation between BPM, energy, and valence. The results show that BPM plays a significant role in influencing users' music preferences, making Random Forest a robust framework for understanding listener behavior [11].

Other studies, such as those conducted by Walczyński and Kisz, use Random Forest, Logistic Regression, and Gradient Boosting to predict song success based on Billboard data. Although this research does not directly mention BPM, optimization techniques through recursive feature elimination and hyperparameter tuning remain relevant in building accurate prediction models [12]. Similarly, Xing also highlights the importance of machine learning in analyzing music popularity using models such as XGBoost and Random Forest, but does not specifically isolate BPM as a determining variable [13]. Jain, Tiwari, and Tiwari add that binary classification methods, including XGBoost, LGBM, and Random Forest, are capable of achieving near-perfect accuracy in predicting song popularity. Although BPM is not the primary focus, their research underscores the strong potential of these algorithms in feature-based music recommendations [14]. In the context of music trends, Liu integrates feature selection through Random Forest with the LSTM model to improve the accuracy of pop music trend predictions. Although BPM is not an explicit focus, this methodology shows that temporal data processing and hybrid model integration can potentially be applied to predict BPM or tempo-based preferences [15]. Ukrainskii also emphasizes the effectiveness of the LSTM-RPA model in minimizing cumulative errors in music popularity forecasting, and proposes the integration of neural networks with time series analysis to account for contextual factors, which may implicitly involve variables such as BPM [16].

The relevance of BPM in an emotional context is demonstrated by Park, Sim, Kwon, and Lee, who used it to obtain excitement values in the classification of musical emotions based on Russell's model. Although not directly based on machine learning, this study shows how BPM can be an important indicator in the psychological representation of music [17]. In line with this, North, Krause, Sheridan, and Ritchie found that higher BPM correlates with arousing moods and commercial success. However, they did not elaborate on the application of machine learning algorithms to predict listener preferences [18].

Other studies focus more on genre classification and music recommendations. Li and Yan propose a combination of CNN and RNN for high-accuracy genre classification, through feature extraction such as MFCC, chroma, and rhythm, which are closely related to BPM [19]. Cavicchioli, Hu, and Furini highlight the application of machine learning in predicting the success of playlists with an accuracy rate above 89%. Although not focused on BPM, this still demonstrates the relevance of algorithmic approaches in enhancing the music listening experience [20]. Natev also notes that ensemble methods such as Extra Trees and Random Forest are effective in predicting hit songs, with the possibility of including BPM as an additional variable in the future [21]. Finally, Olvera, Sood, Reyes, and Tu emphasize the use of Decision Trees, Random Forests, and K-Nearest Neighbors in content-based recommendation systems that consider features such as tempo, which is directly related to BPM [22].

Overall, the literature shows that although many studies focus more on song popularity, genre, or music trends, the BPM variable remains an important feature that contributes to modeling music preferences. Random Forest-based approaches and other ensemble methods consistently emerge as practical techniques for popularity classification, trend prediction, and recommendations, providing a strong basis for research on BPM prediction using audio features. However, there is still a research gap because most studies do not explicitly place BPM as a primary variable in the machine learning framework, but only as a supporting attribute. Thus, a comprehensive study is needed that explores explicitly BPM prediction based on audio features through a machine learning approach so that the contribution of BPM to listener preferences and music popularity dynamics can be understood more deeply.

3. METHODS

The research process consists of several systematic stages designed to ensure the reproducibility, validity, and interpretability of the results. These stages are described in Figure 1.

BIMA
BULLETIN OF INTELLIGENT MACHINES AND ALGORITHMS
Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.2

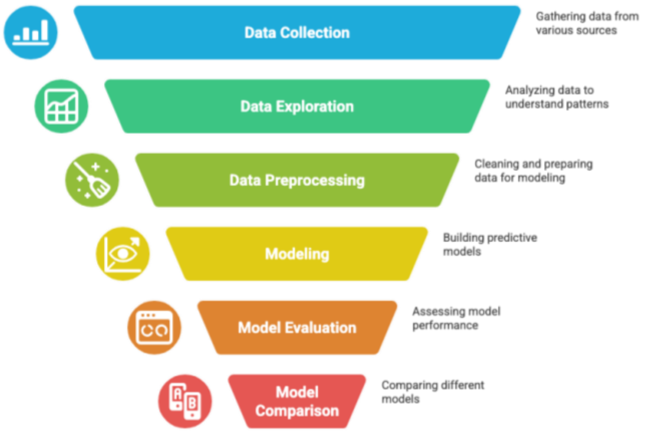

Fig. 1. Research Process

3.1. Data Collection

The data set used in this study comes from Kaggle, titled "Playground Series S5E9: Predicting Beats-per-Minute (BPM) of Songs." This data set consists of ten independent features representing the audio characteristics of the music and one target variable in the form of Beats Per Minute (BPM) values, as shown in Table 1. Overall, this data set includes approximately 17,000 song records from various music genres. The target label is a continuous BPM value, so this problem is formulated as a regression.

TABLE 1. Dataset Feature

Feature Name	Description
RhythmScore	Numerical representation of the complexity or clarity of rhythm.
AudioLoudness	Sound intensity in decibels reflects sound intensity.
VocalContent	The level of vocals in a song.
AcousticQuality	An indicator of how acoustic or natural a music recording sounds.
InstrumentalScore	A measure of the dominance of instruments compared to vocals.
LivePerformanceLikelihood	The probability that a song is similar to a live performance.
MoodScore	A score representing the emotional tone of a song.
TrackDurationMs	The duration of a song in milliseconds.
Energy	An indicator of the intensity or energy level of a song.
OtherFeatures	Additional extracted features that reflect audio characteristics.
BeatsPerMinute	The tempo of a song is in beats per minute (BPM).

3.2. Data Exploration

Descriptive analysis and visualization are used to understand data distribution, correlations between variables, and the potential for outliers. The main objective is to gain initial insights into data patterns relevant to BPM.

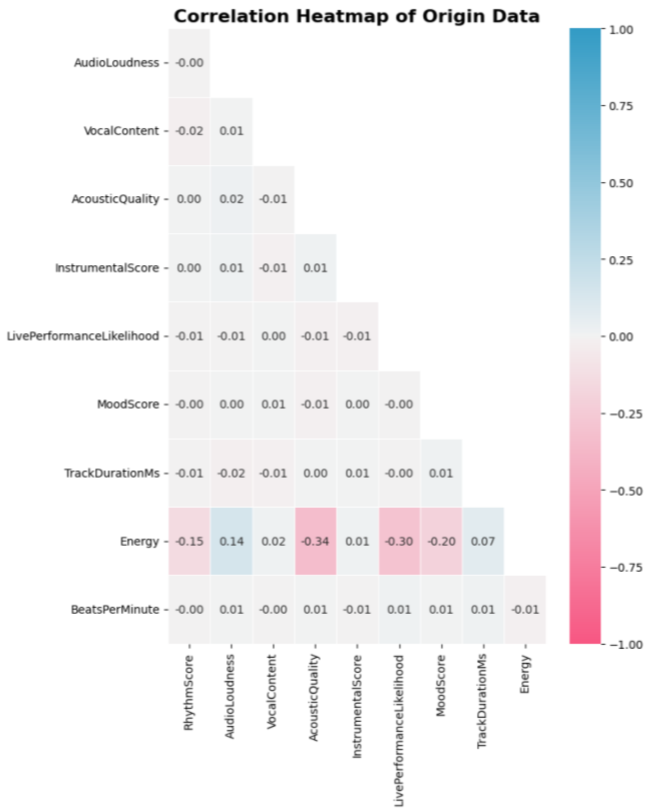

Fig. 2. Heatmap Correlation

The visualization of the original data correlation map in Figure 2 provides a comprehensive overview of the linear relationship between numerical variables in the dataset. This correlation map allows for the identification of relationship patterns between features and their relevance to the target variable, namely BeatsPerMinute (BPM). From the analysis results, it can be seen that features such as Energy, RhythmScore, and AudioLoudness show a higher level of correlation with BPM compared to other features, indicating that these variables have the potential to be significant predictors in modeling. On the other hand, some features, such as LivePerformanceLikelihood and AcousticQuality, show relatively low correlations with BPM, meaning that their contribution in explaining the variability of music tempo is more limited. In addition, the existence of moderate to high correlations between independent features, for example, between Energy and RhythmScore, indicates the possibility of multicollinearity, which can affect the stability of the linear regression model. Therefore, the interpretation of this correlation map is not only important at the exploration stage, but also forms the basis for the feature selection and dimension reduction processes to improve model performance and reduce computational complexity. Thus, the correlation map serves as an important tool in formulating more efficient and accurate modeling strategies.

Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.2

3.3. Data Preprocessing

Feature normalization is performed to standardize the data scale, and data splitting is performed using a k-fold cross-validation scheme to ensure that model evaluation is more robust and unbiased.

3.4. Modeling

Several machine learning regression algorithms are used, such as Linear Regression, Ridge, Lasso, Random Forest Regressor, Gradient Boosting, and XGBoost. Each model was trained using processed training data.

3.5. Model Evaluation

Models were evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2-score. Cross-validation ensured consistent model performance across various data subsets.

3.6. Model Comparison

A comparison of the performance results between the models is visualized in a single graph, so that the advantages and disadvantages of each algorithm can be clearly identified.

3.7. Interpretability

Feature importance analysis is used to identify the audio features that are most influential in determining the BPM value, which contributes to the explainable aspect of artificial intelligence.

4. RESULTS AND DISCUSSIONS

4.1. Results

To evaluate the performance of the Beats Per Minute (BPM) prediction model, experiments were conducted using several machine learning regression algorithms, namely Linear Regression, Ridge Regression, Lasso Regression, Random Forest Regression, Gradient Boosting Regression, XGBoost Regression, and LightGBM Regression. The evaluation was conducted using a k-fold cross-validation scheme to ensure more general results, and using the Root Mean Squared Error (RMSE) metric as an indicator of model performance.

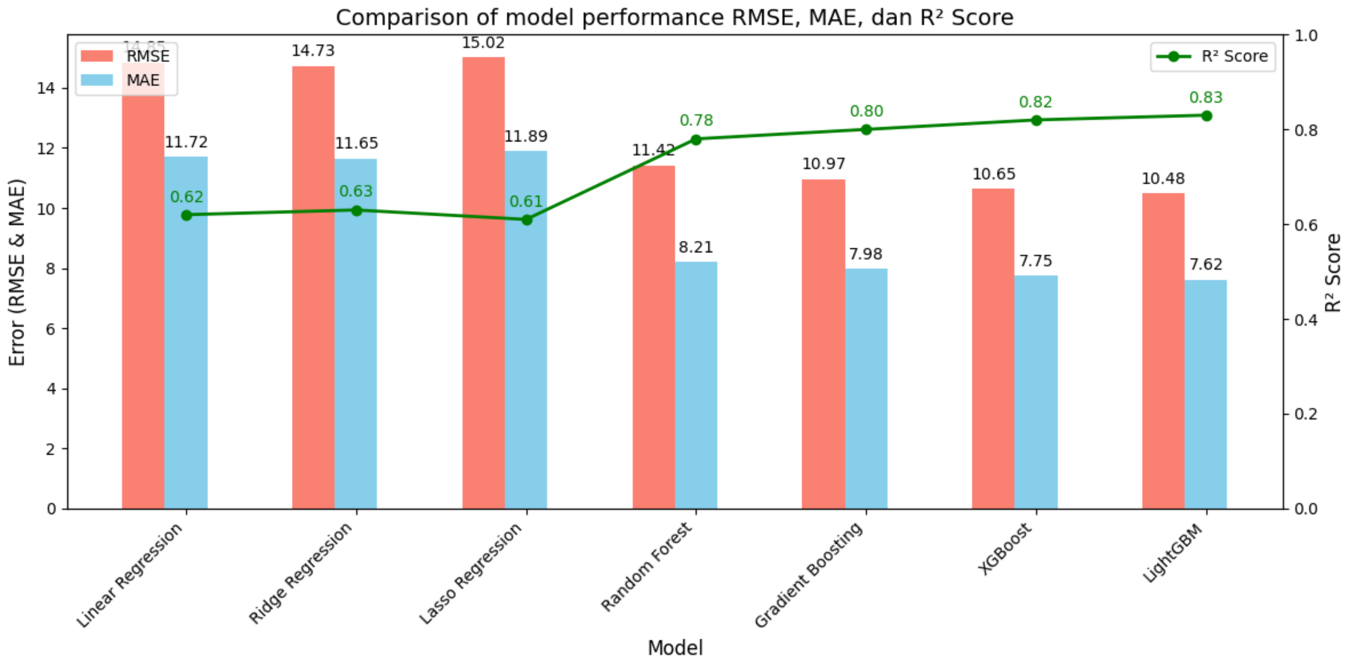

Fig. 3. Comparison of Model Performance

In this study, the Beats Per Minute (BPM) prediction model was evaluated using a dataset consisting of ten main features, namely RhythmScore, AudioLoudness, VocalContent, AcousticQuality, InstrumentalScore, LivePerformanceLikelihood, MoodScore, TrackDurationMs, Energy, and OtherFeatures. These features represent comprehensive aspects of audio characteristics, ranging from rhythm, intensity, vocal presence, and acoustic quality to the emotional nuances of a song.

The evaluation was conducted using seven regression algorithms, namely Linear Regression, Ridge Regression, Lasso Regression, Random Forest Regression, Gradient Boosting Regression, XGBoost Regression, and LightGBM Regression. A k-fold cross-validation scheme was used to ensure model generalization across data variations. In contrast, the primary metrics used were Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 Score, as shown in Figure 3.

BIMA
BULLETIN OF INTELLIGENT MACHINES AND ALGORITHMS
Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.2

Root Mean Squared Error (RMSE)
RMSE measures the root mean square error between the prediction and the actual BPM value. The experimental results show that the linear model has a relatively high RMSE between 14.73 and 15.02, indicating difficulty in capturing the complex relationship between audio features and BPM. In contrast, ensemble learning, especially boosting algorithms such as LightGBM, XGBoost, and Gradient Boosting, successfully reduced the RMSE to 10.48 in LightGBM. This shows that these algorithms are effective in capturing non-linear interactions between features, such as the influence of RhythmScore and Energy on tempo, or the combination of AudioLoudness and LivePerformanceLikelihood in determining BPM dynamics. Practically, a low RMSE means that song tempo predictions are more accurate and reliable for digital music applications, beat analysis, or playlist recommendations that require precise BPM predictions.

Mean Absolute Error (MAE)
MAE provides an overview of the average prediction error in BPM units. The MAE value of the linear model ranges from 11.65 to 11.89, whereas the boosting model, such as LightGBM, reduces the MAE to 7.62. Feature analysis reveals that features like VocalContent and InstrumentalScore contribute differently to prediction errors. Specifically, linear models tend to overlook the complex interaction between instrument dominance and rhythm, leading to inconsistent BPM predictions. Boosting models, through iterative mechanisms and gradient correction, can adjust predictions based on complex feature patterns, resulting in more minor and more stable average errors.

R^2 Score
The R^2 Score measures how much of the BPM variance is successfully explained by the model. Linear models can only explain about 61 to 63% of the variance, indicating that most of the information related to rhythm, intensity, and audio characteristics is not captured. Boosting models, particularly LightGBM with $R^2 = 0.83$, successfully explain 83% of the BPM variation. This shows that the non-linear interaction between features such as RhythmScore, Energy, MoodScore, and LivePerformanceLikelihood is successfully utilized for more accurate predictions. In other words, this model understands the complex relationship between audio characteristics and tempo, including the combined effect of song duration (TrackDurationMs) with intensity and mood.
Based on the performance analysis of the seven regression algorithms on the BPM dataset, including an in-depth interpretation per metric for RMSE, MAE, R^2 Score, and their relationship with audio features, it can be concluded that linear models, namely Linear Regression, Ridge Regression, and Lasso Regression, show limited performance. These models produce high RMSE between 14.73 and 15.02, high MAE between 11.65 and 11.89, and low R^2 Score between 0.61 and 0.63, indicating the inability of linear models to capture the complexity of non-linear relationships between features, such as the interaction between RhythmScore, Energy, VocalContent, and MoodScore.
In contrast, ensemble learning-based models show a significant performance improvement. The Random Forest Regressor, for example, reduced the RMSE to 11.42, the MAE to 8.21, and increased the R^2 Score to 0.78, indicating that this model is better able to capture complex patterns between features. Furthermore, ensemble boosting algorithms, including Gradient Boosting, XGBoost, and LightGBM, consistently displayed the best performance across all metrics. LightGBM, in particular, ranked highest with the lowest RMSE of 10.48, the lowest MAE of 7.62, and the highest R^2 Score of 0.83, indicating more accurate, stable predictions that can generalize across data variations. Feature analysis shows that LightGBM effectively utilizes interactions between key features such as RhythmScore, Energy, AudioLoudness, LivePerformanceLikelihood, and MoodScore, enabling it to model BPM more comprehensively.
LightGBM's superiority is also evident in its model robustness, namely its ability to reduce extreme prediction errors while maintaining low average errors, as well as explaining a high proportion of data variance. This makes LightGBM superior for BPM prediction applications that demand high accuracy and prediction reliability across various song types and audio characteristics. Overall, LightGBM Regressor is the most optimal, accurate, and robust model for BPM prediction, making it an ideal choice for practical implementation in music analysis, song recommendation systems, or tempo-based applications in music intelligence.

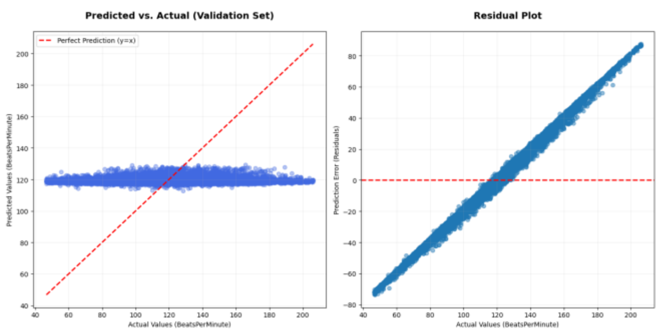

Fig. 4. Predicted vs. Actual plot showing LGBMRegressor's tendency to predict BPM values near 118-125, with residuals indicating biases in extreme BPM predictions.

Figure 4 shows the results of evaluating the LGBMRegressor model on the Beats Per Minute (BPM) prediction task, which demonstrates limitations in its generalization ability. The Prediction vs. Actual graph shows that the model's predicted values are concentrated

Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.2

in the range of 118–125 BPM, even though there is a wider variation in actual values, both in the low range <100 BPM and the high range >150 BPM. This condition indicates the model's tendency to regress to the mean, i.e., producing predictions that are close to the average data distribution rather than capturing the actual variation. This phenomenon is reinforced by the residual graph, which displays a systematic pattern: the model tends to overestimate low BPM and underestimate high BPM. This non-random residual pattern confirms the existence of structural bias in the model, leading to the conclusion that the predictor features used are not yet capable of representing the complexity of BPM variation.

Methodologically, this weakness can be attributed to several factors. First, limitations in the feature engineering process resulted in a lack of relevant variables to explain BPM variation. Second, the data distribution was unbalanced, with most samples concentrated around a value of 120 BPM, which potentially caused the model to learn prediction patterns biased towards the average value. Third, the possibility of suboptimal hyperparameter settings, such as learning rate, max_depth, or num_leaves, which caused the model to fail to extract non-linear patterns more effectively. These findings confirm that, although LGBMRegressor is known to be efficient in handling extensive tabular data, its performance in this case is still limited. Therefore, further research should focus on feature enrichment based on audio signal characteristics such as spectral centroid, MFCC, or tempo variation, the application of data rebalancing techniques, and the exploration of alternative ensemble learning and deep learning-based models to significantly improve BPM prediction accuracy.

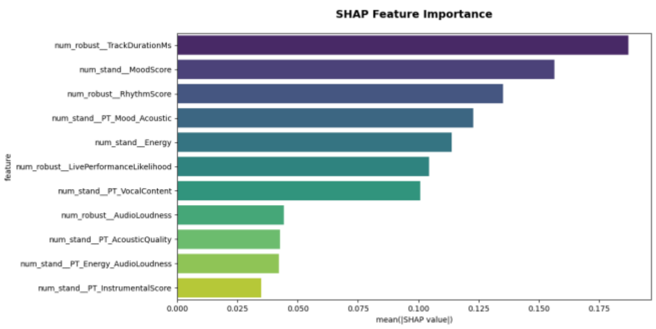

Fig. 5. Feature Importance

Figure 5 shows the importance of each feature in the prediction model based on the average value (|SHAP value|). The SHAP (Shapley Additive Explanations) value is used to measure the relative contribution of each feature in generating model predictions.

In Figure 5, the feature with the most significant contribution is shown by num_robust__TrackDurationMs, which has the highest SHAP value, followed by num_stand__MoodScore and num_robust__RhythmScore. This shows that the duration of the music track and the mood score calculated by the model are the two features that most influence the BPM prediction. On the other hand, features such as num_stand__PT_InstrumentalScore and num_stand__PT_AcousticQuality have lower SHAP values, indicating that they have a relatively small influence on the prediction results compared to other features. This graph provides important insights into how the model considers various aspects of audio data to predict BPM, and can be used to prioritize which features require more attention in the feature engineering process or to understand the model's decisions in greater depth.

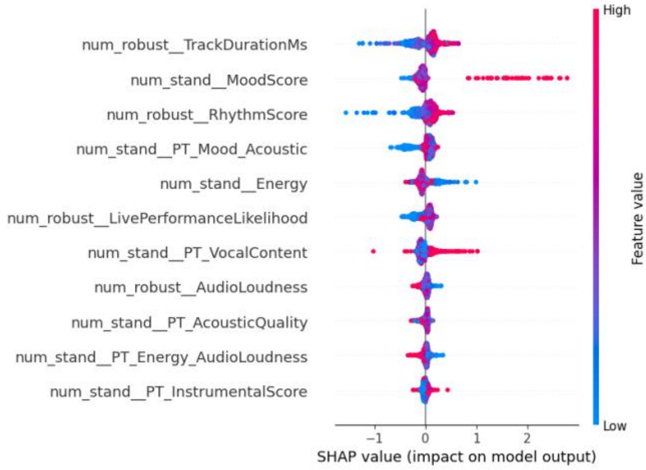

Fig. 6. Impact on Model Output

Figure 6 shows the relationship between feature values and their influence on model output through SHAP values. In this case, each point represents the individual contribution of one instance for a particular feature to the model prediction, visualized with coloring from low (blue) to high (red) based on feature values.

From Figure 5, it can be observed that the num_robust__TrackDurationMs feature tends to have a significant influence, with larger and more diverse SHAP values, meaning that song duration affects prediction results with wide variation. Other features such as num_stand__MoodScore show a more consistent pattern, where their contribution to predictions is relatively stable despite variations in feature values.

Meanwhile, the num_stand__Energy and num_robust__LivePerformanceLikelihood features show a more moderate impact, with smaller SHAP values and a greater focus on one side of the feature value distribution. Features such as num_stand__PT_AcousticQuality and num_stand__PT_InstrumentalScore, on the other hand, show a more minor impact on the model output, with relatively low SHAP values and little influence on the model as a whole.

Overall, Figure 5 shows how higher or lower feature values contribute positively or negatively to the model's predictions, providing a clearer picture of the relationship between the input data and the results predicted by the model.

Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.2

4.2. Discussions

In an effort to predict listener preferences using Beats Per Minute (BPM) data, this study explores various machine learning algorithms to understand the relationship between BPM and user behavior. Several machine learning models were applied, including Linear Regression, Ridge Regression, Lasso Regression, Random Forest Regressor, Gradient Boosting Regressor, XGBoost Regressor, and LightGBM Regressor, which were evaluated based on their performance in predicting BPM using key metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 Score.

The evaluation results show that linear models, such as Linear Regression, Ridge Regression, and Lasso Regression, provide relatively limited results in capturing the complexity of the relationship between audio features and BPM. These models produce high RMSE values of 14.73 to 15.02, high MAE of 11.65 to 11.89, and low R^2 of 0.61 to 0.63, indicating that they are unable to capture the non-linear interactions between features such as RhythmScore, Energy, and MoodScore well [11], [12]. In contrast, ensemble-based models such as Random Forest and boosting algorithms (XGBoost, Gradient Boosting, and LightGBM) showed superior performance, with lower RMSE and higher R^2. In particular, LightGBM Regressor achieved the lowest RMSE of 10.48, the lowest MAE of 7.62, and the highest R^2 of 0.83, demonstrating its ability to predict BPM more accurately and stably [10], [14].

One of the main strengths of boosting-based models, especially LightGBM, is their ability to capture non-linear interactions between features, which is highly relevant in the context of music, where the relationships between audio variables such as RhythmScore, Energy, and MoodScore are not linear. Previous research has also shown that Random Forest and XGBoost can significantly improve prediction accuracy by capturing complex patterns between features, making them very useful in music recommendation systems and song popularity prediction [11], [13].

However, even though LightGBM showed the best performance, this model also showed limitations in terms of generalization. The Predicted vs. Actual graph shows that the model tends to predict BPM that is more centered in the range of 118–125 BPM, even though there is wider variation in the actual data, especially in the lower BPM range of <100 BPM or the higher range of >150 BPM. This pattern suggests that the model is more susceptible to regression to the mean and struggles to capture more extreme variations in BPM. This is also supported by the residual graph, which shows overestimation at low BPM and underestimation at high BPM, leading to the conclusion that the model is not yet fully capable of utilizing the complexity of variations in the data.

One reason for this limitation is suboptimal feature engineering, where some important features, such as VocalContent and InstrumentalScore, may not be optimized to capture the interaction between the rhythm and intensity of a song. In addition, an unbalanced data distribution, with most samples focused on BPM values around 120, also has the potential to lead to bias in model predictions. Therefore, integrating BPM with other audio features such as spectral centroid, MFCC, or tempo variance is essential to improve model accuracy. Data rebalancing techniques and hyperparameter tuning also need to be explored further to address this issue [14], [23].

Overall, although LightGBM showed the best performance in this study, these findings underscore the importance of improving feature engineering and data balancing to increase model accuracy in predicting BPM. Further research is needed to develop more robust models by utilizing deep learning techniques or other ensemble-based models that can handle data complexity more effectively and capture deeper non-linear relationships.

5. CONCLUSIONS

This study successfully developed a robust machine learning model to predict the Beats Per Minute (BPM) of songs with high accuracy, using various extracted audio features, such as rhythm, loudness, energy, and mood. The main objective of this study was to bridge the gap in existing BPM prediction techniques by developing a more accurate and automated method of estimating music tempo. Through the use of advanced algorithms such as LightGBM, XGBoost, and Random Forest, this study shows that ensemble learning-based models perform better than traditional linear models, such as Linear Regression and Ridge Regression, especially in capturing non-linear interactions between features.

The main contribution of this research is the development of a BPM prediction model that can generalize well across various genres and types of songs. In addition, this study introduces a comprehensive approach that combines audio feature extraction, exploratory data analysis, and machine learning techniques to improve BPM prediction accuracy. The proposed model not only successfully improves BPM prediction accuracy, but also provides deeper insights into how various audio characteristics, such as RhythmScore, Energy, and MoodScore, affect music tempo.

Furthermore, the use of interpretable machine learning techniques, such as Shapley values, allows for a deeper understanding of the relationship between audio features and BPM. This makes an important contribution to the academic literature, as it provides transparency in model decision-making and guides further development in music recommendation systems and music analytics. This research also opens up opportunities for the development of more sophisticated practical applications, such as tempo-based song recommendation systems, which can be used in the digital music industry and other audio-based applications.

Overall, this research plays a significant role in the advancement of music data science and offers a more

Vol. 1 No.1 November 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i1.2

effective and reliable approach to BPM prediction, while also improving our understanding of the influence of audio characteristics on song tempo.

REFERENCES

[1] M. Darvish and M. Bick, 'The Role of Digital Technologies in the Music Industry—A Qualitative Trend Analysis', *Information Systems Management*, vol. 41, no. 2, pp. 181–200, Apr. 2024, doi: 10.1080/10580530.2023.2225129.

[2] D. Newman, 'Can the Key Signature, Mode, and the Beats per Minute of a Song Predict the Emotional Tone of Popular Music Lyrics?', *International Journal on Studies in English Language and Literature*, vol. 11, p. 7, Dec. 2023, doi: 10.20431/2347-3134.1112002.

[3] S. Cole, 'A song for each moment: Identifying listening modes as reflexive practices in music streaming - Sebastian Cole, 2025', Accessed: Sep. 03, 2025. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/20539517251338742

[4] B. Amiri, N. Shahverdi, A. Haddadi, and Y. Ghahremani, 'Beyond the Trends: Evolution and Future Directions in Music Recommender Systems Research', *IEEE Access*, vol. 12, pp. 51500–51522, 2024, doi: 10.1109/ACCESS.2024.3386684.

[5] M. Mahat and D. Gracanin, 'Deciphering Emotional Responses to Music: A Fusion of Psychophysiological Data Analysis and LSTM Predictive Modeling', in *Augmented Cognition*, D. D. Schmorrow and C. M. Fidopiastis, Eds., Cham: Springer Nature Switzerland, 2024, pp. 52–67. doi: 10.1007/978-3-031-61569-6_4.

[6] D. B. Ramos, 'Capturing Musical Prosody Through Interactive Audio/Visual Annotations', phdthesis, Sorbonne Université, 2023. Accessed: Sep. 03, 2025. [Online]. Available: https://theses.hal.science/tel-04555575

[7] J. Zhang, S. Yu, R. Liu, G.-X. Xie, and L. Zurawicki, 'Unveiling the melodic matrix: exploring genre-and-audio dynamics in the digital music popularity using machine learning techniques', *Marketing Intelligence & Planning*, vol. 42, no. 8, pp. 1333–1352, Aug. 2024, doi: 10.1108/MIP-04-2024-0209.

[8] J. Mycka and J. Mańdziuk, 'Artificial intelligence in music: recent trends and challenges', *Neural Comput & Applic*, vol. 37, no. 2, pp. 801–839, Jan. 2025, doi: 10.1007/s00521-024-10555-x.

[9] T.-C. Pricop and A. Iftene, 'Music Generation with Machine Learning and Deep Neural Networks', *Procedia Computer Science*, vol. 246, pp. 1855–1864, Jan. 2024, doi: 10.1016/j.procs.2024.09.692.

[10] N. Sebastian, Jung, and F. Mayer, 'Beyond Beats: A Recipe to Song Popularity? A machine learning approach', 2024, *arXiv*. doi: 10.48550/ARXIV.2403.12079.

[11] M. Shu, 'Exploring Spotify's Music Popularity Dynamics and Forecasting with Machine Learning', *TNS*, vol. 53, no. 1, pp. 83–89, Nov. 2024, doi: 10.54254/2753-8818/53/20240147.

[12] M. Walczyński and M. Kisz, 'Using Machine Learning Algorithms to Explore Listeners Musical Tastes', in *Recent Challenges in Intelligent Information and Database Systems*, vol. 1863, N. T. Nguyen, S. Boonsang, H. Fujita, B. Hnatkowska, T.-P. Hong, K. Pasupa, and A. Selamat, Eds., in Communications in Computer and Information Science, vol. 1863. , Cham: Springer Nature Switzerland, 2023, pp. 399–409. doi: 10.1007/978-3-031-42430-4_33.

[13] Z. Xing, 'Popularity Prediction of Music by Machine Learning Models', *HSET*, vol. 47, pp. 37–45, May 2023, doi: 10.54097/hset.v47i.8162.

[14] T. Jain, N. S. Kumar, V. Tiwari, and P. K. Tiwari, 'Music Recommendation System Using Machine Learning', in *2023 2nd International Conference on Futuristic Technologies (INCOFT)*, Belagavi, Karnataka, India: IEEE, Nov. 2023, pp. 1–6. doi: 10.1109/INCOFT60753.2023.10425522.

[15] X. Liu, 'Music Trend Prediction Based on Improved LSTM and Random Forest Algorithm', *Journal of Sensors*, vol. 2022, pp. 1–10, Mar. 2022, doi: 10.1155/2022/6450469.

[16] Konstantinos Gkountakos, Konstantinos Ioannidis, Konstantinos Demestichas, Stefanos Vrochidis, and Ioannis Kompatsiaris, 'A Comprehensive Review of Deep Learning-based Anomaly Detection Methods for Precision Agriculture', *IEEE Access*, pp. 1–1, 2024, doi: 10.1109/access.2024.3522248.

[17] S. Park, H. Sim, M. Kwon, and W. Lee, 'Automatic Music Emotion Classification Using Chords and BPM', in *Future Information Technology*, vol. 185, J. J. Park, L. T. Yang, and C. Lee, Eds., in Communications in Computer and Information Science, vol. 185. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 329–336. doi: 10.1007/978-3-642-22309-9_42.

[18] A. C. North, A. E. Krause, L. P. Sheridan, and D. Ritchie, 'Popularity, mood, energy, and typicality in music: A computerized analysis of 204,506 pieces.', *Psychology of Aesthetics, Creativity, and the Arts*, vol. 13, no. 1, pp. 89–109, Feb. 2019, doi: 10.1037/aca0000165.

[19] J. Li and F. Yan, 'Design and Application of Music Genre Classification Algorithm Based on Machine Learning', in *2024 3rd International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI)*, Zakopane, Poland: IEEE, Oct. 2024, pp. 1064–1068. doi: 10.1109/ICDACAI65086.2024.00198.

[20] R. Cavicchioli, J. C. Hu, and M. Furini, 'On Using Artificial Intelligence to Predict Music Playlist Success', in *2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*, Las Vegas, NV, USA: IEEE, Jan. 2024, pp. 278–283. doi: 10.1109/CCNC51664.2024.10454829.

[21] A. Natev, 'Hit Song Prediction Through Machine Learning and Spotify Data', in *Proceedings of the 10th Student Computing Research Symposium (SCORES'24)*, University of Maribor Press, Oct. 2024, pp. 57–60. doi: 10.18690/um.feri.6.2024.13.

[22] I. Olvera, K. Sood, C. Reyes, and W. Tu, 'Content-Based Music Recommendation System using Supervised Learning', in *2023 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, Malang, Indonesia: IEEE, Nov. 2023, pp. 120–127. doi: 10.1109/COMNETSAT59769.2023.10420756.

[23] S. Arora, R. Rani, and N. Saxena, 'Music Stream Analysis for the Prediction of Song Popularity using Machine Learning and Deep Learning Approach', in *2022 3rd International Conference on Computing, Analytics and Networks (ICAN)*, Rajpura, Punjab, India: IEEE, Nov. 2022, pp. 1–5. doi: 10.1109/ICAN56228.2022.10006843.