

Music Genre Classification Using Modified K-Nearest Neighbor (MK-NN)

I Nyoman Yusha Tresnatama Giri^{a1}, Luh Arida Ayu Rahning Putri^{a2}, Gst. Ayu Vida Mastrika Giri^{a3}, I Gusti Ngurah Anom Cahyadi Putra^{a4}, I Made Widiartha^{a5}, I Wayan Supriana^{a6}

^aInformatics Department, Faculty of Mathematics and Natural Sciences, University of Udayana
South Kuta, Badung, Bali, Indonesia

¹yusatresnatama11@gmail.com

²rahningputri@unud.ac.id

³vida.mastrika@cs.unud.ac.id

⁴anom.cp@unud.ac.id

⁵madewidiartha@unud.ac.id

⁶wayan.supriana@unud.ac.id

Abstract

The genre of music is a grouping of music according to their resemblance to one another and commonly used to organize digital music. To classify music into certain genres, one can do it by listening to the music one by one manually, which will take a long time so that automatic genre assignment is needed which can be done by a number of methods, one of which is the Modified K-Nearest Neighbor.

Modified K-Nearest Neighbor method is a further development of its former method called K-Nearest Neighbor method which adds several additional processes such as validity calculations and weight calculations to provide more information in the selection class for the testing data.

Research to find the best H value shows that the H = 70% of the training data is able to produce an accuracy of 54.100% with K = 5 and the proportion ratio of test data and training data is 20:80 (fold 5). The best H value is then used for further testing, which is to compare the K-Nearest Neighbor method with the Modified K-Nearest Neighbor method using two different proportions of test data and training data and each proportion of data also tests a different K value. The results of the classification comparison of the two methods show that the Modified K-Nearest Neighbor method, with the highest accuracy of 55.300% is superior to the K-Nearest Neighbor method with the highest accuracy of 53.300%. The two highest accuracies produced in each method were obtained using K = 5 and the proportion ratio of test data and training data is 10:90 (fold 10).

Keywords: Classification, K-Nearest Neighbor, Modified K-Nearest Neighbor, Fold

1. Introduction

Music genres are a commonly used way of organizing digital music [8]. Classification of genres can be easily done by listening directly to the music files manually. This manual classification has been carried out previously by experts named Tzanetakis and Cook in 2002. Currently, with the increasing number of music circulating, manual genre assignment will take a long time, so it is necessary to automatically assign genres that can help, reduce or replace roles of humans in giving genres to music [4]. Music genre classification is an important thing that has been studied for many years by the Music Information Retrieval (MIR) community since 2002 and until now, the problem of classifying music genres still continues in the Music Information Retrieval Evaluation eXchange (MIREX) which is an annual evaluation program held by an organization called the International Society for Music Information Retrieval (ISMIR).

There are several classification methods, including K-Nearest Neighbor, Support Vector Machine, Naive Bayes, and others. Of the many existing classification methods, one that is often used is K-Nearest Neighbor (K-NN). This method is able to predict the class, which in this study is called the

genre, from the training data and classify the test data with K nearest neighbors [9]. In terms of accuracy, the K-NN method can still be improved because until now, improvements to the K-NN method continue to be carried out in research [8]. One of the methods resulting from the improvement of the K-NN method is the Modified K-Nearest Neighbor (MK-NN) method. The basic difference that distinguishes MK-NN from K-NN is the addition of a validation function on training data and weight voting on all test data using data validity [15]. With the validation and weighting process on the MK-NN method, it will produce better accuracy if the value of both processes is high. With these two processes, it is hoped that the Modified K-Nearest Neighbor (MK-NN) method can correct any deficiencies in the accuracy calculation process in the K-NN method [15].

In the classification of music genres, the first thing to do before classifying is to extract features from the music itself. In this study, there are 5 features extraction methods to be used for extraction which is MFCC, chroma frequencies, spectral centroid, spectral roll-off, and zero crossing rate. The five features used are from research reference which conducted to classify music and all the five features have a better result in classifying lot of music data [12].

In a previous research, study about music genre classification, has been conducted by [12] using 5 feature extraction methods (MFCC, chroma frequencies, spectral centroid, spectral roll-off, and zero crossing rate) to classify songs to 9 different genres. In this study, there are 3 classification methods used with different accuracy results namely K-Nearest Neighbor reaching 64%, Linear Support Vector Machine reaching 60%, and Poly Support Vector Machine reaching 78%.

In other studies, conducted by [10], compares two methods which is K-NN and MK-NN in identifying dental and oral diseases. This study uses 6 classes and proves that the training data is 100 and the test data is 30 and the value of $K = 3$, the MK-NN method can identify types of dental and oral diseases by reaching 76.66% while K-NN reaches 30%. From this study, MK-NN having higher accuracy than K-NN. This is due to the calculation of the validity value which will affect the weight voting and also the accuracy of the MK-NN itself.

Based on both studies above, authors aim to compare the K-NN method with the MK-NN method on the classification of music genres that utilize the GTZAN dataset according to the [12] as reference hoping that with the use of MK-NN method, the accuracy of K-NN method can be improved.

2. Research Methods

In this research, the audio data that will be used is GTZAN dataset which has 10 genres with 100 songs for each genre. The process of how the system will works can be seen in Figure 1.

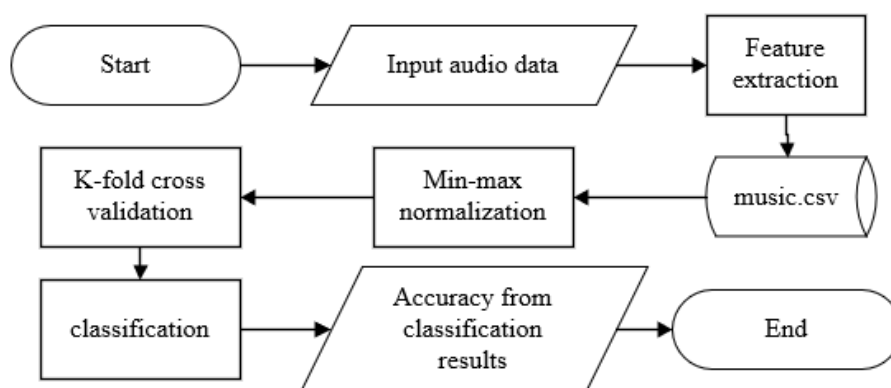


Figure 1. Flowchart of music genre classification

The process starts from feature extraction to get the features of all songs from GTZAN dataset. The features obtained will be saved into a dataset. Next step is whatgmai to normalize the dataset using min-max normalization technique. Normalized data then will be divided equally by using K-fold cross validation technique so that each data have chances to become test data and train data. After cross validation method, data will be classified using both K-Nearest Neighbor and Modified K-Nearest

Neighbor to produce an output in the form of genre prediction which will be calculated to see how high the accuracy will be.

2.1. Research Data

Data used in this study is secondary data obtained through a website called MARSYAS which provides access to the GTZAN dataset. This GTZAN dataset has been collected and generalized by experts named Tzanetakis and Cook in 2002. The GTZAN dataset is still used by researchers in several studies, as an example of research conducted by [10] which demonstrate the fusion of visual and audio features to improve classification performance using the GTZAN dataset. Another example is research conducted by [7] using GTZAN dataset, they introduce Learning Vector Quantization (LVQ) that combined with Self Organizing Map (SOM) based on feature of entropy of wavelet coefficients which results a better accuracy than using LVQ alone.

The GTZAN dataset has a collection of 1000 songs with a duration of 30 seconds for each song. There are 10 different genres with 100 songs for each genre. All song collections are 22050Hz Mono 16-bit in *.wav format.

2.2. Feature Extraction

The goal of this step is to get a new dataset in the form of numeric values of each feature extracted from each song. There are 5 extraction features that will be used namely MFCC, chroma frequencies, spectral centroid, spectral roll-off, and zero crossing rate. Several libraries are involved for feature extraction namely scipy, librosa, and pandas.

a. Mel Frequency Cepstrum Coefficient (MFCC)

MFCC is a series of short-term power spectrum in an audio file. The MFCC models the characteristics of the human voice. The feature vector output from this extraction reaches up to 39 feature vectors [14]. In this study, 13 feature vectors were taken [12].

b. Chroma Frequencies

Chroma frequencies is one of the features that discretizes the spectrum into chromatic notes or keys and represents each note or key totaling 12. The number of feature vectors obtained is obviously will be 12 feature vectors. This feature provide a strong way to describe the similarities between one audio to another [12].

c. Spectral Centroid

This feature characterizes the signal spectrum and indicates the location of the center of gravity of the magnitude spectrum. In human perception, this is like giving the bright impression of a sound. The spectral centroid can be evaluated as a weighted average of the spectral frequencies. The higher the value of this centroid, the higher the brightness of the high frequency sound [12].

d. Spectral Roll Off

Spectral roll off can be defined as the M bin frequency below where 85% of its magnitude distribution is concentrated. In addition to the Spectral centroid, Spectral Roll Off is also a measurement of the spectral shape of the audio [12].

e. Zero Crossing Rate

Zero crossing rate is the number of times a wave crosses 0. Usually very well used for audio that has percussion instruments such as metal and rock genres [12].

2.3. Min-Max Normalization

Because different features obtained from feature extraction method has the possibility of producing a range of data that is not well distributed, then normalization is carried out first before going to the classification stage. Min-max normalization can improve data that is outliers so as to facilitate the calculation process in the classification later [1]. The order of the min-max normalization method is:

- a. Get the maximum value of each feature in a data.
- b. Get the minimum value of each feature in a data.
- c. Apply equation (1) for each data in each existing feature.

$$norm(x) = \frac{x - minValue}{maxValue - minValue} \quad (1)$$

Where $norm(x)$ is the x value of the i -th data on the i -th feature that has been normalized. The $minValue$ and $maxValue$ variables are the minimum value of the i -th feature and the largest value of the unnormalized i -th feature, and x is the unnormalized data.

2.4. Classification

Classification is done after we have a dataset either before or after performing the feature selection.

a. K-Nearest Neighbor

K-Nearest Neighbor works by classifying test data by studying its proximity to training data. This class of test data is obtained from the majority of classes in K data, K is the number of data with the closest distance or can also be referred to as the closest neighbor of the test data [11]. There are many distance calculations available in the K-Nearest Neighbor method, one of which is Euclidean. The purpose of the calculation is to define the distance between the two points which is the point in the training data (x) and the point in testing data (y). Calculation of euclidean distance can be done with equations (2).

$$d(xi, yi) = \sqrt{\sum_{i=0}^n (xi - yi)^2} \quad (2)$$

Where d is the distance between the point on the training data x and the testing data y to be classified. x , y , and i represent attribute and n is the dimension of the attribute.

b. Modified K-Nearest Neighbor

The Modified K-Nearest Neighbor (MK-NN) algorithm is the development of the K-Nearest Neighbor method with the addition of several processes, namely the calculation of validity values and weight calculations [13].

The MK-NN method is not much different from the K-NN which also has a distance calculation process and calculates the proximity between the test data and the training data, the amount of which depends on the K parameter. to each result from the closest distance between the test data and the training data. In the K-NN algorithm, the most class of distance calculations will be chosen as the result of the K-NN classification, but of course the classification results are not necessarily correct. Therefore, each result of the distance calculation is given a weight so that the largest weight will be selected as the classification class. In addition to calculating the weights, it is also necessary to calculate the similarity between the training data and calculate the validity value to support the weight calculation process later [3].

Each data in the training data must be validated before carrying out the next process. The validity of each data depends on each of its neighbors. The validation process is carried out for all data on the training data which then, the validity value generated from the process will be used as more information to carry out the weight calculation process later (Wafiyah et al, (2017). Calculation of validity can be seen in equation (3).

$$Validity(x) = \frac{1}{H} \sum_{i=0}^H S(label(x), label(Ni(x))) \quad (3)$$

Where:

H is the number of points or nearest neighbors. In this case H can be determined by the researchers themselves [5] or can be equated in value with the K value in the K-NN method [6].

label(x) is class x

label(Ni(x)) is the class from the i-th point closest to x

S is used to calculate the similarity between point x and the i-th data from the nearest neighbor of x. This S function can be defined in equation (4).

$$S(a, b) = \begin{cases} 1 & (a = b) \\ 0 & (a \neq b) \end{cases} \quad (4)$$

Where:

a is the class of the training data x

b is the class of the i-th training data closest to x

The weight voting stage is carried out after getting the validity value and the closest Euclidean distance from the i-th test data. Weight voting can be calculated by equation (5).

$$W_{(i)} = Validitas(i) \times \frac{1}{de + a} \quad (5)$$

Where:

$W_{(i)}$ is the weight of the i-th neighbor (training data)

$Validity(i)$ is the validity value of the i-th neighbor (training data)

de is the value of the Euclidean distance from the test data to the i-th neighbor (training data)

a is the smoothing parameter [5].

Calculation of weight voting is needed in the MK-NN method to provide additional information in determining the class of test data.

2.5. K-Fold Cross Validation

K-fold cross validation is one of the techniques that can be used to divide the proportion between test data and training data. K-fold cross validation can be referred to as rotation estimation which is used to make model predictions and estimate how accurate a predictive model is when run in practice. The K-fold cross validation technique breaks the data into K parts of data sets of the same size in order that each data has the opportunity to become test data and training data. This technique can reduce bias in the data. Training and testing were carried out K-times [2].

3. Result and Discussion

3.1. Comparative Test of H Parameters Using the Modified K-Nearest Neighbor Method

The test results for the first test scenario are to determine the effect of different H values on accuracy and find the best H value for calculating the validity of the MK-NN method which can be seen in Figure 2.

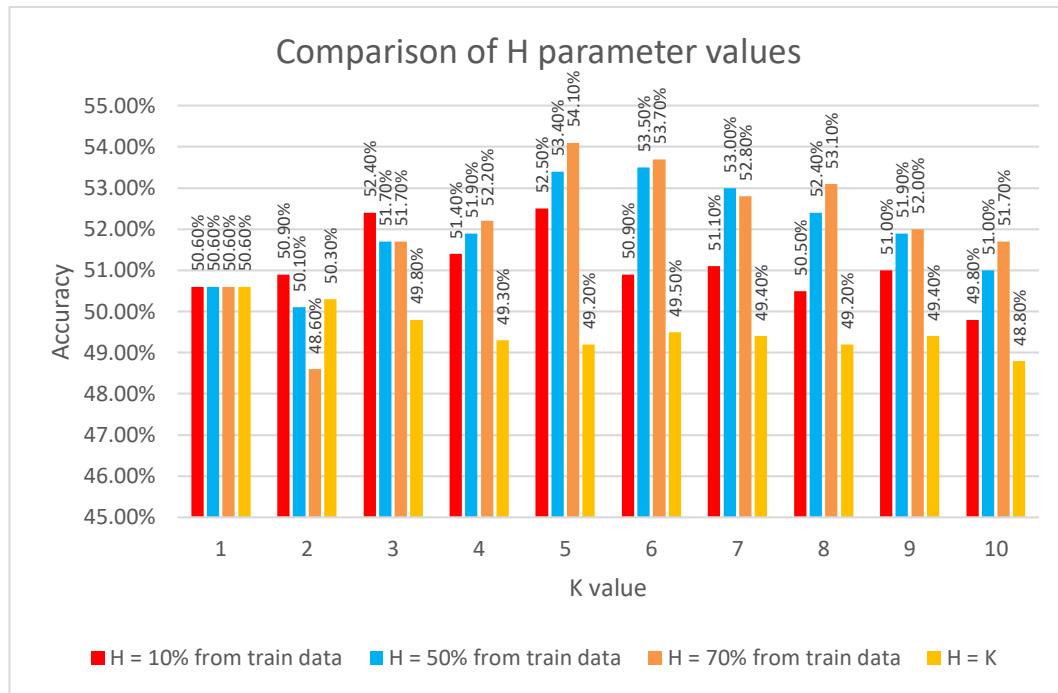


Figure 2. Bar Graph of Test Results to Determine Best H Value for MK-NN Method

Parameter H with a value of 10% from the training data got the lowest to the highest accuracy with an accuracy range of 49,800% - 52,500%, the highest accuracy of 52,500% was obtained using K = 5.

Parameter H with a value of 50% of the training data got the lowest to the highest accuracy with an accuracy range of 50.100% - 53.500%, the highest accuracy of 53.500% was obtained using K = 6.

Parameter H with a value of 70% from the training data got the lowest to the highest accuracy with an accuracy range of 48.600% - 54.100%, the highest accuracy of 54.100% was obtained using K = 5.

The H parameter with the same value as the K value gets the lowest to the highest accuracy with an accuracy range of 48.800% - 50.600%, the highest accuracy of 50.600% is obtained by using K=1. When viewed from the highest average, the best H parameter is obtained with a value of 70% from the training data with K = 5, namely with an accuracy of 54.100%, so that this H value will then be used to calculate the validity of the second and third test scenarios.

3.2. Comparison Test of K Parameters with a Proportion Ratio of 20:80 (Fold 5) Using the K-Nearest Neighbor Method and the Modified K-Nearest Neighbor Method

The test results for the second test scenario is to see the effect of different K parameters using the proportion of test data and training data 20:80 (fold 5) can be seen in the bar graph in Figure 3.

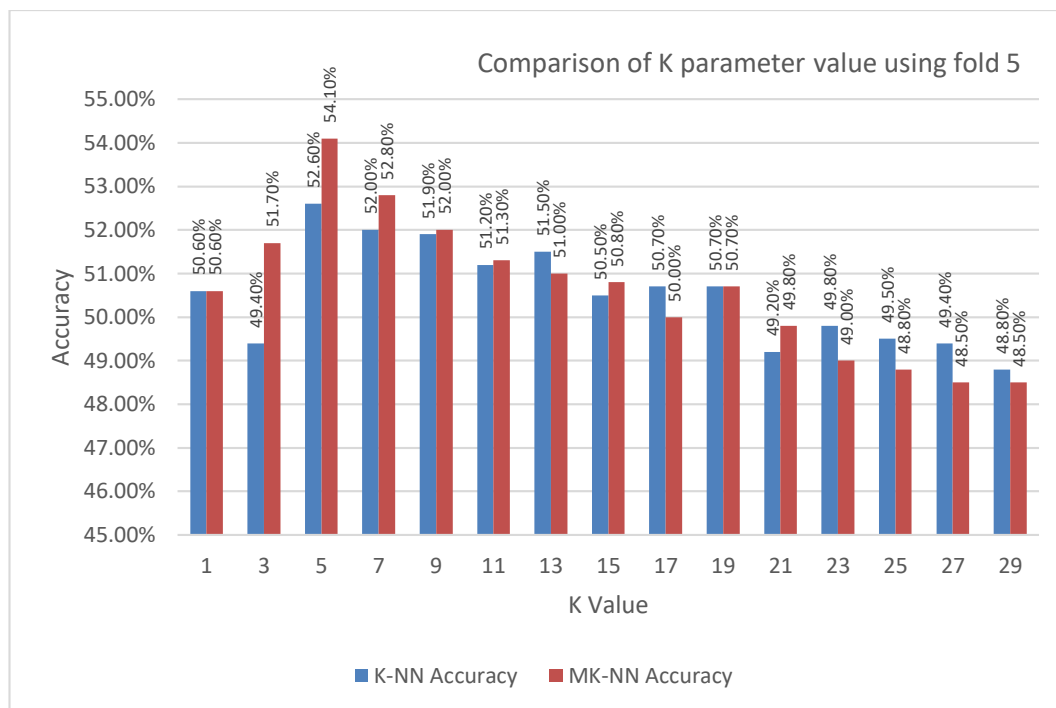


Figure 3. Bar Graph of Test Result Using Fold 5

Results shown on Figure 3 indicate that the K-NN method gets accuracy from the lowest to the highest with an accuracy range of 48.800% - 52.600% and the MK-NN method gets accuracy from the lowest to the highest with an accuracy range of 48.500% - 54.100%

The highest accuracy for the K-NN method with fold 5 was obtained at the value of K = 5 with an accuracy of 52.600% and the lowest accuracy was obtained at the values of K = 29 with an accuracy of 48.800%.

The highest accuracy for the MK-NN method with fold 5 was obtained at the value of K = 5 with an accuracy of 54.100% and the lowest accuracy was obtained at the value of K = 27 and 29 with an accuracy of 48.500%.

The bar graph in Figure 3 shows that the accuracy obtained at the value of K = {1, 3, 5, 7, 9, 11, 13, 15, 17, 19} still does not show a consistent decrease and the highest peak of accuracy is at K = 5 for the K-NN and MK-NN methods. Accuracy begins to show a gradual decrease at the value of K = 21 and so on. The total average accuracy of the K-NN method is 50.520%, while the total average accuracy of the MK-NN method is 50.640% which proves that the MK-NN method outperforms the K-NN method.

3.3. Comparison Test of K Parameters with a Proportion Ratio of 10:90 (Fold 10) Using the K-Nearest Neighbor Method and the Modified K-Nearest Neighbor Method

The test results for the second test scenario is to see the effect of different K parameters using the proportion of test data and training data 10:90 (fold 10) can be seen in the bar graph in Figure 4.

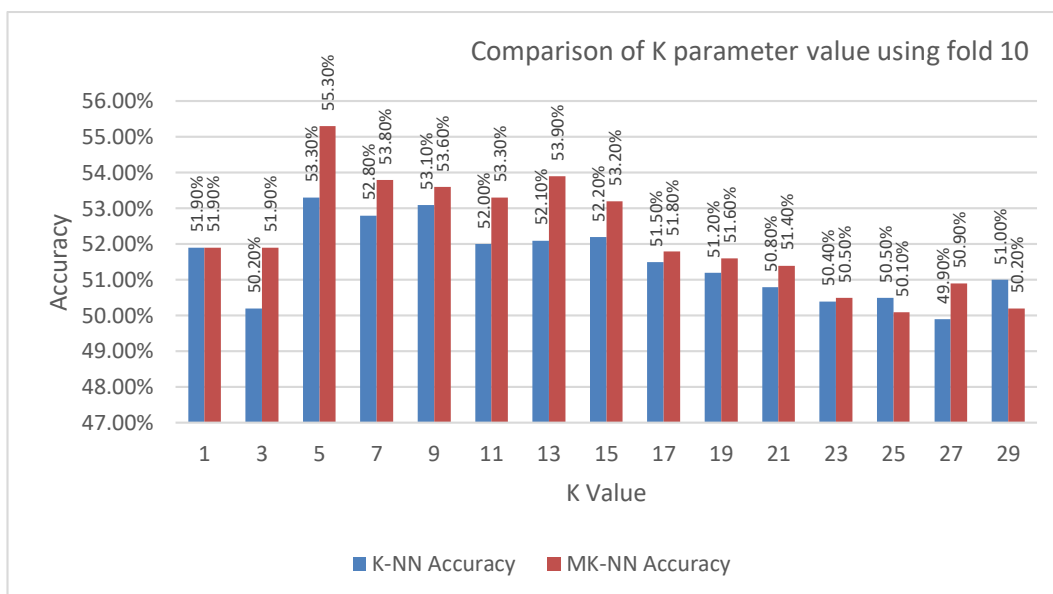


Figure 4. Bar Graph of Test Result Using Fold 10

Results shown on Figure 4 indicate that the K-NN method gets accuracy from the lowest to the highest with an accuracy range of 49.900% - 53.300% and the MK-NN method gets accuracy from the lowest to the highest with an accuracy range of 50.100% - 55.300%

The highest accuracy for the K-NN method with fold 10 was obtained at the value of K = 5 with an accuracy of 53.300% and the lowest accuracy was obtained at the values of K = 27 with an accuracy of 49.900%.

The highest accuracy for the MK-NN method with fold 10 was obtained at the value of K = 5 with an accuracy of 55.300% and the lowest accuracy was obtained at the value of K = 25 with an accuracy of 50.100%.

The bar graph in Figure 4 shows that the accuracy obtained at the value of K = {1, 3, 5, 7, 9, 11, 13, 15} still does not show a consistent decrease and the highest peak of accuracy is at K = 5 for the K-NN and MK-NN methods. Accuracy begins to show a gradual decrease at the value of K = 17 and so on. The total average accuracy of the K-NN method is 51.527%, while the total average accuracy of the MK-NN method is 52.227% which proves that the MK-NN method outperforms the K-NN method, same as previous scenario using fold 5.

3.4. Test Result Analysis

In Figure 2, several values for the H parameter are tested to find the best value for calculating the validity of the Modified K-Nearest Neighbor method and from the results of the first scenario testing that tests this H parameter, the best H value is sought by looking at the highest accuracy generated from several K values that have been determined for each H value tested. The highest average was obtained with a value of H = 70% of the training data using K = 5, this indicates that for this study, with a large number of training data (800 training data), a large H value tends to provide a better validity value. in the Modified K-Nearest Neighbor method which is able to increase the accuracy obtained from the method. The best H value is then used for testing the next test scenario.

In Figure 3, a test was conducted with several K values with a ratio of the proportion of test data and training data of 20:80 and in Figure 4, a test was carried out with several K values with a ratio of the proportion of test data and training data of 10:90. From the two tests, the overall accuracy results are superior to the Modified K-Nearest Neighbor method which obtains the highest accuracy, namely 55.300% with a K = 5 value and using a data proportion of 10:90 while the highest accuracy by the K-

Nearest Neighbor method is 53.300% with value of $K = 5$ and by using the proportion of data 10:90. Both tests show that the proportion of test data and training data is able to affect the classification results obtained from the K-Nearest Neighbor and Modified K-Nearest Neighbor methods. The K value can also affect the classification results where the MK-NN method is able to maintain its superiority compared to the K-NN method when given a large K value even though both methods continue to experience a decrease in accuracy where this can be caused by weight calculations that strengthen the MK-NN method in selecting the prediction class from the tested data.

4. Conclusion

This study has succeeded in classifying using the K-Nearest Neighbor and Modified K-Nearest Neighbor methods. Several conclusions can be drawn based on the results of research that has been carried out. The first test scenario is to know the effect of the value on the H parameter for validation calculations on the Modified K-Nearest Neighbor method show that the H value can affect the accuracy of the method. The best H value was 70% of the training data, the use of this H value was able to get the highest accuracy of 54.100% with a $K = 5$ value and the ratio of the proportion of test data and training data was 20:80 in the Modified K-Nearest Neighbor method. The second test is to know the effect of the value on the K parameter and the proportion of different test data and training data on the K-Nearest Neighbor and Modified K-Nearest Neighbor methods indicate that the K parameter and the proportion of data are able to affect the accuracy of the two methods. The highest accuracy obtained in the Modified K-Nearest Neighbor method is 55.300% using the value of $K = 5$ and the ratio of the proportion of test data and training data is 10:90, while the highest accuracy obtained by the K-Nearest Neighbor method is 53.300% using K value and the ratio of the proportion of test data and training data are the same. The third scenario test is the comparison of the results in classification test between the K-Nearest and Modified K-Nearest Neighbor methods which show that overall, the Modified K-Nearest Neighbor method is able to outperform the K-Nearest Neighbor method in terms of accuracy and both methods get the highest accuracy of each, namely 53,300%. for K-Nearest Neighbor and 55.300% for Modified K-Nearest Neighbor using the value of $K = 5$ and the ratio of the proportion of test data and training data 10:90.

References

- [1] D. A. Nasution, H. H. Khotimah, N. Chamidah, "Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-NN" *Computer Engineering System and Science*, Vol. 4, No. 1, p. 78-82, 2019.
- [2] F. Tempola, M. Muhammad, A. Khairan, "Perbandingan Klasifikasi Antara KNN dan Naïve Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation" *Jurnal Teknologi Informasi dan Ilmu Komputer*, Vol. 5 Issue.5, 2018.
- [3] F. Wafiyah, N. Hidayat, R.S. Perdana, "Implementasi Algoritma Modified KNearest Neighbor (MKNN) untuk Klasifikasi Penyakit Demam" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2017.
- [4] Gst. A. V. M. Giri, "Klasifikasi dan Retrieval Musik Berdasarkan Genre (Sebuah Studi Pustaka)" *Jurnal Ilmiah Ilmu Komputer Universitas Udayana*, 2017.
- [5] H. Parvin, H. Alizadeh, B. Minati, "A Modification on K-Nearest Neighbor Classifier" *Global Journal of Computer Science and Technology*, Vol. 10, Issue 14, 2010.
- [6] I. Agustin, Y.N. Nasution, Wasono, "Klasifikasi Batubara Berdasarkan Jenis Kalori Dengan Menggunakan Algoritma Modified K-Nearest Neighbor (Studi Kasus: PT.Pancaran Surya Abadi)" *Jurnal EKSPONENSIAL*, Vol. 10, Issue.1, 2019.

- [7] L. A. A. R. Putri, S. Hartati, "Klasifikasi Genre Musik Menggunakan Learning Vector Quantization dan Self Organizing Map" *Jurnal Ilmiah ILMU KOMPUTER Universitas Udayana*, Vol. 9, No. 1, p. 14-22, 2016.
- [8] L. Nanni, Y.M.G. Costa, A. Lumini, M.Y. Kim, S.R. Baek, "Combining visual acoustic features for music genre classification" *Expert System With Applications*, p. 1-10, 2016.
- [9] M. Holeña, P. Pulc, M. Kopp, "Classification Methods for Internet Applications" *Poland: Springer International Publishing*, 2020.
- [10] M.R. Ravi, Indriati, S. Adinugroho, "Implementasi Algoritma Modified KNearest Neighbor (MKNN) Untuk Mengidentifikasi Penyakit Gigi dan Mulut" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2019.
- [11] N. Fetra, M. Irsyad, "Aplikasi Pencarian Chord dalam Membantu Penciptaan Lagu Menggunakan Algoritma Fast Fourier Transform (FFT) dan Metode Klasifikasi K-Nearest Neighbor (KNN)" *Jurnal CoreIT*, Vol. 1, Issue.2, 2015.
- [12] N. M. Patil, M. U. Nemade, "Music Genre Classification Using MFCC, K-NN, and SVM Classifier" *Computer Engineering in Research Trends*, Vol. 4, Issue 2, p. 43-47, 2017.
- [13] N. Wahyudi, S. Wahyuningsih, F.D.T. Amijaya, "Optimasi Klasifikasi Batubara Berdasarkan Jenis Kalori dengan menggunakan Genetic Modified K-Nearest Neighbor (GMK-NN) (Studi Kasus: PT Jasa Mutu Mineral Indonesia Samarinda, Kalimantan Timur)" *Jurnal EKSPONENSIAL*, 2019.
- [14] S.A. Majeed, H. Husain, S.A. Samad, T.F. Idbeaa, "Mel Frequency Cepstral Coefficients (MFCC) Feature Extraction Enhancement in the Application of Speech Recognition: A Comparison Study. *Journal of Theoretical and Applied Information Technology*" *Journal of Theoretical and Applied Information Technology*, 2015.
- [15] T. H. Simanjuntak, W. F. Mahmudy, Sutrisno, "Implementasi Modified K-Nearest Neighbor Dengan Otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai" *Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 1, No. 2, p. 75-79, 2017.