# Music Recommendation and Classification System Using Acoustic Properties

Engin Deniz ÇAĞLAR, Ahmet YAMAN
Bilgisayar Mühendisliği Bölümü
Yıldız Teknik Üniversitesi, 34220 Istanbul, Türkiye
{l1117080,l1118607}@sdt.yildiz.edu.tr

***Özetçe* —Modern yaşamın gelişmesi ile birlikte teknolojiye olan ihtiyaç her geçen gün artmaktadır. Her alanda olduğu gibi bu teknolojik yeniliklere ayak uydurabilmek de insanlık için büyük önem arz etmektedir. Teknoloji, beraberinde verileri meydana getirmektedir. Bu bilgileri işleyip anlamlı hala getirmek de makine öğrenmesi algoritmalarıyla sağlanabilmektedir.**

**Bu çalışmada makine öğrenmesi yöntemleriyle müzik türü sınıflandırma sistemi üzerinde çalışıldı. Bu kapsamda sayısal işaret işleme teknikleri kullanarak GTZAN veri setindeki ses dosyalarından zaman frekans analizi yöntemlerini kullanıp özellikler çıkartıldı. Bu sürecin ardından sınıflandırma yöntemlerinde kullanıldı. Boyut indirgeme ve öznitelik seçimi yöntemleri kullanılarak sonuçlar karşılaştırıldı.**

**Sonuç olarak geldiğimiz noktada Sinir Ağı algoritması sayesinde %72 başarı oranına ulaşıldı.**

***Anahtar Kelimeler—Machine learning algorithms, music recommendation system, classification system, using acoustic properties, audio, music, signal processing, spectral features, MFCC, CHROMA CQT***

***Abstract*—With the development of modern life, the need for technology is increasing day by day. As in every field, keeping up with these technological innovations is great importance for humanity. Technology creates data with it. Processing this information and making it meaningful can also be achieved with machine learning algorithms.**

**In this study, the music genre classification system is obtained with machine learning methods. In this context, by using digital signal processing techniques, features were extracted from the audio files in the GTZAN dataset using time-frequency analysis methods and used in classification methods. The results were compared using dimension reduction and feature selection methods.**

**As a result, we have reached a success rate of 72% thanks to the Neural Network algorithm.**

## I. Introduction

We can witness technological developments in order to keep up with the new order brought by life. These technological developments, which cover every field, also create smart systems. In today's conditions, automation systems and even artificial intelligence integrated systems are used in many fields. Thanks to artificial intelligence applications, machines make our work easier by analyzing data and drawing meaningful conclusions from them.

In this project, we worked on music genre classification with machine learning applications. Feature extraction was performed from the audio files in the GTZAN dataset using digital signal processing techniques. 11 different machine learning algorithms were trained with these features. Although the results of machine learning algorithms differ, it was aimed to increase the performance results of the algorithms that gave the highest results. PCA and forward selection methods were used to increase performance and simplify features. Thanks to PCA, dimensional reduction operations were performed by compressing the data. With forward selection method, although the number of features was reduced by 90%, results close to the first estimates were obtained. This study was discussed in detail with the feature extraction methods mentioned in Chapter 5. By using feature extraction methods, 276 features of 1000 audio files in the GTZAN dataset were extracted. These features were used for training purposes in machine learning algorithms mentioned in Chapter 7. Derivative, feature selection and dimension reduction methods were used by making improvements in the obtained features. Model performance evaluation metrics were used to test the success of machine learning algorithms. Accuracy, precision, recall and F1 scores were obtained from the values found in the complexity matrix. Cohen's Kappa coefficient was used to test whether the estimates were random. In the 9th chapter, the results are obtained from the given experiments.

As a result, various machine learning algorithms were trained with parameter optimization using feature extraction, size reduction, feature selection methods on GTZAN data set. Although the size reduction and feature selection methods allowed us to reduce the cost and express our dataset with fewer features, they did not have a positive effect on the success of the models. The project was completed by reaching the highest success rate we achieved with the Neural Network algorithm and the dataset obtained after feature extraction, including both derivatives.

## II. Related Works

In this section, literature reviews of previous studies on music sets are included.

In 2002, the article "Musical Genre Classification of Audio Signals" was published by George Tzanetakis. The collected data set was used for music genre classification and this data set was shared with the entire academic community. Although it is the first study with the GTZAN dataset, the performance results of 61% (Non-real-time) and 44% (Real-time) were obtained. [1]

As mentioned in the article "Improved Music Genre Classification with Convolutional Neural Networks", more successful results were obtained by first making changes on the GTZAN dataset. 1000 music files in the data set consist of 30 seconds. Weibin Zhang and his team were able to achieve more successful results by taking a three-second sample of each. Artificial neural networks were used to extract the attributes of the data set, and thus the success rates were 87.4% with max and average-pooling methods. [2]

"Evaluation of Feature Extractors and Psycho-Acoustic As mentioned in the article "Transformations for Music Genre Classification", studies were conducted on 3 different data sets, named GTZAN, ISMIRrhythm and ISMIRgenge. Although a 72.85% success was achieved by using mean values with the features extracted from the GTZAN dataset, this success rate was achieved by using mean values in the ISMIRgenre dataset. It has been increased up to 78.53%. It has been observed again that the selected training set can affect the performance. [3]

As mentioned in the article "Aggregate features and ADABOOST for music classification", feature extraction methods were applied using the MIREX 2005 dataset. With AdaBoost, one of the machine learning algorithms used, the overall average success rate of 82.34% was achieved. [4]

In the article "Music Genre Classification using Machine Learning Techniques", the audio files in the Audio Set dataset were used. By processing 7 different music genres and a total of 40540 sound files, training was provided using the artificial neural network and 89.1% success was achieved by using spectrogram-based models. [5]

## III. Dataset

GTZAN is a dataset containing various music genres collected in 2000-2001. The dataset consists of 1000 audio tracks, each 30 seconds long. It contains 10 species, each represented by 100 pieces. All tracks are 22050 Hz Mono 16-bit audio files in .wav format.

### *A. Music Genres:*

Blues, classical, country, disco, hiphop, jazz, metal, pop, raggae, rock.

### *B. Music Genre Classification Approaches*

There are several methods to classify this dataset. Of these approaches we use: Spectral Centroid, Spectral Bandwith, Spectral Contrast, Spectral Flatness, Spectral Rolloff, Zero Crossing Rate, Polynomial Features, Root Mean Square Energy, Mel-Frequency Cepstral Coefficients, Chroma Constant-Q Transform.

## IV. Feature Extraction

### *A. Spectral Centroid*

Indicates where the center of mass of the spectrum is. It has a strong relationship with the brightness of the sound. Spectral Centroid is calculated by the weighted mean of the frequencies present in the signal. [6]

$$\frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$$

x(n) is the weighted frequency value or magnitude, n is the frequency range we are in, and f(n) is the center frequency in the range.

### *B. Spectral Bandwidth*

It defines half the maximum peak of the wave width of the audio signal. [7]

$$(\sum_k S(k)(f(k)-f_c)^p)^{\frac{1}{p}}$$

### *C. Spectral Contrast*

Spectral contrast is expressed as the difference in magnitude between the peaks and valleys in the spectrum. [8]

In a study conducted by Tsinghua University, it was seen that the spectral contrast feature provides superior success in classification of music genres compared to MFCC features. [9]

### *D. Spectral Flatness*

Spectral flatness (tonality coefficient) is a measure used to measure how similar a sound is to noise as opposed to being tone-like. A high spectral flatness (close to 1.0) indicates that the spectrum is similar to white noise. It is usually converted to decibels. [10] It is calculated by dividing the geometric mean of the power spectrogram by the arithmetic mean of the power spectrogram.

### *E. Spectral Rolloff*

Spectral rolloff is the frequency below which a specified percentage of the total spectral energy. The frequency after the percentile is reached is considered as the Rolloff Frequency. [11]

### *F. Zero Crossing Rate*

Zero crossing rate is the rate at which a signal crosses the zero line, that is, the rate of sign change. It is used for speech recognition and retrieval of music information. Makes it easy to detect percussion instruments. It is easy to calculate and can be used for noise detection. [12]

### *G. Polynomial Features*

In simple and multiple linear regression, the relationship between dependent variable and independent variables was linear. However, in real life, the relationship between variables may not be linear in some situations. In such cases, Polynomial Regression can be used to model the relationship. With the help of polynomial features, coefficients of fitting an nth-order polynomial to the columns of a spectrogram are obtained. [13]

### H. Root Mean Square Energy

The energy of a signal corresponds to the total magnitude of the signal. This magnitude is not enough for us to understand how loud the signal is. For audio signals, the RMSE helps us calculate how loud the signal is. [14]

$$RMSE = \sqrt{\frac{1}{N}\sum_{n}|x(n)|^2}$$

### I. Mel-Frequency Cepstral Coefficients

The Mel frequency scale is a scale that shows how the human ear perceives the change in sound frequencies. MFCC (Mel-Frequency Cepstral Coefficients) is the expression of the short-time power spectrum of the audio signal on the Mel scale. [15]

Frequency to Mel scale conversion formula;

$M = 1125 \times \ln(1 + (f \div 700))$

$M \rightarrow$ Mel Scale

$f \rightarrow$ Frequency (Hz)

### J. Chroma Constant-Q Transform

Unlike the Fourier transform, the Chroma Constant-Q Transform uses frequencies divided into logarithmic intervals like the Mel Scale. Chroma CQT, on the other hand, shows the energy of the 12 pitches in western music throughout the signal as a vector with 12 features. [16]

## V. Size Reduction and Forward Selection Methods

### A. Principal Component Analysis (PCA)

Principal component analysis is a kind of technique used for dimensional reduction that allows us to define the data set with reduced and tranfsormed data. Its main purpose is to keep the data set with the highest variance, while aiming to reduce size. It provides the reduction of the number of dimensions, by finding the most import features among the data in big data sets. Although reducing the features may cause data loss, it can be quite efficient according to the time and processing capability gained.

On the basis of PCA, it compresses multidimensional data into low-variable data with basic properties.

### B. Forward Selection

The abundance of features (variables) in the data set brings us various difficulties. Excess data requires having powerful hardware. There are various ways to avoid this hardware cost and shorten the time spent in the training process by using clever approaches in this process. If we can selectively reduce the number of data we have, we can avoid these costs.

*1) The p-value:* The p-value, which is used in machine learning algorithms as it is used in most of the statistical modeling, represents the levels of significance. This value is modeled between 0 and 1. The level of significance increases as it approaches 0. P-values less than 0.05 can be used for interpretation. The forward selection method starts with the selection of the variable that will contribute the most to the model. The lower the p-value values used here, the more significant it means.

## VI. Machine Learning Algorithms

### A. Logistic Regression

Logistic Regression is a model for establishing a linear relationship when the number of variables is two or more. Two logical results are produced, true (1) and false (0).

### B. K-Nearest Neighbors

K-NN is a hypothetical progressive, lazy learning algorithm. It memorizes the dataset as opposed to learning the training data. Therefore, when an estimate is desired, it aims to classify according to the close relationship with the previous data.

In the operation of the algorithm, the K value is selected for the number of elements. According to the value of K, the nearest K amount of elements is taken and the distance between the incoming value is calculated using the Euclidean function. Manhattan, Minkowski and Hamming functions can also be used as an alternative to the Euclidean function.

### C. Support Vector Machine

The support vector machine is a classification algorithm that works in logic similar to the linear regression algorithm. Both try to find the best parser that separates the two classes. The SVM algorithm ensures that the two separated classes are set to pass from the farthest place to their elements. It does not need to take parameters. SVM can also be used to classify linear and nonlinear data.

*1) Kernel Trick:* SVM tries to classify data in linear logic, but in some cases it is impossible to do. To get rid of this situation, we use to the kernel trick. If we can create a different dimension, it may be possible to classify it as linear.

### D. Naive Bayes

Naive Bayes classifier is based on Bayes theorem. It is a lazy learning algorithm. Calculates the probability of each state for an element and classifies it according to the highest probability value.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times \ldots \times P(x_n|c) \times P(c)$$

### E. Decision Tree

The decision tree method is a machine learning algorithm used in both classification and regression problems. Decision trees, which are also frequently used in the field of data mining, are generally at a level that can be considered at the human level. Making sense and visualizing data can be easily achieved through decision trees.

The decision tree is a recursively operation. A tree structure is created by starting with a single node and branching out new results.

### F. Random Forest

The Random Forest algorithm can be used in both classification and regression problems, such as a decision tree. It is obtained by creating multiple decision trees instead of creating a single decision tree. If any result is desired, it is found by averaging the decision trees.

### G. Quadratic Discriminant

Quadratic Classifier is a quadratic classification method used to separate measurements of two or more objects and event class. It is a more comprehensive version of the linear classifier.

### H. Linear Discriminant

It is a linear discriminant dimension reduction method. It aims to classify data by finding a linear combination of features. It reduces higher-dimensional components to lower dimensions by modeling differences in groups.

### I. Gaussian Process Regression

Gaussian process regression (GPR) is a non-parametric, Bayesian regression approach. GPR works pretty well on small datasets. It also has the ability to provide measures of uncertainty over estimates.

### J. Neural Network

Artificial neural network is a technique developed by using the human brain in information processing technique. It mimics the working logic of human nervous systems. Learning takes place by using examples with the logic of training. It occurs when the processed data is tried repeatedly until a convergence is found thanks to the connection weights.

Like neurons, the basic unit of the brain, the basic building block of an artificial neural network is a sensor that performs simple signal processing.

The computer with the neural network is taught to perform a task by having the pre-labeled training samples analyzed. Learns to categorize new data by analyzing repetitive patterns in the presented data.

### K. AdaBoost

Boosting algorithms try to obtain a strong learner by combining the weak learners obtained in each iteration.

If the correlation between the two variables is low, the prediction rate is also low. Conversely, the higher the correlation, the higher the prediction rate.

## VII. Model Performance Evaluation Methods

After training a machine learning algorithm, we need to find out how well the model fits our data. There are different model performance evaluation metrics for different machine learning algorithms.

In this project, accuracy, precision, recall and f1 score values were obtained from the values we found by creating a confusion matrix. Cohen's Kappa coefficient was used to test whether the estimates were random.

### A. Confusion Matrix

A confusion matrix is a table used to describe the performance of a classification model on a set of test data for which the actual values are known. The success of the model is related to the quantities of data that are classified as true and false.

- True Positive : The number of cases where the prediction is positive and the actual result is positive.
- False Positive : The number of cases where the prediction is positive and the actual result is negative.
- False Negative : The number of cases where the prediction is negative and the actual result is positive.
- True Negative : The number of cases where the prediction is negative and the actual result is negative.

### B. Accuracy

The accuracy value is found by the ratio of the areas that we classified correctly in the model to the total data set. It was used in conjunction with other metrics because it did not make enough sense on its own.

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

### C. Precision

Precision, on the other hand, calculates how many of the values we predicted as positive are actually positive.

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

### D. Recall

It is the ratio of the number of correctly classified positive samples to the total number of positive samples. [17]

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

### E. F1 Score

The F1 score is the harmonic mean of the precision and recall values and uses the harmonic mean to deal with extreme situations. The main reason for using the F1 score is to make the right model selection in data sets that are not evenly distributed.

$$F_1 = 2 * \frac{precision * recall}{precision + recall} \tag{4}$$

### F. Cohen's Kappa

Cohen's Kappa measures the reliability of agreement between two or more raters. It is a type of non-parametric statistics because the obtained variable is a categorical variable.

When calculating the Kappa coefficient, two different variables are used, Pr(a) and Pr(e). Pr(a) is the sum ratio of the observed agreement for the two raters, and Pr(e) is the probability of this agreement occurring by chance. [18] The formula used to find Cohen's Kappa coefficient is:

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}, \tag{5}$$

| $\kappa$ | Yorum |
|---|---|
| <0 | There's no match. |
| 0.0 - 0.20 | There is insignificant match. |
| 0.21 - 0.40 | There is moderate match. |
| 0.41 - 0.60 | There is a common match. |
| 0.61 - 0.80 | There is significant match. |
| 0.81 - 1.00 | Almost perfect match. |

**Table 1** Cohen's Kappa

## VIII. Experimental Results

In this section, we will see the changes we made to our dataset so that we can compare our results after training machine learning algorithms.

In order to obtain the most successful result in machine learning algorithms, Grid-Search and K-Fold Cross Validation methods were used. Thanks to Grid Search, the parameters with the highest scores in the models were reached. With K-Fold Cross Validation, the entire data set was used as test and training data, to prevent overfitting.

### A. Processes With Dataset

The datasets of extracted features, contain the following attributes:

Spectral Centroid, Spectral Bandwidth, Spectral Contrast, Spectral Flatness, Spectral Rolloff, Zero Crossing Rate, Root Mean Square Energy, MFCC, Chroma CQT, Poly Features

In order to increase model success and test which dataset gives the best results, we created 3 different datasets by taking the 1st and 2nd derivatives of the features:

1) Extracted Attributes (267 Attributes)
2) Extracted Attributes + First Derivative of Attributes (533 Attributes)
3) Extracted Attributes + First Derivative of Attributes + Second Derivative of Attributes (799 Attributes)

We used the mean, standard deviation, median, mod, minimum, maximum, kurtosis and skewness values of the features in all datasets to train the machine learning algorithms.
The numbers from 0 to 9, in the confusion matrices used in the experimental results are respectively refers to the genres of blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae and rock.

### B. Results After Feature Extraction

In the first experiment, no dimension reduction or feature selection was applied on all 3 data sets. The datasets were given to machine learning algorithms with their status after feature extraction. Parameter-optimized 5-fold Cross Validation scores of all three sets are given in the table

| Algrthm | 1. Dataset | 2.Dataset | 3. Dataset |
|---|---|---|---|
| A | 0,567 | 0,586 | 0,589 |
| B | 0,676 | 0,692 | 0,7 |
| C | 0,584 | 0,58 | 0,591 |
| D | 0,456 | 0,46 | 0,476 |
| E | 0,56 | 0,599 | 0,611 |
| F | 0,662 | 0,697 | 0,721 |
| G | 0,276 | 0,206 | 0,205 |
| H | 0,491 | 0,546 | 0,545 |
| I | 0,146 | 0,145 | 0,139 |
| J | 0,661 | 0,69 | 0,704 |
| K | 0,64 | 0,559 | 0,18 |

**Table 2** Best Scores

*1) Best Results:* The accuracy metrics of the models that gave the best results from each dataset in machine learning algorithms were compared.

Among the comparisons, the best result 0.72 was obtained with the Neural Net algorithm using the third dataset.

### C. Effects of Features On Model Scores

The features obtained in this section were given to the SVM and Neural Net algorithms, respectively, without using their derivatives, and their accuracy scores were compared. The effects of the features we used on the classification were determined. Before comparing, we divided the attributes into 4 different classes:

- Spectral Features
- MFCC
- Chroma CQT
- Polynomial Features

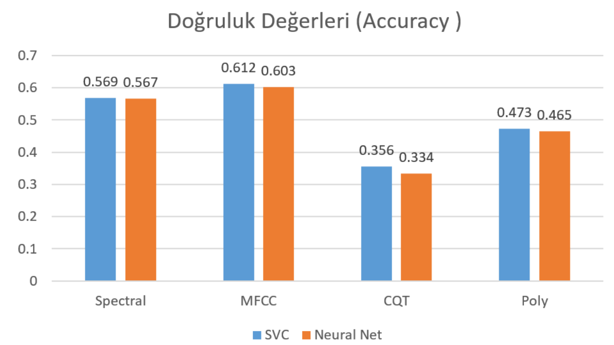


**Figure 1** Feature- Accuracy Graph

As a result of the comparisons, it was seen that the Spectral and MFCC features performed remarkably well in determining the music genre.

*D. Results with Principal Component Analysis*

Then Principal Component Analysis was applied to the datasets and the results from the machine learning algorithms were compared.

*1) Algorithm Results:* Size-reduced datasets were adjusted to represent 95% of the distribution in older datasets.

- When we apply PCA to the first dataset, we can express 95% of the distribution with 75 features.
- When we apply PCA to the second dataset, we can express 95% of the distribution with 155 features.
- When we apply PCA to the third dataset, we can express 95% of the distribution with 216 features.

The Linear Discriminant model performed better on all datasets compared to other models. Principal Component Analysis's use of linear methods for dimension reduction and feature selection has increased the success of linear classification models.

The accuracy results and the parameters of the best scores obtained from the PCA-applied datasets are given in the following Table 3.

*2) Best Results:* The accuracy metrics of the best-performing models were compared with the PCA-applied datasets.

Among the comparisons, the best result 0.68 was obtained with the Linear Discriminant algorithm using the second dataset.

| Algrthm | 1. Dataset | 2. Dataset | 3. Dataset |
|---|---|---|---|
| A | 0,455 | 0,326 | 0,286 |
| B | 0,635 | 0,634 | 0,605 |
| C | 0,337 | 0,155 | 0,143 |
| D | 0,384 | 0,415 | 0,408 |
| E | 0,602 | 0,588 | 0,579 |
| F | 0,605 | 0,622 | 0,616 |
| G | 0,275 | 0,264 | 0,172 |
| H | 0,525 | 0,54 | 0,527 |
| I | 0,316 | 0,168 | 0,165 |
| J | 0,617 | 0,616 | 0,607 |
| K | 0,656 | 0,682 | 0,681 |

**Table 3** PCA Best Scores

*E. Results with Forward Selection*

The fact that there were a lot of variables in our project was both time cost and strong transaction cost for us. We tried to find the most necessary ones among dozens of variables and obtain close estimation results. In this way;

*1) Algorithm Results:* In the dataset with Forward Selection applied, the p value was set to be less than 0.05.

| Algrthm | 1. Dataset | 2. Dataset | 3. Dataset |
|---|---|---|---|
| A | 0,542 | 0,558 | 0,579 |
| B | 0,603 | 0,659 | 0,69 |
| C | 0,533 | 0,591 | 0,594 |
| D | 0,417 | 0,44 | 0,428 |
| E | 0,569 | 0,581 | 0,585 |
| F | 0,598 | 0,649 | 0,684 |
| G | 0,25 | 0,223 | 0,235 |
| H | 0,453 | 0,484 | 0,505 |
| I | 0,456 | 0,445 | 0,438 |
| J | 0,694 | 0,642 | 0,691 |
| K | 0,569 | 0,661 | 0,669 |

**Table 4** Forward Selection Best Scores

- When we apply forward selection to the first dataset, we can reduce the number of features from 267 to 27.
- When we apply forward selection to the second dataset, we can reduce the number of features from 533 to 59.
- When we apply forward selection to the first dataset, we can reduce the number of features from 799 to 56.

Thanks to the forward selection method, we observed an increase in success in Gaussian process, AdaBoost, QDA and Linear Discriminant algorithms. In other algorithms, we obtained values close to the results in our full data set. According to the results of our experiment, we observed that the elimination methods can sometimes produce more successful results. In addition to these, we were able to get close results despite the 90% reduction in the features in the data set. This has saved us a lot of time.

Obtaining results close to PCA results once again showed us that we are on the right track.

## IX. Conclusion

In this project, we aimed to train machine learning algorithms with the music data set we have and to find the algorithm that can classify a music according to ten different classes in the most successful way.

Initially, we took the audio files in the GTZAN dataset and extracted the attributes that would help us classify the audio data such as spectral bandwidth, MFCC, Chroma CQT. We also kept the derivatives of the extracted features in separate datasets to examine their effects on the success of the models. To train our models, we transformed the features in the audio data into more meaningful data with mathematical functions such as mod, median, and standard deviation.

After creating our datasets, we trained 11 different machine learning algorithms with our data. We used Grid Search to optimize the parameters taken by the algorithms. In this way, we found the parameters that gave us the highest score. By using Cross Validation instead of dividing our data into test and training data, we avoided the problem of overfitting when calculating the score.

We implemented Principal Component Analysis method because it is costly to train multiple machine learning algorithms with parameter optimization. Although the algorithms started to learn faster with the decrease in the number of features, this situation did not provide a significant increase in accuracy scores. Since Principal Component Analysis uses a linear method for dimension reduction, the success rates of the Linear Discriminant machine learning algorithm have increased. We achieved results close to previous success values by reducing our feature count by almost 75%.

Then, using the forward feature selection method, it was aimed to get rid of unnecessary features by choosing the features with the highest meaning on the data. The size of the datasets decreased by 90% after feature selection. The scores obtained from the feature selection method were close to scores at PCA method.

As a result, datasets using first and second derivatives, achieved higher performance, regardless of the use of dimension reduction and feature selection methods. No increase in success was observed in experiments where we used forward feature selection. The dataset with the first and second derivatives in the first experiment, which were not changed by size reduction or feature selection methods after feature extraction, achieved the highest success with 72% using the Neural Network algorithm. We will continue to work to increase the success of feature selection.

## References


[1] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on speech and audio processing*, vol. 10, no. 5, pp. 293–302, 2002.

[2] W. Zhang, W. Lei, X. Xu, and X. Xing, "Improved music genre classification with convolutional neural networks." in *Interspeech*, 2016, pp. 3304–3308.

[3] T. Lidy and A. Rauber, "Evaluation of feature extractors and psycho-acoustic transformations for music genre classification," in *ISMIR*, 2005, pp. 34–41.

[4] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. Kégl, "Aggregate features and a da b oost for music classification," *Machine learning*, vol. 65, no. 2-3, pp. 473–484, 2006.

[5] H. Bahuleyan, "Music genre classification using machine learning techniques," *arXiv preprint arXiv:1804.01149*, 2018.

[6] spectralcentroid. [Online]. Available: https://librosa.org/doc/0.7.2/generated/librosa.feature.spectral$_centroid.html$

[7] spectralbandwidth. [Online]. Available: https://librosa.org/doc/0.7.2/generated/librosa.feature.spectral$_bandwidth.html$

[8] D.-N. Jiang, L. Lu, H.-J. Zhang, J.-H. Tao, and L.-H. Cai, "Music type classification by spectral contrast feature," in *Proceedings. IEEE International Conference on Multimedia and Expo*, vol. 1. IEEE, 2002, pp. 113–116.

[9] J. Yang, F.-L. Luo, and A. Nehorai, "Spectral contrast enhancement: Algorithms and comparisons," *Speech Communication*, vol. 39, no. 1-2, pp. 33–46, 2003.

[10] S. Dubnov, "Generalization of spectral flatness measure for non-gaussian linear processes," *IEEE Signal Processing Letters*, vol. 11, no. 8, pp. 698–701, 2004.

[11] "spectralrolloff." [Online]. Available: https://librosa.org/doc/0.7.2/generated/librosa.feature.spectral$_rolloff.html$

[12] "zerocrossingrate." [Online]. Available: https://librosa.org/doc/0.7.2/generated/librosa.feature.zero$_crossing_rate.html$

[13] "polyfeatures." [Online]. Available: https://librosa.org/doc/0.7.2/generated/librosa.feature.poly$_features.html$

[14] "rmse." [Online]. Available: https://librosa.org/doc/0.7.2/generated/librosa.feature.rm

[15] M. Xu, L.-Y. Duan, J. Cai, L.-T. Chia, C. Xu, and Q. Tian, "Hmm-based audio keyword generation," in *Pacific-Rim Conference on Multimedia*. Springer, 2004, pp. 566–574.

[16] "chromacqt." [Online]. Available: https://librosa.org/doc/0.7.2/generated/librosa.feature.chroma$_cqt.html$

[17] C. Coşkun and A. Baykal, "Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması," *Akademik Bilişim*, vol. 2011, pp. 1–8, 2011.

[18] S. Kiliç, "Kappa test," *Psychiatry and Behavioral Sciences*, vol. 5, no. 3, p. 142, 2015.