ASEAN Engineering Journal

Full Paper

MOOD DETECTION BASED ON LAST SONG LISTENED ON SPOTIFY

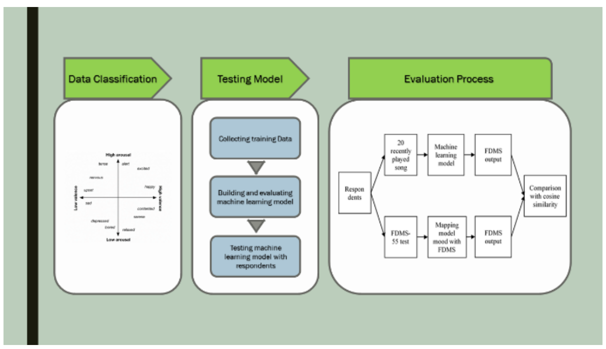
Ridi Ferdiana*, Wiliam Fajar Dicka, Faturahman Yudanto

Department of Electrical and Information Engineering, Universitas Gadjah Mada, 55281, Yogyakarta, Indonesia

Article history
Received
05 April 2022
Received in revised form
03 February 2022
Accepted
05 February 2022
Published online
31 August 2022

*Corresponding author
ridi@ugm.ac.id

Graphical abstract



Abstract

A Song Is One Medium Used To Express Someone's Emotion, Whether As A Performer Or Audience. With The Advancement Of Machine Learning And A Deeper Understanding Of Sentiment Analysis, We Decided To Study Mood Detection Based On The Last Song Listened To**.** One of the direct ways to measure someone's mood is by using a Four-dimensional Mood Scale (FDMS) device. This device categorized mood into four dimensions: low valence, high valence, low arousal, and high arousal. In this article, we used a variation of FDMS adapted to the Indonesian language called FDMS-55 to compare the result from our model. Our model is trained using song data collected from Spotify and Genius using their respective API (Application Programming Interface). We classified manually into a mood class and then processed further using Azure Cognitive Service Text Analytics API. Based on evaluation conducted on the model, the FastTreeOva algorithm produces the highest accuracy both on valence class with 0.8901 and arousal class with 0.9167. The comparison between the model result and respondent's FDMS-55 device result is made with cosine similarity and yields similarity value of 0.770 with 0.103 standard deviation. It is concluded that someone's mood is related to the song they listened to, and our model can precisely predict someone's mood based on the last song they listened to.

Keywords: Four-dimensional mood scale, Multi-class classification, Machine learning, Natural language processing, Sentiment Analysis

1.0 INTRODUCTION

Music has been known to be strongly linked with human emotions. The composer or performer often uses musical features such as rhythm, instrumental, and lyrics to express their emotion. The difference between these features is seen clearly; for example, we can easily distinguish a piece of music or song with a sad tone from one with a happy tone. This is also true for the audience. Someone is more likely to listen to music that is aligned with their current mood or emotion. Furthermore, music can generate specific emotion to the listener according to its tone [1]. In other words, it is entirely possible to detect someone's current mood based on the last song that they listened to. Of course, it is not easy to identify mood or emotion, especially using a computer to do so. Although emotion classification and detection has existed for a long time and has been reliably measured using devices such as the Four-Dimensional Mood Scale (FDMS) [2].

In recent years, the advancement of machine learning technology has opened many paths to innovation and integration between computational technology and various field study such as sentiment analysis that is implemented in this study. Machine learning uses a large amount of data to find a pattern in said data, producing a result that we specified beforehand. This study creates a machine learning model that can detect certain moods or emotions within a song. This model will then be compared to our respondent's actual mood that is taken by the Four-dimensional Mood Scale (FDMS) device. Using cosine similarity, we will then compare our machine learning model's result with the result taken from the FDMS device to find a similarity value between someone's mood and the kind of songs they listened to. If we conclude that we can measure someone's mood based on the song they listened to, we can potentially apply this model in an algorithm on a music platform, such as one that can detect the user's mood based on their current song and choose or recommend the next song accordingly.

2.0 METHODOLOGY

Kim et al. divide mood representation into two grouping methods, namely based on categories and dimensionality [3]. Representation based on categories is also used on MIREX audio mood classification. In that study, the mood is categorized into five groups [4].

Table 1 describes the five groups or clusters of mood categorization and lists each mood inside those groups or clusters. This categorization shows that the mood shifts from the peaceful nature on cluster 1 to more of antagonistic nature on cluster 5.

Table 1 Different categorization of mood

Cluster	Mood
1	Rowdy, Rousing, Confident, Boisterous, Passionate
2	Amiable / Good-natured, Sweet, Fun, Rollicking, Cheerful
3	Witty, Humorous, Whimsical, Wry, Campy, Quirky, Silly
4	Literate, Wistful, Bittersweet, Autumnal, Brooding, Poignant
5	Volatile, Fiery, Visceral, Aggressive, Tense / Anxious, Intense

Thayer applies the other representation in the quadrant that consists of two scaling values, namely "valence" and "arousal." Valence is related to one's happiness, and arousal is related to one's enthusiasm [5]. Another similar model, such as the Tellegen-Watson-Clark model, replaces Thayer's arousal and valence value with positive and negative affect, forming a similar 4-quadrant figure with high and low positive/negative affect [6]. Lucia Martin-Gómez used this model as emotion categorization in their study on sentiment analysis in music [7].

Another study states that emotion can be categorized based on core emotion that he called core affect. This theory explains that emotion consists of two polar that is valence and arousal, like Thayer's and Tellegen-Watson-Clark's model [8]. In the following study, this core affect theory was also related to the previous five-category model [9].

Figure 1 shows that every quadrant has its own set of moods. For example, if someone has a high value of arousal and valence, it can be said that they are feeling excited. The validation of mood classification based on this arousal-valence dimensionality has been confirmed by multiple studies [10], [11]. Bhat et al. also used Thayer's mood model as a foundation to classify mood on Western and Hindi style music based on intensity, rhythm, and pitch. They then simplified Thayer's mood model class classification into 8 of the following class: happy, exuberant, energetic, frantic, sad, depressed, calm, and contentment [12].

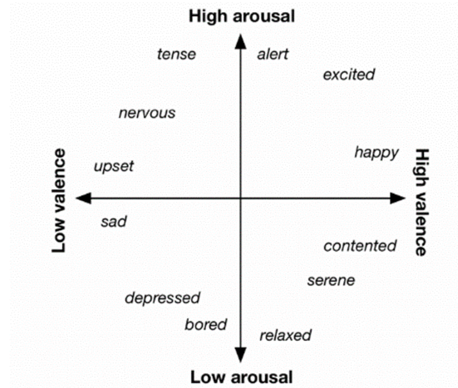


Figure 1 Mood representation based on dimensionality [5].

Similar classification can also be seen on the Tellegen-Watson-Clark model in Figure 2. Again, its X-axis represents negative effect instead of valence, and its Y-axis represents positive effect instead of arousal.

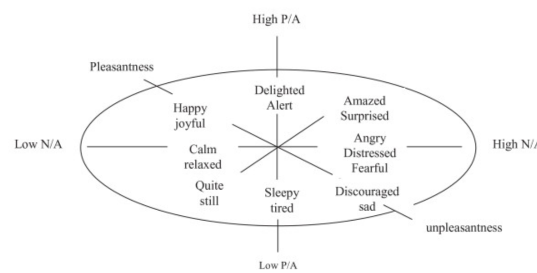


Figure 2 Tellegen-Watson-Clark mood representation based on positive and negative affect [6].

The existence of mood in humans is known to be measurable. One of the devices widely used for this measurement in Indonesia is the Four-Dimensional Mood Scale (FDMS), which measures mood based on positive energy, tiredness, negative activation, and relaxation. FDMS is made based on the core affect theory with the precept of 2 bipolar, valence, and arousal [2].

FDMS measuring device comes in the form of the questionnaire, consisting of various adjectives which point to dimensions specified in Figure 1 and Figure 2. Using this device, respondents must fill a number between 1 to 5, representing how relevant each adjective is to the respondent's emotion. This method has been used and modified to conduct multiple studies on human behaviors such as impulse buying and unhealthy eating [13] [14].

3.0 EXPERIMENT

There are three main steps in this experiment, which can be seen in Figure 3. The first step is collecting training data that will be used for our machine learning model. Next, using those data, we will build and evaluate the model. And the last step is testing and comparing the result of our model with actual respondents.

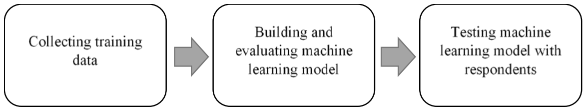


Figure 3 Experiment model steps

Collecting Training Data

We used Spotify API, Genius Lyrics API, and Azure Cognitive Service Text Analytics version 3 for data acquisition. We started by acquisitioning song feature data from Spotify playlists using Spotify Web API. Spotify Web API collected various songs from our playlist and extracted the features from those songs. These features include danceability, energy, key, loudness, the quantity of speech, acoustic and instrumental, liveness, valence, and tempo. We get the data as specified in Table 2.

Table 2 Song features from the dataset

Feature	Data Source	Used to Train the Model
Id	-	No
Artist	Spotify API	No
Song	Spotify API	No
Danceability	Spotify API	Yes
Energy	Spotify API	Yes
Key	Spotify API	No
Loudness	Spotify API	Yes
Qty of Speech	Spotify API	No
Qty of Acoustic	Spotify API	Yes
Qty of Instrumental	Spotify API	Yes
Liveness	Spotify API	No
Valence	Spotify API	Yes
Tempo	Spotify API	Yes
Sent_Positive	Text Analytics v3	Yes
Sent_Negative	Text Analytics v3	Yes
Sent_Neutral	Text Analytics v3	Yes
Class_Label 1	-	No
Class_Label 2	-	Yes

Genius Lyrics API is used to get the lyrics of the song that we collected. There are 2 steps for collecting these lyrics. First, we inputted the name of the song and the artists to find the lyrics page. This step is done using the song data we gathered from Spotify Web API. Then from this page, we downloaded the lyrics by web scrapping.

In the next step, we process the lyrics of our song data using Azure Text Analytic to get the sentiment value of our song. Azure Text Analytics will generate a percentage of positive sentiment, neutral sentiment, and negative sentiment on each song. These data are then saved as dataset files, then used to build the machine learning model. We will then build our machine learning model using the ML.Net Model Builder framework (see Figure 4). These are the data pre-processing step of our experiment:

- Get song features from a selected playlist using Spotify Web API
- Get song lyrics from Genius Lyrics API using song and artist's name
- Send the lyrics into Azure Cognitive Service Text Analytics v3
- Get the response API and use each song's sentiment value (positive, negative, and neutral)

Train And Evaluate The Machine Learning Model

In this research, we train three types of machine learning models. Each machine learning model will be used to classify songs based on the valence and arousal value of the song. Each model is trained with a dataset consisting of 1,037 song data from 17 playlists (see Table 3). Machine learning model training is done using ML.Net Model Builder in Visual Studio 2019 environment on a laptop with Intel Core i5-8265U 1.60GHz processor and 8GB of RAM.

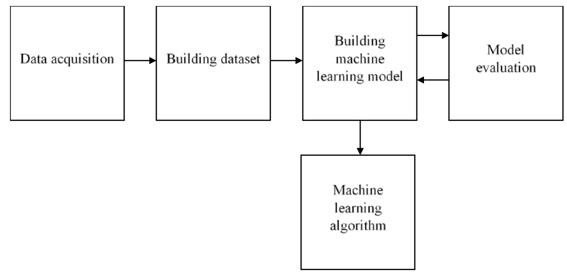


Figure 4 Building and evaluating machine learning model

Table 3 Dataset playlist

Playlist	Arousal Label	Valence Label
Broken Heart	Low Arousal	Low Valence
Good Times	High Arousal	High Valence
No Stress!	Low Arousal	High Valence
90's Love Song	Low Arousal	High Valence
Alone Again	Low Arousal	Low Valence
Happy Hits	High Arousal	High Valence
Slow-Mo Emo	Low Arousal	Low Valence
Life Sucks	Low Arousal	Low Valence
Heart Beats	High Arousal	High Valence
Young Wild Free	High Arousal	High Valence
Rage Beats	High Arousal	Low Valence
Deep Sleep	Low Arousal	High Valence
Powerwalk 120-125	High Arousal	High Valence
Workout	High Arousal	High Valence
Coping with Loss	Low Arousal	Low Valence
Sad Beats	High Arousal	Low Valence
Sad Indie	Low Arousal	Low Valence

ML.Net is an open-source machine learning network to build, train, and deploy machine learning models, which includes sentiment analysis. In this process, we divide our data into an 80:20 split of training data and testing data. ML.Net builder will train our model using three multi-class classification algorithms: FastTreeOva, FastForestOva, and LightGbmMulti. FastTreeOva and FastForestOva use a one-versus-all trainer to train a multi-class classifier using fast tree and fast forest decision tree binary classification model, respectively. The last algorithm, LightGbmMulti uses a boosted decision tree multi-class classification model to train our machine learning model. After training is complete, ML.Net Model Builder will show the evaluation result from each of our models. Then our model will be exported and used to classify actual data.

Testing Machine Learning Model With Respondents

The last phase of our experiment is testing our machine learning model with real respondents. In this phase, respondents are asked to login into their Spotify account via our web application which is specifically built for this testing phase. Spotify API will then collect the respondent's recently played songs. Then we use our trained machine learning model to analyze the respondent's recently played songs to measure its sentiment in the form of valence and arousal output value.

Next, we measure our respondent's recent emotions using FDMS. Our respondents consisted of 51 Indonesian citizens. Therefore, we used the FDMS-55 device to perform the measurement. FDMS-55 is an Indonesian language-adapted variation of FDMS created by Adinugroho [15]. In Table 4, we can see some of the examples of FDMS-55 measured emotions.

Table 4 FDMS Dimension and examples

Dimension	Variable example	Valence & Arousal value
Positive Energy	Active Energetic Enthusiastic Spirited	High valence and high arousal
Tiredness	Bored Tired Lazy Weary	Low valence and low arousal
Negative Activation	Fear Angry Nervous Angst	Low valence and high arousal
Relaxation	Relaxed Calm Quiet Peaceful	High valence and low arousal

There are four classes in FDMS which is positive energy, tiredness, negative activation, and relaxation. FDMS-55 will measure the sum of value on each class as its output. These outputs are represented as a 4-dimension vector. We also have adjusted our machine learning model to generate results in the same 4-dimensional vector. Then, we compare these two results to each other using cosine similarity. This comparison will produce output ranging from 0 to 1 in which 0 refers to minimum similarity and one refers to high similarity. This experiment can conclude whether we can accurately detect someone's emotions using their recently played song.

The output is represented in a 4-dimensional vector. The similarity between the output from FDMS and our machine learning model will be measured with cosine similarity. Cosine similarity is a method to calculate the similarity between two vectors based on the cosine angle between the two vectors without comparing the actual value of both vectors. Cosine similarity can be calculated using

$$\cos(\theta) = \frac{\boldsymbol{A} \cdot \boldsymbol{B}}{\|\boldsymbol{A}\|\|\boldsymbol{B}\|}$$

Cosine similarity value ranges from 0 to 1. The bigger the value, the more similar the value the outputs are. The entire process of the evaluation method with our respondents can be seen in Figure 5.

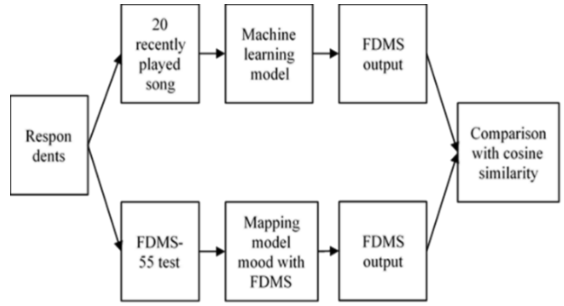


Figure 5 Testing the machine learning model with respondents

4.0 RESULTS

Machine Learning Accuracy

We built the machine learning model with three algorithms, which are FastTreeOva, FastForestOva, and LightGbmMulti. Each algorithm is built using the same dataset of 1,038 songs from 17 playlists. We use 80% of our dataset as training data and 20% as testing data. Using the training data on our models, we can calculate the accuracy when predicting valence class and arousal class with the following formula.

$$Accuracy = \frac{Correct\ Prediction}{All\ Prediction}$$

The model with FastTreeOva algorithm reached the highest accuracy of both valence class with 0.8910 and arousal class with 0.9167. The comparasion of FastTreeOva with the other algorithm can be seen on figure 6. This result shows that this machine learning model can correctly measure valence and arousal value based on song's lyrics. Therefore, it is suitable to compare the FDMS-55 result from our respondents to conclude whether there is a similarity between someone's emotion with the last song they listened to.

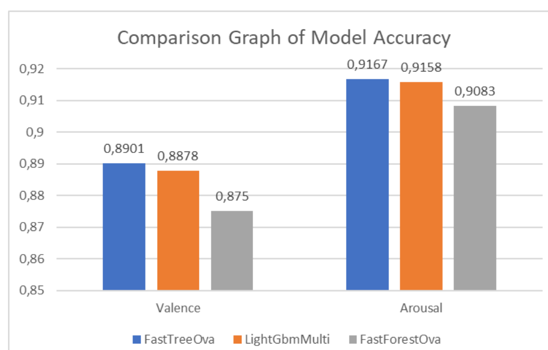


Figure 6 Comparison graph of model accuracy using 3 algorithms

Comparing Machine Learning Result With Fdms-55

We built the machine learning model with three algorithms, which are FastTreeOva, FastForestOva, and LightGbmMulti. Each algorithm is built using the same dataset of 1,038 songs from 17 playlists. We use 80% of our dataset as training data and 20% as testing data. Using the training data on our models, we can calculate the accuracy when predicting valence class and arousal class with the following formula.

5.0 CONCLUSIONS

From our experiment and the test conducted with the test with 51 respondents from this experiment, we can conclude the following:

- The highest accuracy from the machine learning model is 0.8901 for the valence class and 0.9167 for the arousal class. Both results were yielded from the FastTreeOva algorithm. From this result, it can be concluded that the algorithm used for mood classification based on song data is suitable for use.
- The result from the machine learning model test with respondents is that the FDMS output from the model and the FMDS-55 device has a similarity of 0.770 with 0.103 standard deviations based on cosine similarity. From this result, the conclusion is that the machine learning model can be used to measure emotion based on songs almost as well as an FDMS-55 device.

Acknowledgement

The authors would like to thank Universitas Gadjah Mada's academic grants for their contribution in supporting the research and experiment in this paper

References

[1] J. K. Vuoskoski and T. Eerola, 2011 "Measuring Music-Induced Emotion: A Comparison of Emotion Models, Personality Biases, and Intensity of Experiences," *Musicae Scientiae*, 15(2): 159–173., doi: 10.1177/102986491101500203.

[2] T. J. Huelsman, R. C. Nemanick, and D. C. Munz, "Scales to measure four dimensions of dispositional mood: Positive energy, tiredness, negative activation, and relaxation," *Educational and Psychological Measurment*, 58(5): 804–819, 1998.

[3] Y. E. Kim *et al.*, 2010. "Music emotion recognition: A state of the art review," Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010),. *ISMIR 2010*, vol. 11: 255–266,

[4] X. Hu, J. S. Downie, C. Laurier, M. Bay, and A. F. Ehmann, "The 2007 mirex audio mood classification task: Lessons learned," *Proceedings of the 9th International Society for Music Information Retrieval (ISMIR 2008).* 9: 462–467, 2008.

[5] R. E. Thayerm, 1989.*The Biopsychology of Mood and Arousal*. New York, NY, USA: Oxford University Press,

[6] D. Yang and W.-S. Lee, 2004. "Disambiguating Music Emotion Using Software Agents,"

[7] L. M. Gómez and M. N. Cáceres, 2018, "Applying Data Mining for Sentiment Analysis in Music," 198–205. doi: 10.1007/978-3-319-61578-3_20.

[8] J. A. Russel, 1980,"A circumplex model of affect," *Journal of Personality and Social Psychology*, 39(6): 1161-1178. *Journal of Personality and Social Psychology,* [Online]. Available: https://doi.org/10.1037/h0077714

[9] M. S. M. Yik, J. A. Russel, C. K. Ahn, J. M. F. Dols, and N. Suzuki, 2002."Relating the Five-Factor Model of Personality to a Circumplex Model of Affect. In: McCrae R R., Allik J. (eds) The Five-Factor Model of Personaloty Across Cultures," in *International and Cultural Psychology Series*, Boston: Springer,

[10] C. Laurier, M. Sordo, J. Serr̀a, and P. Herrera, "Music mood representations from social tags," Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009), vol. 10, pp. 381–386, 2009.X. Hu, M. Bay, and J. S. Downie, 2007. "Creating a simplified music mood classification ground-truth set," *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR 2007)*, 309–310

[11] A. S. Bhat, V. S. Amith, N. S. Prasad, and D. M. Mohan, "An efficient classification algorithm for music mood detection in western and Hindi music using audio feature extraction," *International Conference on Signal Image Processing Challenge (ICSIP),* 5: 359–364, 2014, [Online]. Available: https://doi.org/10.1109/ICSIP.2014.63

[12] D. H. Silvera, A. M. Lavack, and F. Kropp, 2008,"Impulse buying: the role of affect, social influence, and subjective wellbeing," *Journal of Consumer Marketing*, 25(1): 23–33, doi: 10.1108/07363760810845381.

[13] B. Verplanken, A. G. Herabadi, J. A. Perry, and D. H. Silvera, 2005, "Consumer style and health: The role of impulsive buying in unhealthy eating," *Psychology & Health*, 20(4): 429–441. doi: 10.1080/08870440412331337084.

[14] I. Adinugroho, 2018"Memahami Mood Dalam Konteks Indonesia: Adaptasi Dan Uji Validitas Four Dimensions Mood Scale (Understanding Mood in Indonesian Context: Adaptation and Validity Examination of Four Dimensions Mood Scale)," *SSRN Electron*, 5(2)