Music Genre Classification Using Machine Learning Algorithms

Zekiye Doğan
2017510025

Şükran Özbek
2017510085

***Abstract*—This document is about algorithms made and developed to find the most suitable model on the Music Genre Classification data set. The results and analyzes of these algorithms are also included in this document.**

***Index Terms*—machine learning, music genre, classification, random forest**

INTRODUCTION

The aim of this project is to classify music according to its genres by performing operations on the selected data set. First, the dataset was examined in detail. The distribution of the features, the relations between each other and the target were analyzed. The presence of missing value, outlier and duplicate data in the dataset was checked and its effects on it were investigated. In the continuation of these processes, it was tried to obtain features that could be more effective on the model from the existing features. Classification models were used to classify the data. And since the target value is included in the train dataset, it was supervised. Target gets the value of the music genre.

RELATED WORK

A. Music Genre Classification and Recommendation by Using Machine Learning Techniques [1]

In this study, music genre classification and recommendation are made by determining the acoustic properties of music and using machine learning techniques. By making a comparison between the models used, it was determined which one worked better on the data set and related analyzes were made. According to the study, the highest success worked with the SVM algorithm. Study has been conducted over two steps; determining how features that will be used in recommendation are obtained and developing a service that recommends songs to user requests. In the first part, models were created by extracting features and used in classification. Using deep learning method showed that there is no considerable performance chance on music genre classification. Furthermore, SVM has achieved higher success than the CNN algorithm. As for music recommendations, because there is no objective metric about music recommendation, music recommendation by genre can be used a solution this challenge. While it gives high results in some species, it has been observed that these performances decrease in others.

B. Machine Learning Algorithms -A Review [2]

In this article, a brief review and future prospect of the vast applications of machine learning algorithms has been made. These are also divided into supervised learning, unsupervised learning and semi-supervised algorithms.

As a result Machine Learning can be Supervised or Unsupervised. If we have less amount of data and clearly labeled data for training, it will be better to choose Supervised Learning. Unsupervised Learning generally gives better performance and results for large datasets. If you have a large dataset that is easily accessible, it would be better to look into deep learning techniques.

C. A Hybrid Model For Music Genre Classification Using LSTM And SVM [3]

Machine learning models are hugely effective in predicting music genre classification. Even today, many well-known companies (Spotify, Youtube) use them. In order to take advantage of the time dependent nature of the dataset used in this article, the Long Short Term Memory (LSTM) Neural Network is used for music genre classification and combined with the Support Vector Machine (SVM) classifier to improve its performance. This article, which considers a hybrid model of these two classifiers, states that the individual models achieve an increase in prediction accuracy. This hybrid model is applied to the GTZAN music dataset and compared with the results of the independent LSTM and SVM models. The proposed model exceeded the independent accuracies of LSTM and SVM classifiers with 89 percent accuracy and summarizes the efficient use of each classifier.

D. Music Genre Classification using Machine Learning Algorithms: A comparison [4]

The aim of the project and research in this article is to find a better machine learning algorithm than pre-existing models that predict the genre of songs. In this project, multiple classification models were created and trained on the Free Music Archive (FMA) dataset. The performances of all these models were compared and their results were analyzed in terms of their predictive accuracy.

A few of these models were trained on mel spectrograms of songs along with their vocal features, and a few others were trained on spectrograms of songs only. It was found that one of the models, a convolutional neural network for which only

the spectrograms were given as the dataset, gave the highest accuracy among all the other models.

I. Material and Methods

The dataset is suitable for supervised classification. It also has a multi-label target value. Among the appropriate algorithms, Neural Network, RandomForest, Support Vector Machine and Logistic Regression were selected to use algorithms.

A. Why RandomForest?

Random Forest is a supervised learning algorithm. One of the important advantages of the algorithm is that it can be used in both classification and regression problems. It can be used to identify the most important feature among the features available in the training dataset. It creates random forests and combines them for a more stable and accurate prediction. When splitting a node, instead of looking for the most important feature, it searches for the best feature among a random subset of features. This results in a wide variety, which often results in a better model. It may be possible for the models to be developed to memorize the train and to have over-fitting. As a solution to this situation, it was decided to use the RandomForest algorithm. Random Forest Classifiers facilitate the reduction in the over-fitting of the model and these classifiers are more accurate than the decision trees in several cases.

B. Why Logistic Regression?

The target value of the dataset takes a value of 1(energic music) and 0 on the generated target feature. With this in mind, we chose to use Logistic Regression.Logistic Regression is useful for understanding the effect of independent variables on a single outcome variable. Logistic Regression is a regression method for classification. It is used to classify categorical or numerical data. It works if the dependent variable, namely the result, can only take 2 different values. It thinks that predictors are independent and there is no missing data. It is useful for understanding the effect of independent variables on a single outcome variable.

C. Why Neural Network?

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature. Neural networks can adapt to changing input; so the network generates the best possible result without needing to redesign the output criteria. The concept of neural networks, which has its roots in artificial intelligence, is swiftly gaining popularity in the development of trading systems.

An artificial neural network behaves the same way. It works on three layers. The input layer takes input. The hidden layer processes the input. Finally, the output layer sends the calculated output.

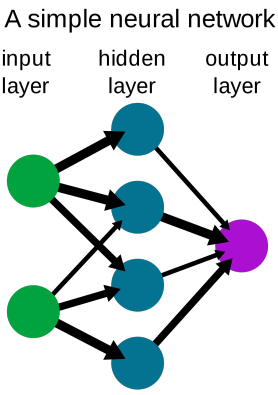

Fig. 1. Simple neural network

D. Why Support Vector Machine?

Another most widely used state-of-the-art machine learning technique is Support Vector Machine (SVM). In machine learning, support-vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. It basically, draw margins between the classes. The margins are drawn in such a fashion that the distance between the margin and the classes is maximum and hence, minimizing the classification error. Since our dataset has a multi-label class value, we chose and applied SVM.

Dataset Information

The dataset used is the “Music Genre Classification” dataset from MachineHack Hackathon. There are both Train and Test datasets. The Train dataset has a total of 17 features and 17,996 examples with target. The test dataset has a total of 7,713 samples with 16 features and target. Some example features: Popularity, Danceability, Energy, Acousticness, Instrumentalness.

Then, a detailed analysis of the data was started. The review was initially started manually. In this analysis, it was examined which data types the data had (categoric or numeric) and it was observed in which range the values were. The meanings of these values were examined.

As a result of manual trials, it was tried to predict which features would be more effective. During manual testing, two instances were randomly selected and a comparison was made between these features.

When viewed from a statistical view, Loudness, Instrumentalness ve Tempo values were fixed. There were features ArtistName and TrackName that were considered ineffective and so they were deleted from the dataset as they would not have any effect on future predictions.

Numerical target values have been converted into categorical data. The aim here is to better understand the relations of the

target feature with other features, and the map process has been applied. If the values remained numeric, it could have a meaning that one music genre is greater than another. This is an undesirable misconception.

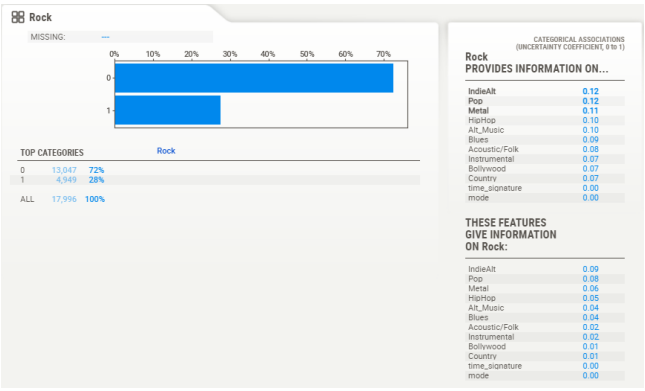

Fig. 2. Rock music genre distribution

Let's take a look at the Target feature review:
Music Genre

- 0. ' Acoustic/Folk '
- 1. ' Alt Music '
- 2. ' Blues '
- 3. ' Bollywood '
- 4. ' Country '
- 5. ' HipHop '
- 6. ' IndieAlt'
- 7. ' Instrumental '
- 8. ' Metal '
- 9. ' Pop '
- 10. ' Rock '

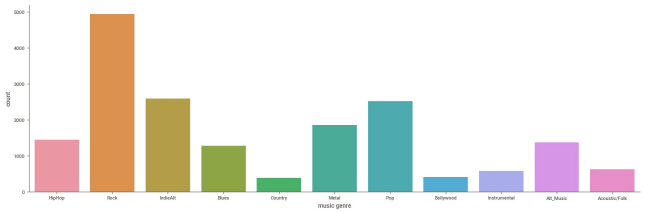

Fig. 3. Music Genre Distribution

The minimum, maximum, mean and standard deviation values of each feature were examined by making a data quality report. One of the reasons for doing this is to determine the features that may contain outliers by looking at which values the data are in, mean and standard deviation.

Then, the analysis report was obtained through the Spyder tool. In this analysis report, the features of each feature were examined in detail and the relationship of these features with other features was observed statistically. The most important features when examining are the correlation relationship of the features with the target (Music Genre), their distribution and how effective they are on the dataset. Thus, an idea was gained on both obtaining information about the feature and manipulating that feature.

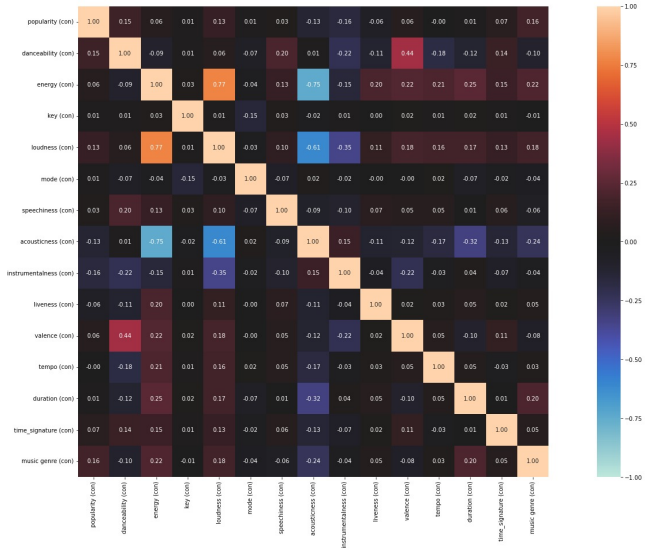

Fig. 4. Correlation Matrix before Data Preprocessing

Then the correlation matrix of the dataset was obtained. This matrix contains both categorical and numerical features. The information became more understandable thanks to the graphic that summarizes the relationships between them and helps visually. Looking at this, it was seen that the 3 features that affected Music Genre the most were Duration(0.70), Acounticness(0.59), Energy (0.56).

First of all, some pre-processes should be applied to the dataset in order to be able to operate on the dataset. After the correlation analysis, it was checked whether there was a missing value in the dataset. Missing values were found in Popularity, Key and Instrumentalness properties. These values are arranged by replacing the median values. Then, it was checked whether there were duplicates on the data and the values were corrected by deleting them.

Range normalization applied in Instrumentalness, Tempo, Loudness and Duration. This is because these properties have higher values than other properties. Because it can create pressure on other data and impair the accuracy of the result to be obtained from the model. The reason for doing this is that it helps us to understand the scattered values and see better results.

Then the boxplot was used to check for an outlier in the data set. Outlier detected in some properties. Finally, outlier analysis was performed and outlier values were arranged with IQR. For discrete features, outliers were determined with the IQR (Inter Quartile Range) method and outlier values that were less than the minimum thresold value were set to the minimum value, and those that were greater than the maximum were set to the maximum value.

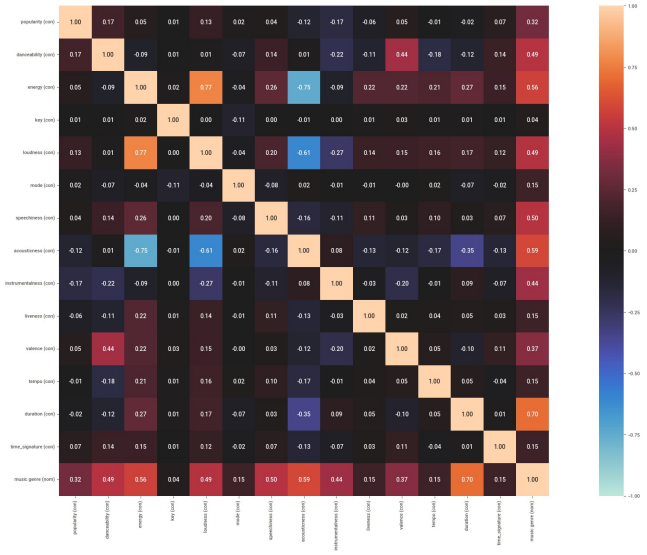

Fig. 5. Correlation Matrix before Data Preprocessing

EXPERIMENTAL RESULTS

E. Part -1

The dataset used is suitable for multi-class classification. Research has been done on the models that can be applied on the dataset and it has been decided on four different models. These models are Random Forest, Neural Network, Logistic Regression and Support Vector Machine. The success measurement of the models was made via Spyder. The dataset was cleaned and made ready for use on the model. After the target feature of the dataset is determined, it is removed from the dataset. Afterwards, the data set was divided into 70Random Forest model has many parameters. After many trials, n-estimators value was determined as '1000' and criterion value was determined as 'entropy' in order to get the best efficiency from the model. Neural Network is one of the models that gives the best results. The model has been trained many times, and with parameter values (max-iter = 500, solver="adam",shuffle = True), the model gives the best results. The Logistic Regression model has a structure that supports multi-class and returns the best result for this data with parameters (multi-class='multinomial', solver='sag',max-iter=1500,C=100). Support Vector Machine supports multi-class, similar to Logistic Regression. It gives the best results with (C=10, kernel = "linear") parameters. OnevsOneClassifier is another way to do multi-class classification. OnevsOneClassifier tries to find the best result by comparing each label of the target with each other. This structure is referenced in the Support Vector Machine model.

F. Part -2

After applying the processes mentioned in part 1, we decided that the results were insufficient and that every improvement made could not exceed a certain threshold, and we started to do part 2.

One of the multi-label classification solutions was to convert the target feature to binary, so we started part2. As a result of our dataset and our research, we divided our target into 2 values from the energic music genre. Of these, the energic music we determined corresponds to the value of 1.

We made this decision as follows, first we integrated our music genre class into the correlation matrix and looked at their relationships with all the features.

We've looked at it from many angles, and "Rock, Metal, Instrumentalness, Acoustic-Folk, Bollywood, Blues" has formed the energic music genre. These are also based on the features of "energy, loudness, instrumentalness, acoustic". After that, the models valid in part1 were applied. The Random Forest model has many parameters.

After many trials, n-estimators value was determined as '500' and criterion value 'entropy' in order to get the best efficiency from the model.

Neural Network is one of the models that gives the best results. The model has been trained many times and with parameter values(max-iter = 400), the model gives the best results.

The Logistic Regression model has a structure that supports multiple classes and returns the best result for these data with parameters (solver='saga').

The Support Vector Machine supports multiple classes, similar to Logistic Regression. It gives the best results with default parameters.

	Part 1				Part 2			
	R. F	N. N	L. R	S.V.M	R. F	N. N	L. R	S.V.M
accuracy	0.5147	0.5264	0.5019	0.5032	0.7783	0.7479	0.6842	0.6836
precision	0.58	0.54	0.51	0.53	0.78	0.75	0.69	0.69
recall	0.56	0.58	0.54	0.53	0.78	0.75	0.68	0.67
f1-score	0.57	0.55	0.52	0.52	0.78	0.75	0.68	0.67
MAE	2.1032	1.9790	2.1093	2.0933	0.2216	0.2520	0.3157	0.31634
MSE	12.2821	11.3809	12.3740	12.1772	0.2216	0.2520	0.3157	0.31634
RMSE	3.5045	3.3735	3.5176	3.4895	0.4708	0.5020	0.5619	0.5624
R^2	-0.1879	-0.1007	-0.1968	-0.1778	0.1089	-0.0130	-0.2693	-0.2715

Fig. 6. Results

Looking at the result table, it has been decided that the best model for multi-class classification is Random Forest, even though its accuracy is lower than Neural Network, considering f1-score and precision values.

CONCLUSION

REFERENCES

[1] ELBIR, Ahmet, et al. Music genre classification and recommendation by using machine learning techniques. In: 2018 Innovations in Intelligent Systems and Applications Conference (ASYU). IEEE, 2018. p. 1-5.

[2] MAHESH, Batta. Machine Learning Algorithms-A Review. International Journal of Science and Research (IJSR).[Internet], 2020, 9: 381-386.

[3] FULZELE, Prasenjeet, et al. A hybrid model for music genre classification using LSTM and SVM. In: 2018 Eleventh International Conference on Contemporary Computing (IC3). IEEE, 2018. p. 1-3.

[4] CHILLARA, Snigdha, et al. Music genre classification using machine learning algorithms: a comparison. Int Res J Eng Technol, 2019, 6.5: 851-858.