Design and Analysis System of KNN and ID3 Algorithm for Music Classification based on Mood Feature Extraction

Made Sudarma[1], I Gede Harsemadi[2]

[1]Department of Electrical and Computer Engineering, Faculty of Engineering, Udayana University, Bali, Indonesia
[2]Department of Information System, STMIK-STIKOM Bali, Indonesia

Article Info	ABSTRACT
  	Each of music which has been created, has its own mood which is emitted, therefore, there has been many researches in Music Information Retrieval (MIR) field that has been done for recognition of mood to music. This research produced software to classify music to the mood by using K-Nearest Neighbor and ID3 algorithm. In this research accuracy performance comparison and measurement of average classification time is carried out which is obtained based on the value produced from music feature extraction process. For music feature extraction process it uses 9 types of spectral analysis, consists of 400 practicing data and 400 testing data. The system produced outcome as classification label of mood type those are contentment, exuberance, depression and anxious. Classification by using algorithm of KNN is good enough that is 86.55% at k value = 3 and average processing time is 0.01021. Whereas by using ID3 it results accuracy of 59.33% and average of processing time is 0.05091 second.

Corresponding Author:

Made Sudarma,
Department of Electrical and Computer Engineering
Faculty of Engineering,
Udayana University,
Jimbaran, Badung, Bali, Indonesia.
Email: imasudarma@gmail.com

1. INTRODUCTION

There is direct relationship between music and emotion [1]. Music has many advantages to human's body and soul, music can calm even inspiring the person who listen to the music. Psychologist stated that there is part in human's brain that feels music is close to the brain part that related to human's emotional expression. Researchers of Music Information Retrieval (MIR) field developed algorithm to detect mood to music automatically. They tried to use emotion in music to get conventional metadata such as style, similarity, genre, and mood. In research performed by Song, et al. [2] stated that each of music created has its own emitted emotion energy. Thayer [3] proposed 2 dimensions model that mapping mood in music, this 2 dimensions approach designate theory that stated that mood is caused by two factors, those are stress (happiness and anxious) and energy (relax and energetic). From this 2 dimensions model it is divided into 4 clusters those are contentment, exuberance, depression, and anxious.

In various research related to MIR, it used various kinds of data mining method for grouping including data classification and clustering such as C4.5 [4], decission tree [5], Support Vector Model [1], Artificial Neural Network [6], Self Organization [7], K-Means [8], [9], etc. Classification process by using this data mining algorithm was initiated with pre-processing stage. In this stage music part to be used is the refrain, which the part with frequent words and notes repetition and this in the part that most determine mood type include in the music [10]. This part with 30 seconds duration [1], [6] with format of *.wav mono channel is furthermore processed by using signal processing of Fast Fourier Transform (FFT) and nine types

of Spectral Analysis Features, with the purpose to change signal from time domain to frequency domain to make easy feature extraction process to get a set of feature value that will be processed as practicing data and testing data in classification process using K-Nearest Neighbor (KNN) and Iterative Dichotomiser 3 (ID3) algorithm. KNN is known as strong algorithm in classification process of testing data to practicing data that has much noise, so it is effective if it uses big practicing data. Whereas, ID3 algorithm uses concept from information entropy, which has superiority in performing searching through in all existing classification possibilities in decision tree.

This research has purpose to be able to build a system that can classify music into four mood parameters and to find out accuracy level and processing time performed by using KNN and ID3 algorithm. The second part of this research paper will discuss about literature study, methodology system will be discussed in the third part, research result is presented on the fourth part, and on the fifth part is the conclusion will and system development in the future.

2.   COMPREHENSIVE THEORETICAL BASE

The following will explain about supporting theories that will become the theory base of this research, among others:

2.1. Thayer's 2 Dimensions Model

Mood is the condition of emotion that relatively last long. Mood is different with simple emotion where emotion is less specific, less intense, and less likely to be triggered by stimulus or certain event [3]. In 1989 Robert Thayer suggested a mood model in two dimensions, it offered a simple method but effective to represent the mood.

Thayer adopted a different approach from Hevner's model. Thayer's model suggested that mood depends on two factors, those are: stress (happiness and anxiety) and energy (calm and energy) combined in two dimensions axis that formed four different quadrants, those are: contentment, representing music mood type of calm and happy; depression, representing music mood type of anxiety and depression; exuberance, representing music mood type that refers to happiness and energetic; and anxiety, representing music type of panic, anxiety and confused. For more detail can be seen in Figure 1.
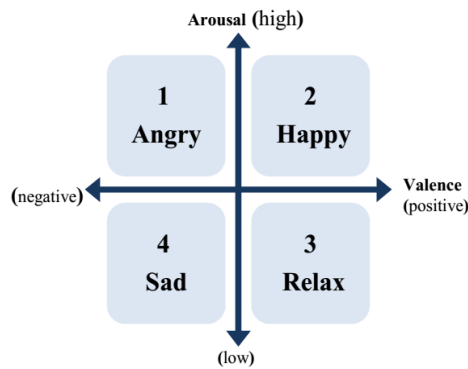

Figure 1. Thayer's emotion dimension model [3]

One of weaknesses in this model is the low of granularity which is less able to define different emotion. On other side, the superiority of this model is the simplicity in categorizing mood that can lower meaning ambiguity.

2.2. Fast Fourier Transform

The *Fast Fourier Transform* is an algorithm to count transformation of Fourier discrete quickly and efficiently. Due to continuous signals in communication system, for continuous signals case such as audio signal can use Fourier transformation. The Fourier transformation is defined by Equation 1:

$$s(f) = \int_{-\infty}^{\infty} s(t)e^{-j2\pi ft} \, dt \tag{1}$$

where s(f) is signal in frequency domain, s(t) is signal in time domain, and $e^{-j2\pi ft}$ is constant of a signal's value, f is frequency and t is time.

FFT (*Fast Fourier Transform*) is one of methods for transformation of audio signal in time domain into signal in frequency domain, which means that the audio recording process is kept in digital form in the form of audio spectrum wave with frequency base so it is easier in analyzing recorded audio frequency spectrum. On the other hand, this FFT implementation helps in the processing of filtering input signal properly to be frequency signal [11], [12].

2.3. Feature Extraction

The Feature Extraction is a process carried out to take characteristics of an input data. Furthermore, this data will become base to do a certain job in a different stage. The features produced will determine class of incoming input signal. Feature extraction involves input analysis from audio signal. In *Music Information Retrieval,* some researchers agreed that feature extraction play more important role from other phases whether for music classification purpose of for music introduction purpose. Some feature extraction method used in this research among others [13]:

First, based on statistical property from audio signal, where audio feature is analyzed based on audio signal block's length and tone level resulted from the extraction process. In this case audio feature value was obtained by using *spectral Skewness*and *kurtosis* analysis.

1. *Spectral Skewnes* is asymmetric probability distribution measurement from real value random variable in this case is the audio signal spectrum. This feature shows whether there is or there is not spectrum that skewed towards average range of certain arithmetic value. For sample of N value it forms Equation 2, where skewness value is:

$$Skewness = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{N} \cdot \sum_{n=0}^{N-1}(x(n) - \bar{x})^3}{\left(\frac{1}{N} \cdot \sum_{n-0}^{N-1}(x(n) - \bar{x})^2\right)^{3/2}} \tag{2}$$

In above Equation $\bar{x}$ shows the average of magnitude, m_3 is sample at third spectrum center, and m_2 is variant sample.

2. *Spectral Kurtosis* shows centralization of audio spectrum that can be used to show tone level counted at fourth iteration. To count *spectral kurtosis* it uses Equation 3 and Equation 4 those are:

$$m_4 = \int (f - \mu)^4 . p(f) dx \tag{3}$$

And kurtosis value is:

$$\gamma = \frac{m_4}{\sigma_4} \tag{4}$$

where, μ = mean and σ = deviation standard.

Second, audio feature is obtained based on spectral shape, in this case it can be found out based on the timbre (sound/audio color), pitch (high-low of tone), and loudness (powerful-weak of sound). To get audio feature value in this spectral shape it is obtained by *spectral centroid, rolloff, slope, spread, decrease,* and *flux*.

1. *Spectral Centroid* is the existing spectrum brightness indicator, and it shows gravity centre spectral. Spectral centroid shows sound's clarity level. Spectral centroid can be counted by using Equation 5, 6, 7 as the following:

$$\mu = \int x . p(f) dx \tag{5}$$

Where to count *p(f)*:

$$p(f) = \frac{A(f)}{\sum_f A(f)} \tag{6}$$

And to count *A(f)*:

$$A(f) = |f[x(t)]| \tag{7}$$

The x variable is frequency from collected sample. Whereas variable *p(f)* is probability to observe *f*.

2. *Spectral Rolloff* is bandwidth measure from n block which is analyzed from audio sample. Spectral Rollof is defined as frequency below accumulation of STFT (*short time Fourier transform)* power that reach percentage of certain κ value between 0.85 (85%) or 0.95 (95%). Equation 8 for spectral rolloff is as the following:

$$Spectral\ Rollof = \sum_{k=0}^{i} |X(k,n)| = \kappa.\sum_{k=0}^{\kappa/2} |X(k,n)| \tag{8}$$

3. *Spectral Slope* represents amount of slope from energy spectral as frequency function, with assumption that amplitude spectrum follows the linear model, which is:

$$A(k) = mk + b \tag{9}$$

Slope m is counted by using linear regression of Equation 10

$$m = \frac{\frac{K}{2}\sum_{k=0}^{\frac{K}{2}-1} kA(k) - \sum_{k=0}^{\frac{K}{2}-1}\sum_{k=0}^{\frac{K}{2}-1} A(k)}{\frac{K}{2}\sum_{k=0}^{\frac{K}{2}-1} k^2 - \left(\sum_{k=0}^{\frac{K}{2}-1} k\right)} \tag{10}$$

In the Equation 10 *K* is total amount of frequency value, *A(k)* is spectral magnitude with frequency index *k.*

4. Spectral Spread, also often called as instant or a moment bandwidth spectral, this spectral describes concentration of spectrum power around spectral centroid. It can be meant as deviation standard value of spectrum power around spectral centroid. Equation 11 as the following:

$$Spectral\ Spread = \sqrt{\frac{\sum_{k=0}^{\frac{\kappa}{2}-1}(k - vSC(n))^2.|X(k,n)|^2}{\sum_{k=0}^{\frac{\kappa}{2}-1} |X(k,n)|^2}} \tag{11}$$

5. Spectral Decrease is amount of decrease of spectral amplitude. This Equation came from perception and more related to frequency according to human's perception. The formula is:

$$Spectral\ Decrease = \frac{1}{\sum_{n=0}^{N-1} x(n)}.\sum_{n=0}^{N-1} \frac{x(n) - x(0)}{N-1} \tag{12}$$

In this Equation 12 $x(n)$ represents frequency or magnitude quality value of n value.

6. *Spectral Flux* is used to measure change of spectral shape, which is defined as difference of average between frames on audio spectrum respectively.

$$Spectral\ Flux = \frac{\sqrt{\sum_{k=0}^{\kappa/2-1}(|X(k,n)| - |X(k,n-1)|)^2}}{\kappa/2} \tag{13}$$

Spectral flux can be meant as basic approach to be able to see roughness level on audio spectrum.

Third, audio feature is obtained based on audio's signal properties, where audio feature to be analyzed is based on tone along the audio signal, which describes harmonies in music. To obtain audio feature value based on this signal properties it uses spectral flatness.

7. *Spectral Flatness* is tone distribution measure and spectral power in audio spectrum. *Spectral Flatness* is counted by dividing average geometric of spectrum power by average arithmetic from the spectrum power. *Spectral Flatness* is used in measurement of all spectrums. Equation 14 is as the following:

$$Spectral\ Flatness = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\frac{\sum_{n=0}^{N-1} x(n)}{N}} \qquad (14)$$

In this Equation, *x(n)* value is magnitude of n value of spectrum power with length N.

2.4. K-Nearest Neighbor (K-NN)

K-Nearest Neighbor algorithm or usually called K-NN is a data classification method that works relatively simpler compare to other data classification method. This algorithm tries to classify new data which the class is still unknown by choosing data of some k located to the nearest of new data [14]. The most class of nearest data of some *k* is chosen as class predicted for the new data. Generally, k is determined in odd amounts to avoid the emerging of same distance amounts in classification process [15].

K-NN takes decision that new data d includes in C class based on some nearest neighbor of d. If distance eucliden is used as measurement of closeness then d will become the center of hypersphere with radius r equal to the eucliden distance. What has to be done is to increase r so hypersphere loads k data. Class for data d is given based on the amount of the most class members appear in hypersphere centered on the d.

The near of far of neighbor usually is counted based on Eucliden distance with Equation 15 as the following:

$$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}$$

$$d = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \qquad (15)$$

Classification of K-NN is carried out by searching of k numbers if nearest neighbor and choosing class with the most k_i in class ω_i

2.5. Iterative Dichotomizer Tree (ID3)

ID3 is classification algorithm by using regulation induction method which is used to produce model of a data collection that is developed based on supervised learning system by J. Rose Quinlan in 1986 [16]. Briefly, ID3 algorithm is explained as the following:
1. To take out all unused attributes and the entropy is calculated which is correlated to data training.
2. Choose attribute where it has minimum entropy data.
3. Make node with the content of those attributes.

Basic method of ID3 is to choose attributes for classification by using statistic method started from the upper tree. The method to choose attribute is by using statistical property called information gain, which is defined to determine the measure of value of an attribute. Previously, entropy value of an object classified in the tree should be tested. Entropy is measure of information theory that can find out characteristic of impurity and homogeneity of data collection. From the entropy value, then information gain (IG) value of each attributes is counted. Entropy value is defined as the following:

$$E(S) = \sum_{i}^{N} -P_i * \log_2(P_i) \qquad (16)$$

where P_i is ratio of C_i class in set of sample data S = {$x_1, x_2, \dots x_k$}

$$P_i = \frac{\sum x_k \in C_i}{S} \qquad (17)$$

From entropy formula above it can be concluded that definition of entropy (S) is number of byte estimated that is needed to be able to extract a class (+ or -) from a number of random data in a sample room

S. Entropy can be said as necessity of byte to state a class. The smaller entropy value the better it is to be used in extracting a class.

After entropy value is obtained for a data collection, we can measure effectiveness of an attribute in classifying data. This effectiveness measure is called information gain. Mathematically, information gain of an attribute A is stated as the following:

$$G(S, A) = E(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} E(S_v) \qquad (18)$$

where:

a. Quality $W_i = \frac{|S_v|}{|S|}$ is ratio of data with attribute v in sample set
b. A: attribute; V: a possible value for attribute A
c. Values (A): possible set for attribute A
d. $|S_v|$: amount of sample for value v
e. $|S|$: total amount of data sample
f. $E(S_v)$: *entropy* for samples that have value v

3. RESEARCH METHODOLOGY

Music mood classification system flow generally can be seen in Figure 2. From a number of obtained mood category it came from social annotation/tag from site. Furthermore it is carried out music file pre-processing, by using the refrain part from the music. Music clip refrain duration is determined to be only 30 seconds and furthermore it is kept with format of .wav with mono audio channel. Extraction process is started by changing input music file signal to be frequency domain by using *Fast Fourier Transform* (*FFT*) method. FFT is used for audio signal transformation in time domain to be signal in frequency domain. This music file produced by FFT furthermore entering feature extracting stage by using spectral analysis. Result from *spectral analysis* (*spectral centroid, spectral skewness, spectral rolloff, spectral slope, spectral kurtosis, spectral spread, spectral decrease, spectral flux,* and *spectral flatness)* is a set of feature set value consists of 9 attribute values for each of music file. These 9 attribute values are the values that become special characteristic of music file that is used as training data and testing data for classification process using the KNN and ID3 algorithm.
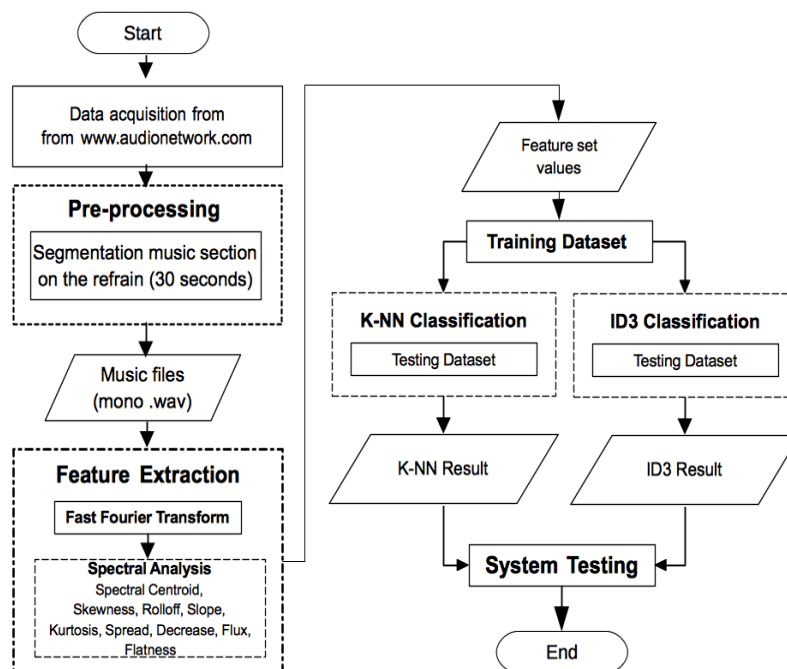

Figure 2. Music Classification System Flow to Mood

System testing is performed to measure percentage of system's classification accuracy by using both of these classification algorithms. Testing is carried out with benchmarking that refers to reference standard to object that will be compared. In this case compiling of training data and testing data have been classified based on mood type in site, those are *contentment*, *exuberance*, *depression* and *anxious*. With total of training data is 400 files for the fourth mood type and testing data is 400 files.

We calculate the ratio of the level of accuracy of classification by category mood using K-NN and ID3 algorithm defined as the following:

$$\text{accuracy} = \frac{\text{the amount of correct data test}}{\text{the amount of data test based on mood}} \times 100\% \qquad (4.1.)$$

For classification time of both algorithms are counted when user pushes the "classification" button on the system, so it will result label of music mood classification.

4. RESULTS AND ANALYSIS
4.1. Characteristics of Training Data Feature

In these experiments we did a training phase using 400 audio files have a source on annotation/tag of music expertise in site www.audionetwork.comwhich are used in this research as training data, which is spread to be 100 audio files of category 1 (Contentment), 100 audio files of category 2 (Exuberance), 100 audio files of category 3 (Depression), and 100 audio files of category 4 (Anxious). Each audio file used in the training phase and classification has a duration of 30 second clips. In the pre-processing stage is an audio file is converted into WAV format with mono type, sampling rate of 44.1 kHz. The system interface of data training developed.Generally, in an audio file that is used as training data it has 10 attributes value such as 9 attributes from spectral analysis and 1 attribute as code of music mood category. This will be kept in a training dataset file.

400 records in this training data has comparative feature characteristic values of the type of mood. Figure 3 shows that in a training data formed it have highest feature values respectively in spectral rolloff, spectral centroid, spectral spread and spectral kurtosis. This highest value in each of mood can be seen in mood of contentment, exuberance, and anxious, whereas the lowest value is in the mood of depression. The exception is in spectral kurtosis where the highest spectral feature values are respectively in the mood of contentment, depression, and anxious, and the lowest is in the mood of exuberance.

Overall tranining data then stored in a file called training dataset for the next phases of classification process using K-NN and ID3.
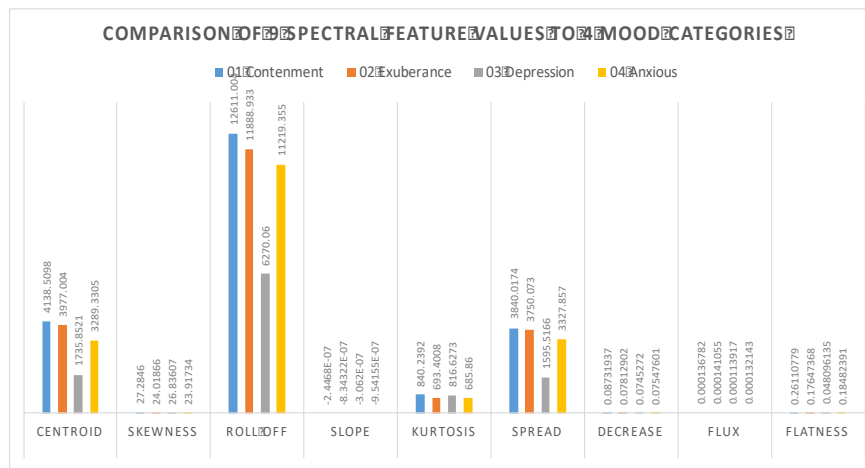

Figure 3. Comparison of 9 spectral feature values to 4 mood categories

4.2 Testing Result of K-Nearest Neighbor Classification

Classification interface has function as interface for user to perform classification process by using KNN or ID3 method by using training dataset which has been input previously. The difference is that in

KNN interface there is a textbox that is used to input the KNN neighboring k value. Data input in this system is the audio data with format mono *.wav. The training data that is used in both of these interfaces are a set of feature set value result from spectral analysis that has source from audio file with four types of mood category. This training data is used as reference data to find out level of classification accuracy and time needed in classification process in music file input to mood. The followings are appearance for interface of KNN and ID3 classification system (Figure 4, 5).

Both of these interfaces have 4 panels consist of, first panel, the Music Mood Data Processing where including of several buttons that are used as trigger of classification stages. Second, the graphic panel result of signal sampling, consist of original audio signal sampling and signal sampling of audio result of FFT. The third is the graphic panel of original audio signal spectrogram and audio signal spectrogram result of FFT. The fourth is panel of classification result that consists of classification result label and classification processing time.

For classification result using KNN algorithm at 40-400 this testing data is carried out with k value that is 1 to 16, it is obtained percentage of average classification accuracy for highest value is in k value = 3 for 86.55%, and average processing time per music file is 0.01021 seconds. This result can be seen in Figure 6 and Table 1. Whereas for classification result using ID3 algorithm uses 40-400 testing data and it is obtained percentage of average classification accuracy for 59.33%, and average processing time per music file is 0.05091 seconds.
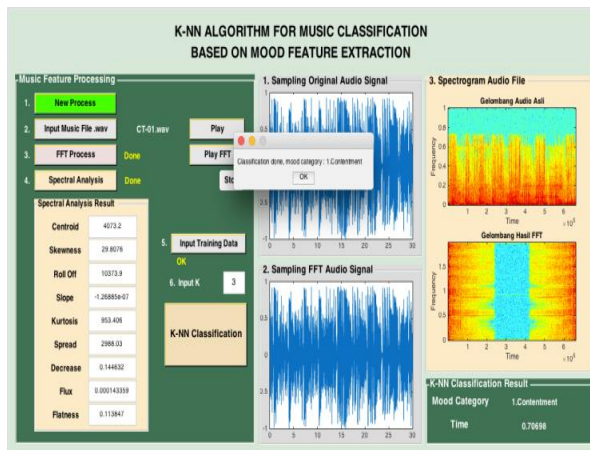

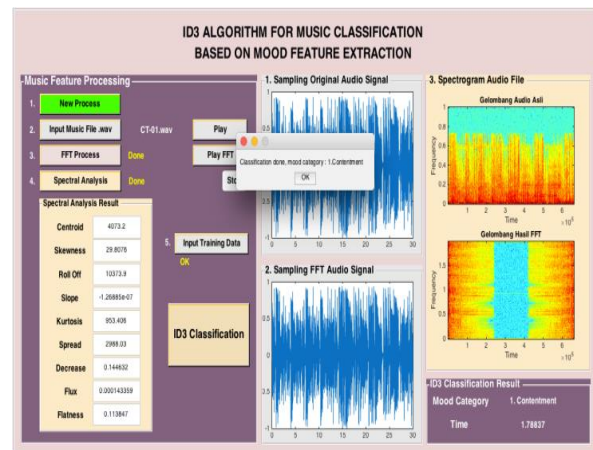Figure 4. Interface of KNN classification system
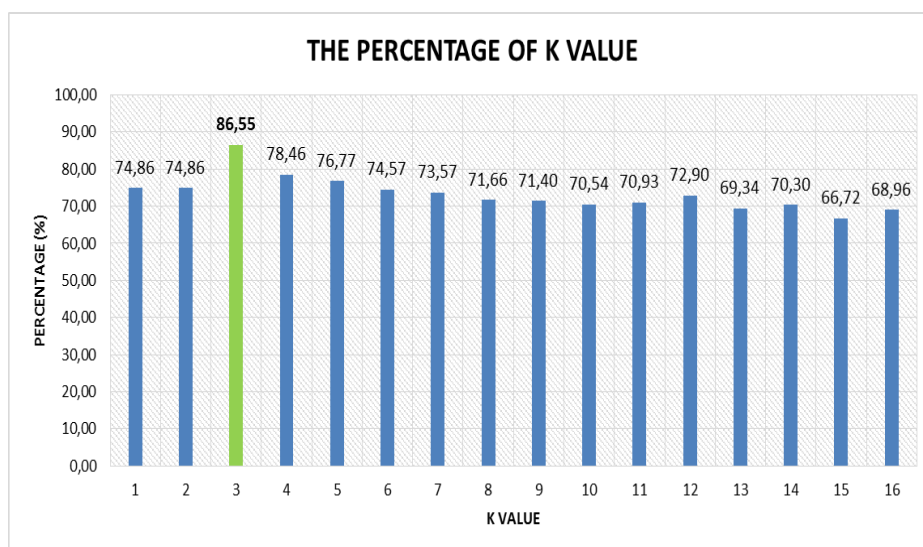

Figure 5. Interface of ID3 classification system

Figure 6. Result of classification accuracy uses K-NN

Table 1. Accuracy and Average Classification Processing Time of KNN

No	Amount of testing data	KNN		ID3	
		Accuracy of $k = 3$	Average Time (seconds)	Accuracy	Average Time (seconds)
1	40	87,5 %	0.01059	67.5 %	0.06915
2	80	85 %	0.01084	65 %	0.04746
3	120	89,2 %	0.01091	60.8 %	0.05443
4	160	89,3 %	0.01047	60.6 %	0.04752
5	200	87,5 %	0.01005	59 %	0.04336
6	240	87,5 %	0.01038	57.1 %	0.05473
7	280	88,7 %	0.00991	55 %	0.05133
8	320	85 %	0.00974	55.6 %	0.04888
9	360	84,7 %	0.00964	56.4 %	0.04691
10	400	84 %	0.00956	56.2 %	0.04530
Accuracy and average classification		86,55 %	0.01021 seconds	59.33 %	0.05091 seconds

5.   CONCLUSION

Generally, result of system classification built by using KNN algorithm has been able to perform music classification to mood by using Thayer's model with percentage of accuracy is good enough which is 86.55%, and average processing time is 0.01021 seconds. Whereas system with ID3 algorithm produced classification accuracy for 59.33% and average processing time is 0.050911275 seconds. Suggestion for further system development, in music file pre-processing stage it can be tried to use file with format of stereo *.wav with more optimal refrain duration. Moreover, the K-NN algorithm has weakness in determining accurate k value to produce the best system accuracy. Therefore, it is expected that in further system development optimizing is carried out in choosing the k value parameter automatically.

REFERENCES

[1]   S. Pouyanfar and H. Sameti, "Music emotion recognition using two level classification", in *2014 Iranian Conference on Intelligent Systems (ICIS)*, 2014, pp. 1–6.
[2]   Y. Song, S. Dixon, and M. Pearce, "Evaluation of Musical Features for Emotion Classification", in *ISMIR*, 2012, pp. 523–528.
[3]   Thayer R.E., *The Biopsychology of Mood and Arousal*. Oxford University Press, 1990.
[4]   V. Hampiholi, "A method for music classification based on perceived mood detection for Indian bollywood music", in *Proceedings of World Academy of Science, Engineering and Technology*, 2012, p. 507.
[5]   B.G. Patra, D. Das, and S. Bandyopadhyay, "Automatic music mood classification of Hindi songs", in *Sixth International Joint Conference on Natural Language Processing*, 2013, p. 24.
[6]   M.B. Mokhsin, N.B. Rosli, S. Zambri, N.D. Ahmad, and S. Rahah, "Automatic Music Emotion Classification Using Artificial Neural Network Based on Vocal and Instrumental Sound Timbres", *J. Comput. Sci.*, vol. 10, no. 12, pp. 2584–2592, Dec. 2014.
[7]   K.C. Dewi and L.A.A.R. Putri, "Music Mood Player Implementation Applied In Daycare Using Self Organizing Map Method", *J. Buana Inform.*, vol. 2, no. 2, 2011.
[8]   A. Setiawan, "Analisis Klasifikasi Suara Berdasarkan Gender dengan Format Wav Menggunakan Algoritma K-Means", *Sains Dan Teknol.*, vol. 2, no. 2, 2009.
[9]   G. Vyas and M.K. Dutta, "Automatic mood detection of indian music using MFCCs and K-means algorithm", in *Seventh International Conference on Contemporary Computing (IC3)*, 2014, pp. 117–122.
[10] S. Oh, M. Hahn, and J. Kim, "Music mood classification using intro and refrain parts of lyrics", in *International Conference on Information Science and Applications (ICISA)*, 2013, pp. 1–3.
[11] R.Y. Sipasulta, A.S. Lumenta, and S.R. Sompie, "Simulasi Sistem Pengacak Sinyal Dengan Metode FFT (Fast Fourier Transform)", J. Tek. Elektro Dan Komput. UNSRAT, vol. 3, no. 2, pp. 1–9, 2014.
[12] S. Gopinathan, R. Kokila, and P. Thangavel, "Wavelet and FFT Based Image Denoising Using Non-linear Filters", *IJECE*, vol. 5, no. 5, 2015.
[13] A. Lerch, *An introduction to audio content analysis: applications in signal processing and music informatics*. Hoboken, NJ: Wiley, 2012.

[14] M. Abdar, S.R.N. Kalhori, T. Sutikno, I.M.I. Subroto, and G. Arji, "Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases", *IJECE*, vol. 5, no. 6, pp. 1569–1576, 2015.

[15] Eko Prasetyo, Data Mining : Konsep dan Aplikasi Menggunakan Matlab. Yogyakarta: ANDI, 2012.

[16] Vasudevan, "Iterative Dichotomiser-3 Algorithm in Data Mining Applied to Diabetes Database", *J. Comput. Sci.*, vol. 10, no. 7, pp. 1151–1155, Jul. 2014.

BIOGRAPHIES OF AUTHORS

Dr. Ir. Made Sudarma, M.A.Sc.
Department of Electrical and Computer Engineering
Faculty of Engineering, Udayana University
Udayana Campus Rd, Jimbaran, Badung
Bali - INDONESIA
Email: imasudarma@gmail.com

I Gede Harsemadi, S.Kom., M.T.
Department of Information System
STMIK STIKOM BALI
Jl. Raya Puputan No. 86, Denpasar
Bali 80234, Indonesia. Tel./Fax: +62361 244445
Email: gedeharsemadi@yahoo.com