Music Mood Prediction Based on Spotify's Audio Features Using Logistic Regression

Marvin Ray Dalida*, Lyah Bianca Aquino*, William Cris Hod*, Rachelle Ann Agapor*,
Shekinah Lor Huyo-a†, Gabriel Avelino Sampedro‡
*College of Computer Science, Technological University of the Philippines, Manila, Philippines
†Research and Development Center, Philippine Coding Camp, Manila, Philippines
‡College of Computing and Information Technologies, National University, Manila, Philippines
Email: {marvinray.dalida, lyahbianca.aquino, williamcris.hod, rachelleann.agapor} @tup.edu.ph;
sbhuyoa@up.edu.ph; garsampedro@ieee.org

2022 IEEE 14th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) | 978-1-6654-6493-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/HNICEM57413.2022.10109396

***Abstract*—Music influences our mood. Individuals have experienced music personally where their emotions become involved, allowing the tempo or lyrics of the music to impact them. Music not only provides entertainment but also helps boost overall well-being. Over the years, music streaming platforms have become popular for the way music is delivered and music queues have been tailored for the listener. In these applications, machine learning has been for music recommendation. In this paper, an innovative approach for modeling a track's mood based on audio components from the Spotify application program interface (API) available on the Spotify PH market will be explored. This paper will focus on the performance evaluation of the the application of logistic regression in predicting the mood of a song. The validation of the model uses the stratified k-fold cross-validation and evaluation confusion matrix for analyzing the model. The primary expected outcome of the system is to predict the song's mood based on the 12 features of Spotify audio components. The model will be the basis of future studies to identify the best factor that affects the track's mood based on its audio features.**

***Index Terms*—application program interface (API), k-fold cross-validation, logistic regression, evaluation confusion matrix.**

I. Introduction

People's lives are so ingrained that their music tastes reflect their moods and who they are. It is reasonable to assert that one of the most prevalent hobbies of people is listening to music. Music helps people freely express their sentiments and emotions, so music is vital in our lives. Today, surfing the world wide web to listen to music has become seemless and subscription-based music streaming platforms brought down the cost to listen to music. Digital and streaming music have steadily superseded the traditional business model of physical recordings in the recorded music business during the previous 17 years. Online music services like SoundCloud, YouTube, and Spotify Music attract more attention from the general public and increase the revenue generated by the digital and streaming music industries.

Among the various music streaming platforms, Spotify emerges as one of the most popular platforms where people worldwide stream songs legally. Since its inception in 2008, Spotify continues to be a pioneer in this field. Spotify has introduced a new genre and mood filter that allows users to quickly sort their "Liked Songs" collection for every occasion and mood [1]. They recently received a patent for a system that examines human voices and background noises to produce song recommendations based on, among other things, emotional state, gender, age, or accent. However, several users and artists are concerned that Spotify's new technology could jeopardize user privacy, and artists have labeled the technology invasive and argued that it discriminates [2].

While the idea seemed revolutionary, the music business clearly needs the use of data science disciplines like machine learning, data mining, and recommender systems to effectively and safely achieve their objectives. Machine learning is a branch of artificial intelligence and computer science that employs data and algorithms to simulate how humans learn, progressively increasing its accuracy. The objective is to comprehend data structures and modify them for models so that people can readily comprehend and use them [3]. In other words, machine learning is used to solve complex problems using statistical modeling and data processing. Given the current situation, Spotify's application program interface (API) will support and enable all of the platform's features, allowing complete access to all of the music data offered for investigating and enhancing future applications in that space. One enhancement of the API, would be to add mood filters. Mood can be classified into moods through a deep learning approach through a machine learning approach and may be classified as either happy, calm, energetic, or sad [4].

The paper aims to analyze the machine learning component in predicting the track's mood using the audio features extracted by the Spotify API. Generally, the paper will outline the approach to answering the problem using a supervised learning model to classify music tracks' moods. The music tracks contain songs and instrumental music. The musical genres included are pop, rock, R&B, folk, jazz, metal, electronic, classical, and instrumental. Specifically, this paper investigates this subject and proposes a method based on supervised learning. Furthermore, the proposed model shall be evaluated based on the performance of the logistic regression model generated using the Sklearn of Python, identify the features

that significantly affect the prediction, and increase listeners' satisfaction in finding a relationship between their existing emotions and the track's mood.

II. Review of Related Literature

This section contains relevant literature and studies from other countries about the importance of music mood prediction in the community as well as studies on music mood prediction using alternative models.

A. Music Mood Prediction Using Other Models

Spotify provides developers access to information on numerous audio features for each song on its platform via an API. According to Spotify's description, these attributes are approximated and calculated for each tune. Variables such as loudness and tempo may be represented in numeric values - decibels (dB) and beats-per-second (bps). On the other hand, other aspects, such as "Acousticness," "Energy," and "Valence," are nominal and descriptive in nature and their rankings based on Spotify's algorithmic computations. "Acousticness," "Danceability," "Duration," "Energy," "Instrumentalness," "Key," "Liveness," "Loudness," "Mode," "Speechiness," "Tempo," and "Valence" are all provided audio features [5]. Panda et al. analyzed the usefulness of the Spotify API audio features to the music emotion recognition sector. According to their reserach, three of the 12 Spotify API properties - energy, valence, and acousticness were significantly relevant to emotion classification [6].

Another study by Yang asserts that effective music extraction is a crucial area for further study in emotion identification [7]. The researcher applies the backpropagation network of a neural network in music extraction. The short-term energy, short-term average amplitude, short-term autocorrelation function, short-term zero-crossing rate, frequency spectrum, amplitude spectrum, and phase spectrum are the best grounds for classifying musical qualities. According to test results from the propsed extraction method, the use of the artificial bee colony algorithm-optimized back propagation network builds the music sentiment classifier to have a more significant recognition impact than other classifiers.

Another study by Kumar and Daiya explores the possibility of using easily accessible audio metadata, such as artist and year, to enhance the performance of mood classification models in order to increase the capacity to recognize mood [8]. The researchers' primary goal is to predict the mood of the music, but they also use multi-task learning to predict the music's metadata (such as the artist and year). Their study discovers that multimodal mood prediction does not need feature engineering since deep learning-based models outperform conventional methods for mood detection. The performance is constantly improved by using their multi-task learning approach to the current Convolutional Neural Network model for mood classification.

Another study by Ridoean et al. discovers that the best MPEG-7 audio properties are audio power and harmony in music mood classification features. The audio strength and harmonious tone of the music influences the label, although the results deteriorate when the Audio Spectrum Projection attributes are combined with Audio Power and Audio Harmonicity for categorization [9]. The researchers' findings indicate that the accuracy of the classifier employing the Audio Power and Audio Harmonicity characteristics is best for labeling moods such as furious, joyful, and sad.

Due to its outstanding performance, machine learning is becoming increasingly popular across all industries, and music is constantly a part of our life [10], [11]. The SVR model is designed for the MER using a practical supervised framework with an auto-encoder-based optimized SVR model. A support vector regression model based on KMBSO is employed for emotion classification. The ideal SVR parameters are chosen using the KMBSO approach. According to their findings, this system performs mood recognition substantially better than other current methods. The performance for lyrics and karaoke is discovered to be less than the song, although the suggested SVR KMBSO has produced superior results for songs.

B. Significance of Music's Mood in the Community

In scientific studies, music is seen to affect the mood of people. The sounds of music in harmony and the loudness affects daily behaviors, thoughts, and actions. A research by Koelsch asserts that music may influence each of the five instinctual subsystems or elements of emotion: (1) When we evaluate music, we experience emotions like pleasure or repulsion, which have an impact on our cognitive emotions. (2) Strong action effects brought on by music, such as dancing or moving to the beat, can also alter the motivating element. (3) Peripheral psychological activity, such as alertness or relaxation, may be altered by music, which impacts the physiological component. (4) When producing or listening to music, a person's facial expression of emotion is influenced. The expression component is impacted. Finally, music (5) elicits emotions like happiness, being moved, courage, grief, tension, relaxation, etc [12]. The research claims that music might alter every part of the brain connected to emotion. A person's social connection and sense of belonging in society may be expressed via music in addition to their emotions. As a result, feelings sparked by or associated with music are natural, and these feelings can be intense. This makes music a unique instrument for examining emotions.

According to Shukla, listening to sad music usually generates three reactions: authentic, comforting, and uplifting sorrow (positive valence); sadness (negative valence); and sweet sorrow (positive valence). Self-selected music (rather than prescribed) can also aid in regulating unpleasant emotions driven by other demanding tasks [13]. The author's findings indicate that heavy metal music can also be a beneficial way to process anger. People frequently use it to control their emotions.

On the other hand, according to a study conducted by Heshmat, music can give listeners significant emotional reactions like thrills and shivers. In interactions with music, happy feelings predominate [14]. Dopamine and other neurotransmitters

connected to rewards may be released in response to enjoyable music. One may easily alter their mood or release stress by listening to music. People use music to regulate, amplify, and lessen unfavorable emotional states in their daily lives (e.g., stress, fatigue).

III. Methodology

This section includes the acquisition, modeling, and evaluation of the data. Since the research revolves around music, the data to be processed will be audio files acquired using the Spotify API. The first objective is to create a dataset and this can only be done through data acquisition processes. Once data has been obtained, data modelling is needed to ensure that the data can be efficiently and effectively processed by the algorithms [15]. Finally, performance evaluation must be done on the models to check if the system produces the desired outputs.

A. Data Acquisition

Before analyzing anything, data must be first obtained. The data collected must be classified in order to create a classification model. The dataset shall comprise of audio files with labels that specify the mood of the music. The collected data consists of 1,500 music labeled into four moods: energetic, calm, sad, and happy. Each mood playlist has 375 tracks that will be used to extract audio features to create the data set.

The data is cleansed by removing the duplicated tracks in each mood category to prevent mislabelling when combining multiple data sources. Processing was done by applying a logical code that returns the replicated data set, which is eliminated one by one until there are no more replications.

1) Dataset: Once the data has been collected, will be segregated into playlists. Each playlist's uniform resource identifier (URI) is copied and passed as an argument to Spotify's "playlist tracks" function to get the list of every track's URI. Each URI to Spotify's "audio features" function is supplied as an argument to extract the audio features of every track.

For each tune, the Spotify API provides its audio characteristics. The 1,500 Spotify music songs that make up the comma-separated values (CSV) file comprise the audio features from the four playlists. Spotify's API gives the ability to extract several audio features from a track. Table 1 shows the emulated descriptions in the developers' documentation analysis of Spotify's API that was used for this analysis. Table III-A1 shows the different feautures and their respective descriptions.

B. Data Modelling

Once the data has been collected, the next objective would be to create a data model. Creating a data model for the machine learning algorithm would include splitting the dataset, training the model, getting the accuracy of the training dataset, validating the model, and evaluating the model. This subsection shall tackle the whole data modelling process.

TABLE I
Feature Ranking

Audio Feature	Description
Danceability	Danceability describes how suitable a track is for dancing.
Energy	Energizing music has a quick, loud, and raucous feeling. For example, death metal has high energy.
Key	the key that the song is in. Using the conventional pitch class notation, integers correspond to pitches.
Loudness	The intensity level of the sound of the track decibels (dB).
Mode	Mode indicates the modality (major or minor) of a track.
Speechiness	Speechiness identifies if lyrics are being sung on a track.
Acousticness	Acousticness refers to the confidence measure of a track, wherein a track with a CM of 0.0 to 1.0 is acoustic.
Instrumentalness	Predicts whether a track contains no vocals.
Liveness	Detects if a live audience is present on a track.
Valence	Valence refers to the positive tone of the track.
Tempo	The tempo refers to the number of beats per minute (BPM) of a track.
Time Signature	An estimated overall time signature of a track.

1) Splitting the Dataset: Sklearn's model selection approach provides the framework for data analysis and measurement. When producing a forecast, this technique enables the development of precise findings. The train split function is used to divide the dataset into training and testing in order to do this. A training set containing 80% of the data and a testing set with the remaining 20% comprise the data set. An X-train feature matrix and a Y-train vector label are features of the training dataset. The training set would be used to train the system, similar to how children are given pictures with labels when learning new words. The testing set would be used to test the accuracy of the system, similar to how children are given tests after being taught words.

2) Training the Model: Before fitting and training the model, Logistic Regression must be imported first from Sklearn's linear models to create a model and fit/train it using the X-train and Y-train. To get the training accuracy of the model, the researchers made the model predict the y-value after being provided with an x- and y-training set. The system would then use this large number of data with labels and the model would try to "learn" or understand how each y-value was derived.

3) Validation of the Model: In the validation of the model, a 10-fold cross-validation with stratified sampling is implemented. One fold is used as the test set in each of the 'k' iterations, while the remaining folds are used for training. The created model is validated using Sklearn's stratified K-fold cross-validation. The accuracy score of each iteration is shown in Figure 3. Then the mean accuracy score is acquired.

Sklearn's feature selection function recursive feature elimination (RFE) could rank all the audio features by their importance. First, the model is trained on the initial set of features. The importance of each component is obtained through any

specific attribute or callable. Then, the minor key features are pruned from the current features. The procedure is repeated on the pruned set until the desired number of selected features is eventually reached.

TABLE II
FEATURE RANKING

Ranking	Audio Features
1	Energy
2	Acousticness
3	Valence
4	Instrumentalness
5	Speechiness
6	Danceability
7	Liveness
8	Mode
9	Loudness
10	Time Signature
11	Key
12	Tempo

4) Evaluating the Model: To assess the model's accuracy, a heatmap confusion matrix is displayed using the Seaborn Library and Matplotlib. The data is shown in two dimensions on the heatmap. The Seaborn library offers a high-level data visualization interface that allows us to create our matrix while the data values are represented in the graph by colors.

IV. PERFORMANCE EVALUATION

The dataset consists of 1500 tracks classified into four moods, with 375 tracks each. After running the machine learning model, the results of the test shall display the different predictions made by the system. The results would include correct predictions of correct items or true positive (TP); correct predictions of incorrect items or true negative (NP); incorrect predictions of correct items or false positive (FP); and incorrect predictions of incorrect items or false negative (FN). The results of each are seen in the confusion matrix in Figure 1.

To further evaluate the results, metrics such as accuracy, precision, recall, and f1-scores are obtained. The accuracy of the model illustrates how well the model performs in correctly identifying labels. The precision of the model demonstrates how well the model performs in identifying correct values. The recall demonstrates the positivity rate of the model. Lastly, the f1-score takes into account both precision and recall. The equations to indicate the metrics mentioned are as follows:

$$accuracy = \frac{TN + TP}{TP + FP + TN + FN} \tag{1}$$

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

Fig. 1. The confusion matrix shows the number of correct and incorrect classification, per mood.

TABLE III
PERFORMANCE EVALUATION

Mood	accuracy	precision	recall	f1 score
Calm	0.9866	0.9494	0.9036	0.9259
Energetic	0.9733	0.8806	0.8194	0.8489
Happy	0.9400	0.7662	0.8310	0.7973
Sad	0.9600	0.8442	0.8784	0.8610
AVERAGE	0.9600	0.8442	0.8784	0.8610

$$F1 = 2 * \frac{precision * recall}{precission + recall} \tag{4}$$

In statistics, the reduction in variance is connected with a model's accuracy. It gauges how near the values are to one another. Recall is a statistic that assesses real positives. It demonstrates the accuracy of the readings. The accuracy also considers the actual positives and negatives. It gauges how effectively your system recognizes the proper values, which might be positive or negative. In addition to the accuracy, precision, recall, and f1 score, the receiver operating characteristic (ROC) curves is used to further analyze the created model. ROC curves are widely used to graphically depict the relationship/trade-off between clinical sensitivity and specificity for every cut-off for a test or a combination of tests. The closer the curve gets to the upper left, the better. Figure 2 presents the ROC curve of the results.

The whole dataset is composed of the audio attributes of 1,500 tracks. Of the 1,500 tracks, 1,200 are used in the training process and 300 in the testing phase. The system's algorithm used is logistic regression, and it performs with an average accuracy of 0.9645, a precision of 0.8601, a recall of 0.8581, and an f1 score of 0.8583. In addition, the results have been observed not to overfit or under, thus validating the application of the model in music classification.

V. CONCLUSION

This paper offered an analysis of the suitability of the Spotify API audio features to the music emotion recognition

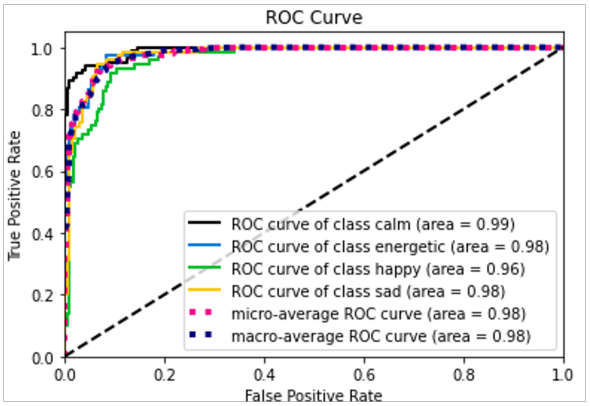

Fig. 2. ROC Curve

field. The researchers conclude that among the 12 audio features extracted for each track, the five significant factors that affect mood prediction are energy, acousticness, valence, instrumentalness, and speechiness. Conversely, the three minor audio features that affect the result of the mood are time signature, key, and tempo. This allows the conclusion that the logistic regression model only predicts the mood of the music track based on the audio features. The model disregards the song's lyrics and analyzes the numerical value of the audio components of the track. So, if one wants to listen to happy music, a higher positive valence of the track would get him to feel that way.

While everyone reacts differently to music, one thing is sure: it plays an essential role in one's life. What works for one person may not suit others – but music in all its forms will stand to benefit and entertain people. In any genre, myriad benefits of music can motivate a person's focus on a particular event or predominantly set the mood to ease stress. Knowing the track's emotion significantly affects listeners, depending on certain occurrences, may it be putting someone in a better perspective or generally wanting to change what they feel. Predicting the right mood of the song that suits a person's emotion could psychologically improve one's feelings and change how one perceives the world.

Future research could examine predicting the mood based on the lyrics might prove an essential area for future research to give more accurate results and be enjoyed more by the listeners. Primarily, the model will be the basis of future studies in identifying the best factor that affects the track's mood based on its audio features.

ACKNOWLEDGMENT

This research work was funded by National University, Philippines, through the National University Researcher Program (2021F-1T-02-MLA-CCIT).

REFERENCES

[1] I. Siles, A. Segura-Castillo, R. Solís, and M. Sancho, "Folk theories of algorithmic recommendations on spotify: Enacting data assemblages in the global south," *Big Data & Society*, vol. 7, no. 1, p. 2053951720923377, 2020.

[2] M. Bartlett, F. Morreale, and G. Prabhakar, "Analysing privacy policies and terms of use to understand algorithmic recommendations: the case studies of tinder and spotify," *Journal of the Royal Society of New Zealand*, pp. 1–14, 2022.

[3] Z.-H. Zhou, *Machine learning*. Springer Nature, 2021.

[4] R. Ferdiana, W. F. Dicka, and F. Yudanto, "Mood detection based on last song listened on spotify," *ASEAN Engineering Journal*, vol. 12, no. 3, pp. 123–127, 2022.

[5] L. Spear, A. Milton, G. Allen, A. Raj, M. Green, M. D. Ekstrand, and M. S. Pera, "Baby shark to barracuda: Analyzing children's music listening behavior," in *Fifteenth ACM Conference on Recommender Systems*, 2021, pp. 639–644.

[6] R. Panda, H. Redinho, C. Gonçalves, R. Malheiro, and R. P. Paiva, "How does the spotify api compare to the music emotion recognition state-of-the-art?" in *Proceedings of the 18th Sound and Music Computing Conference (SMC 2021)*. Axea sas/SMC Network, 2021, pp. 238–245.

[7] J. Yang, "A novel music emotion recognition model using neural network technology," *Frontiers in Psychology*, p. 4341, 2021.

[8] R. Kumar and M. Dahiya, "Multi-task learning with metadata for music mood classification," *arXiv preprint arXiv:2110.04765*, 2021.

[9] J. A. Ridoean, R. Sarno, D. Sunaryo, and D. R. Wijaya, "Music mood classification using audio power and audio harmonicity based on mpeg-7 audio features and support vector machine," in *2017 3rd International conference on science in information technology (ICSITech)*. IEEE, 2017, pp. 72–76.

[10] B. L. Sturm, O. Ben-Tal, Ú. Monaghan, N. Collins, D. Herremans, E. Chew, G. Hadjeres, E. Deruty, and F. Pachet, "Machine learning research that matters for music creation: A case study," *Journal of New Music Research*, vol. 48, no. 1, pp. 36–55, 2019.

[11] G. A. R. Sampedro, D. J. S. Agron, G. C. Amaizu, D.-S. Kim, and J.-M. Lee, "Design of an in-process quality monitoring strategy for fdm-type 3d printer using deep learning," *Applied Sciences*, vol. 12, no. 17, p. 8753, 2022.

[12] S. Koelsch, "Investigating the neural encoding of emotion with music," *Neuron*, vol. 98, no. 6, pp. 1075–1079, 2018.

[13] A. Shukla, "The Importance Of Music: When and Why we listen to music - Cognition Today — cognitiontoday.com," https://cognitiontoday.com/the-importance-of-music-when-and-why-we-listen-to-music/, [Accessed 07-Oct-2022].

[14] S. Heshmat, "Music, emotion, and well-being," *Psychology Today*, vol. 25, 2019.

[15] G. A. Sampedro, R. G. C. Kim, Y. J. Aruan, D.-S. Kim, and J.-M. Lee, "Noise filtering mobile application for speech enhancement using a redundant convolutional encoder-decoder," in *2021 1st International Conference in Information and Computing Research (iCORE)*. IEEE, 2021, pp. 34–38.