A Data Mining apporach for identifying song taste: A complete predictive model

Imtiajul Islam[†], Md. Alvi Tazwar[‡],Faria Tabassum[ǂ],
[†‡*]Department of Computer Science and Engineering, East West University, Dhaka-1212, Bangladesh
Email: [†]iimtiajul@gmail.com, [‡]atd.mondol@gmail.com, [ǂ]faria.yera@gmail.com,

Abstract—music is more than a sound to ear. People all across the world of different cultures, languages, and lifestyles love music. It causes dopamine to be released from the brain that's why people love music so much. It can be so diverse, so meaningful, and such a great escape from the real world. In this paper we present a model that predicts the taste of songs of human based on the attribute of the song. If attribute is provided then it gives the prediction if the human like the song or not.

Keywords—
clustering,prediction,attribute,song,taste,target,algorithms etc.

I. INTRODUCTION

Machine learning is one of crucial part of any analysis. With the help of any mining process, we can come into any conclusion which will help us to predict any result. To reach a decision about anything we need process of mining as it will help to predict the futuristic outcome. Music is one of our core sources of entertainment. There are lot of scope of determining any possible outcome in music industry. As there are lot of music genre and types in the whole world so music companies always keep an eye which type of music is listening by the music lovers most in the days, To come to a solution a large number of dataset is needed and with the help of several algorithms we can come into solution that whether a music can be loved by the listeners or not.

II. RELATED WORK

Research on predictive web perfecting has involved the important issue of log file processing and the determination of user transactions (sessions) from it3. Several approaches have been proposed towards this direction. The paper [1] presents an approach to classifying students in order to predict their final grade based on features extracted from logged data in an education web-based system. They design, implement, and evaluate a series of pattern classifiers and compare their performance on an online course dataset [2].

III. DATA PREPARATION

The data set used for the prediction is fetch from famous music listening application Spotify. All the songs are fetched from the 2017 list. In the dataset there are 16 columns and 13 are attributes. To prepare the data we used an open source library called pandas which is a powerful tool of python. In pandas, we import the data from CSV file and the tool easily convert and prepare the data for testing and make the date useful by converting data in tabular form of matrix. The general process of importing is,

```
import numpy as np
spotify  = pd.read_csv('Spotifydata.csv')
```

Another library is used for numpy which is another useful tool in python. It is mainly used for computing in scientific calculation and also used in analysis of N –number of object of array. The general syntax is given below,

```
. import numpy as np.
type_level = np.where(spotify['target']=1, 0, 1)
```

IV. OBJECTIVES AND GOALS

- Identify proper dataset for finding song taste.
- Cleaning the dataset.
- Applying the algorithms and test the dataset.
- Finding pattern of which algorithm is working effectively.

V. DATA PRE-PROCESSING

To prepare a data for testing there are some pre-processing work is needed to prepare the data. As all the dataset are not clean and there may be ambiguous information that may hamper the whole work. The dataset may be inconsistent and dirty so before preparing it for we need to clean the data by reducing deducting and recoiling the noises.

VI. MATERIALS AND METHODS

In the present paper, to come into a solution, there are several algorithms we use like decision tree, naïve bays a priori and svm. Decision tree learning uses a decision tree (as a predictive model) to go from observations about an item to conclusions about the item's target value. A tree structure is maintained to identify the results. It is one of the predictive techniques used in statistics, and machine learning. Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.[3] Naïve Bayes classification technique is based on Bayes' Theorem with an assumption of independence

among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naive'.[4]

VII. Results and Analysis

This section contains the result and analysis of the tested dataset under certain algorithms. At first we use decision tree algorithm and use the dataset for creating tree. Which mainly select the output based on attributes.

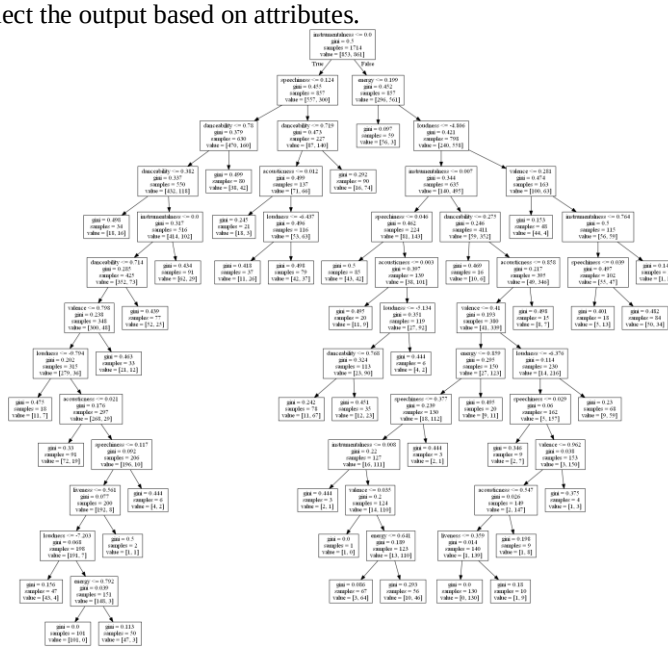

Fig: decision tree.

Now lets see the prediction of the test as well as the results we found on the decision tree algorithm.

```
Out[10]: array([1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1,
                1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1,
                0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 0,
                0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0,
                1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0,
                0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0,
                1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1,
                1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0,
                1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0,
                1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0,
                1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1,
                1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1,
                0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1,
                0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0], dtype=int64)
```

Fig: Testing instances

```
***This is Confusion Matrix***

[[114  53]
 [ 33 103]]
Accuracy: 71.6 %
Precision: 66.0 %
Recall score: 75.7 %
```

Fig: Results from decision tree

Next we applied Naïve Bayes algorithm to find the prediction value of a music and identify the confusion matrix we also analyzed the accuracy and precision of the dataset.

```
***This is Confusion Matrix***

[[138   4]
 [  0 161]]

***Predicting new test instances***

[0. 1. 1. 0. 0. 1. 0. 1. 1. 0. 1. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 1. 1. 1.
 0. 0. 0. 0. 1. 0. 1. 1. 1. 1. 1. 1. 0. 1. 0. 1. 0. 0. 1. 0. 1. 1. 0. 0.
 1. 1. 1. 1. 0. 0. 1. 1. 1. 1. 1. 1. 0. 1. 0. 0. 1. 1. 0. 1. 1. 1. 1. 0. 1.
 1. 1. 0. 1. 1. 0. 1. 0. 1. 0. 1. 1. 1. 1. 1. 0. 0. 1. 1. 1. 0. 0. 0. 0.
 1. 1. 1. 0. 0. 0. 0. 1. 0. 1. 0. 1. 0. 1. 1. 1. 1. 1. 0. 0. 0. 1. 1. 0.
 0. 1. 0. 0. 0. 1. 1. 1. 0. 0. 1. 1. 1. 1. 1. 1. 1. 0. 0. 0. 0. 0. 0. 1.
 0. 0. 1. 1. 0. 1. 1. 0. 1. 1. 0. 1. 1. 0. 0. 1. 0. 1. 1. 1. 1. 0. 1. 1.
 1. 1. 0. 1. 1. 1. 1. 1. 1. 1. 1. 0. 0. 1. 1. 0. 0. 1. 0. 0. 1. 0. 0. 0.
 1. 1. 0. 1. 0. 1. 1. 0. 0. 1. 1. 0. 0. 0. 1. 1. 1. 0. 1. 0. 0. 1. 1. 1.
 0. 1. 1. 1. 1. 1. 1. 0. 1. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 1.
 1. 0. 1. 0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 1. 1. 0. 1. 1.
 1. 1. 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 1. 1. 1. 1. 1. 1. 0. 1. 1. 1. 1. 1.
 1. 1. 0. 0. 1. 1. 1. 0. 1. 0. 0. 1. 1. 1. 0. 1.]
Accuracy:  98.68
Precision:  98.79
training time:  4.203100506624864
testing time:  0.7001611949417723
recall score: 98.59
```

Fig: Naïve Bayes Reults

VIII. Result and Analysis

The goal of the prediction based project was to identify which algorithm is working effectively in certain dataset. We calculate and analysis several distinct attributes to compare between 2 algorithms so that we can check which algorithm is working perfectly in the for the song test dataset. We measure the confusion matrix and with this confusion matrix we calculated the accuracy, precision, and recall score. After calculating these score we can definitely come into a solution. Let's look into that,

Feature name	Decision tree	Naïve bias
Accuracy	71.6%	98.68%
Precision	66%	98.79%
Recall score	75.7%	98.59%

Fig: comparison

So after watching the table we can definitely say that in this specific data set the decision tree algorithm is working less effectively than naïve bias algorithm in every aspects.

IX. Conclusion

The paper describes an efficient approach for finding songs taste based on several algorithms. It creates a broad door of

knowledge for song companies. The music industry is quite big and there is lot of scope of further research. The future of the machine learning is in the hand of efficient algorithms because without efficient algorithms the whole world it sector will fall down drastically.

REFERENCES

[1]. *R. Cooley, B. Mobasher and J. Srivastava: "Data Preparation for Mining World Wide Web Browsing Patterns", Knowledge and Information Systems, Vol. 1, No. 1, pp. 5–32, Feb. 1999. [16] M.S. Chen, J.S. Park and P.S. Yu: "Efficient Machine learning for Path Traversal Patterns", IEEE Transactions on Knowledge and Data Engineering, Vol. 10, No. 2, pp. 209–221, Apr. 1998*

[2] *Behrouz.et.al., (2003) Predicting Student Performance: An Application of Machine learning Methods with The Educational Web-Based System Lon-CAPA © 2003 IEEE, Boulder, CO*

[3] *wiki. [Online] [Cited: 8 27, 2018.] https://en.wikipedia.org/wiki/Decision_tree_learning.* [4]*Morgan, A. 2012, Production Methodologies, Techniques and Guidelines for ModernComputer Game Testing, Ashley Morgan.*

[4] ***Rowlings.*** *Analytics. Vidhya. [Online] [Cited: 8 27, 2018.] https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/* [6]Williams, L. 2006, *Testing Overview and Black-box Testing Techniques, ,*

[5] *2.* ***SVM.*** *Star learning . [Online] [Cited: 8 21, 2018.] http://www.statsoft.com/textbook/support-vector-machines.*

[6]*wiki. [Online] [Cited: 8 27, 2018.] https://en.wikipedia.org/wiki/Apriori_algorithm.*

.