Music Genre Classification using ML

Manogna Katakam, M. Krishna Prasanna[3]
[1] Student, Information Technology, B V Raju Institute of Technology, Telangana, India
[2] student, Information Technology, B V Raju Institute of Technology, Telangana, India
[3] Assistant Professor, Information Technology, B V Raju Institute of Technology , Telangana, India

ABSTRACT

This research work has trained and compared the proposed models on GTZAN dataset. It consists of 1000 audio files each having 30 seconds duration. There are 10 classes (10 music genres) each containing 100 audio tracks. Each track is in .wav format. It contains audio files of the following 10 genres: Blues, Classical, Country, Disco, Hip-hop, Jazz, Metal, Pop, Reggae, Rock.There are various methods to perform classification on this dataset. We will use K-nearest neighbors' algorithm.

Keyword: GTZAN dataset, genre, KNN Algorithm

1. INTRODUCTION

1.1 Machine Learning

Machine learning is the field of study that allows computers to learn without being explicitlyprogrammed. Using machine learning, we do not need to provide explicit instructions to Computers for reacting to some special situations. We need to provide training on computersto find real-time solutions for specific problems. The chess game is a famous example wheremachine learning is being used to play chess. The code lets the machine learnand optimizes itself over repeated games.

Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for building mathematical modelsand making predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, Facebook auto-tagging, recommender system, and many more.

1.1.1 Types of Machine Learning Algorithms

- *Supervised Learning Algorithm:*

Supervised learning is a type of Machine learning in which the machine needs external supervision to learn. The supervised learning models are trained using the labelled dataset. Once the training and processing are done, the model is tested by providing sample test data to check whether it predicts the correct output.

The goal of supervised learning is to map input data with the output data. Supervised learning is based on supervision, and it is the same as when a student learns things under the teacher's supervision. An example of supervised learning is spam filtering.

Supervised learning can be divided further into two categories of problem:

1) Classification
2) Regression

Examples of some popular supervised learning algorithms are Simple Linear regression, Decision Tree, Logistic Regression, KNN algorithm, etc.

- *Unsupervised Learning Algorithm:*

It is a type of machine learning in which the machine does not need any external supervision to learn from the data, hence called unsupervised learning. The unsupervised models can be trained using the unlabeled dataset that is not classified, nor categorized, and the algorithm needs to act on that data without any supervision. Inunsupervised learning, the model does not have a predefined output, and it tries to find useful insights from the huge amount of data. These are used to solve the Association and

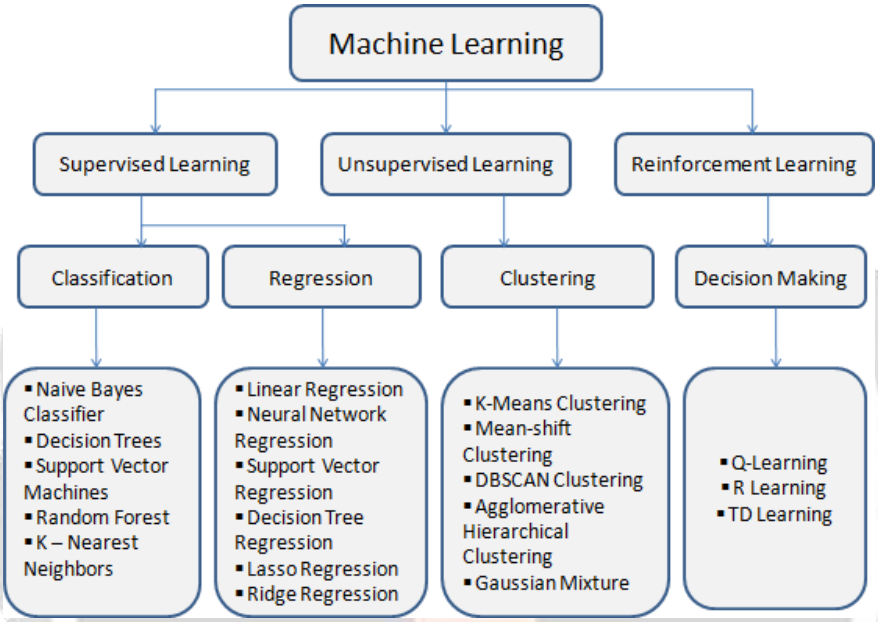
Clustering problems.

Hence further, it can be classified into two types:

1) Clustering
2) Association

Reinforcement Learning:

In Reinforcement learning, an agent interacts with its environment by producing actions, and learns with the help of feedback. The feedback is given to the agent in theform of rewards, such as for each good action, he gets a positive reward, and foreachbad action, he gets a negative reward. There is no supervision provided for theagent. Q-Learning algorithm is used in reinforcement learning.



1.1.2 Some Popular Machine Learning Algorithms

Linear Regression:

Linear regression is one of the most popular and simple machine learning algorithms that is used for predictive analysis. Here, **predictive analysis** defines prediction of something, and linear regression makes predictions for *continuous numbers* such as **salary, age, etc.** It shows the linear relationship between the dependent and independent variables and shows how the dependent variable(y) changes according to the independent variable (x). It tries to best fit a line between the dependent and independent variables, and this best fit line is known as the regression line.

Logistic Regression:

Logistic regression is the supervised learning algorithm, which is used to predict categorical variables or discrete values. It can be used for the classification problems inmachine learning, and the output of the logistic regression algorithm can be either Yesor NO, 0 or 1, Red or Blue, etc. Logistic regression is like the linear regression

except how they are used, such as Linear regression is used to solve the regression problem and predict continuous values, whereas Logistic regression is used to solve the Classification problem and used to predict the discrete values.

Decision Tree Algorithm:

A decision tree is a supervised learning algorithm that is mainly used to solve classification problems but can also be used for solving regression problems. It can work with both categorical variables and continuous variables. It shows a tree-like structure that includes nodes and branches and starts with the root node that expands on further branches till the leaf node. The internal node is used to represent the features of the dataset, branches show the decision rules, and leaf nodes represent the outcome of the problem.

1.2 Deep Learning

- Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain—albeit far from matching its ability—allowing it to "learn" from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy.
- Deep learning drives many AI applications and services that improve automation, performing analytical and physical tasks without human intervention. Deep learning technology lies behind everyday products and services (such as digital assistants, voice-enabled TV remotes, and credit card fraud detection) as well as emerging technologies (such as self-driving cars).

1.3 Existing Systems

Historically, attempts made by others to build music genre classification systems have yielded fine but not extraordinary results.

There were models which used the following:

- K-Means Clustering
- Convolutional Neural Networks

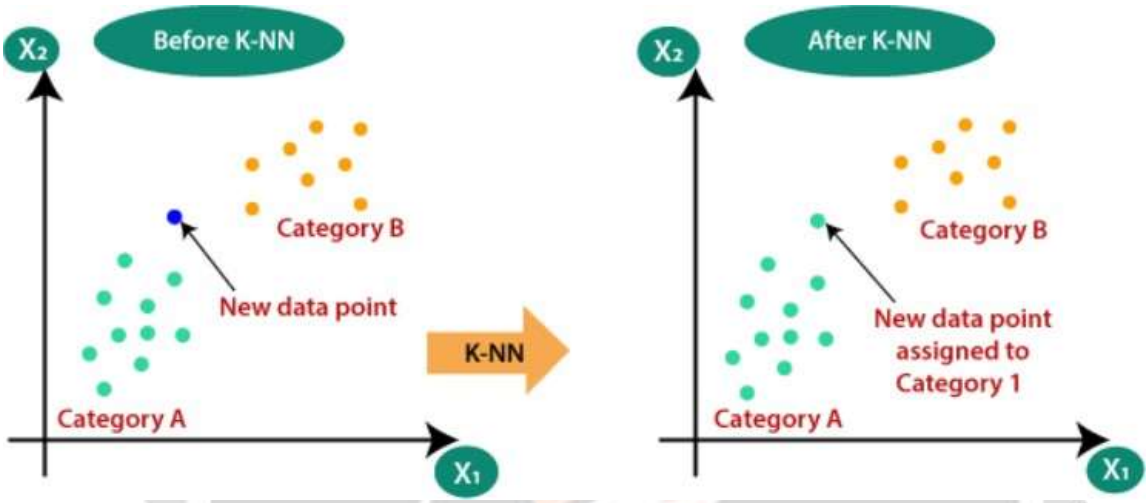
METHODOLOGY

We used the K-Nearest Neighbors algorithm because various research proves it is one of the best algorithms to give good performance and till time along with optimized models organizations uses this algorithm in recommendation systems as support.

KNN Algorithm:

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most like the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.

- It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much like the new data.

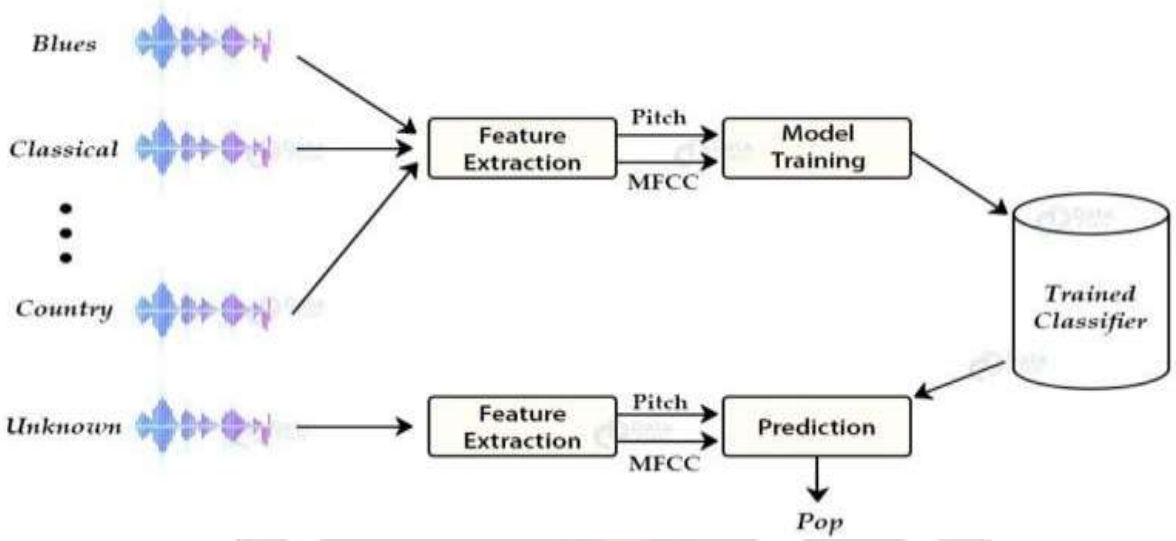


1.4 GTZAN Dataset

The GTZAN Database is used to enter data into the system because it is a collection of free accessible songs from many genres. The collection is made up of thousands of audio tracks dividedinto 10 different genres. This database includes blues, classical, country, disco, hip-hop, jazz, metal,pop, reggae, and rock music. Music data on the GTZAN database is taken at 22050 Hz and lasts forabout 30seconds, a total of 22020 x 30 = 661500 samples. For each smooth window of 2048 samples, with a change of 1024samples, as calculated during the study, all the results provided below are rated at more than ten runs, and the accuracy of the sections was selected as metric performance metrics.

Music Genre	**Number of Songs**
Blues	1000
Classical	1000
Country	1000
Folk	1000
Hip-Hop	1000
Jazz	1000
Metal	1000
Pop	1000
Reggae	1000
Rock	1000
Total	**10000**

1.5 Architecture

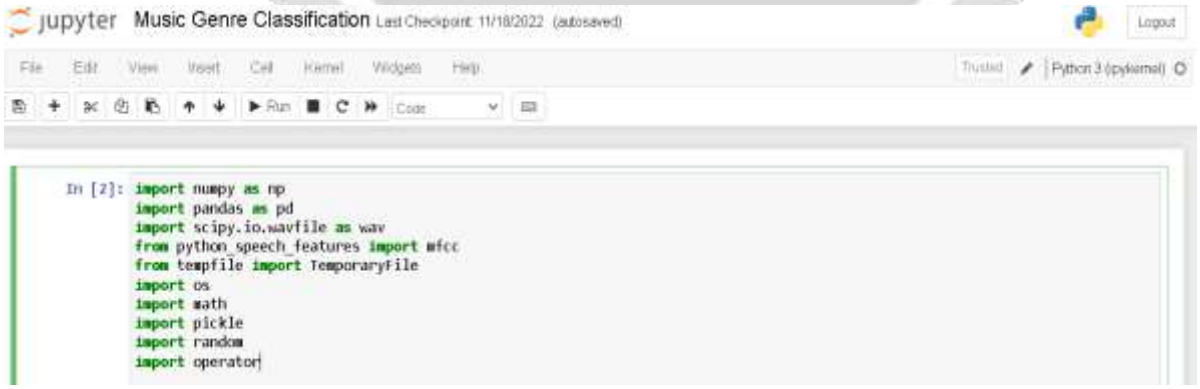


1.6 Advantages Over the Existing System

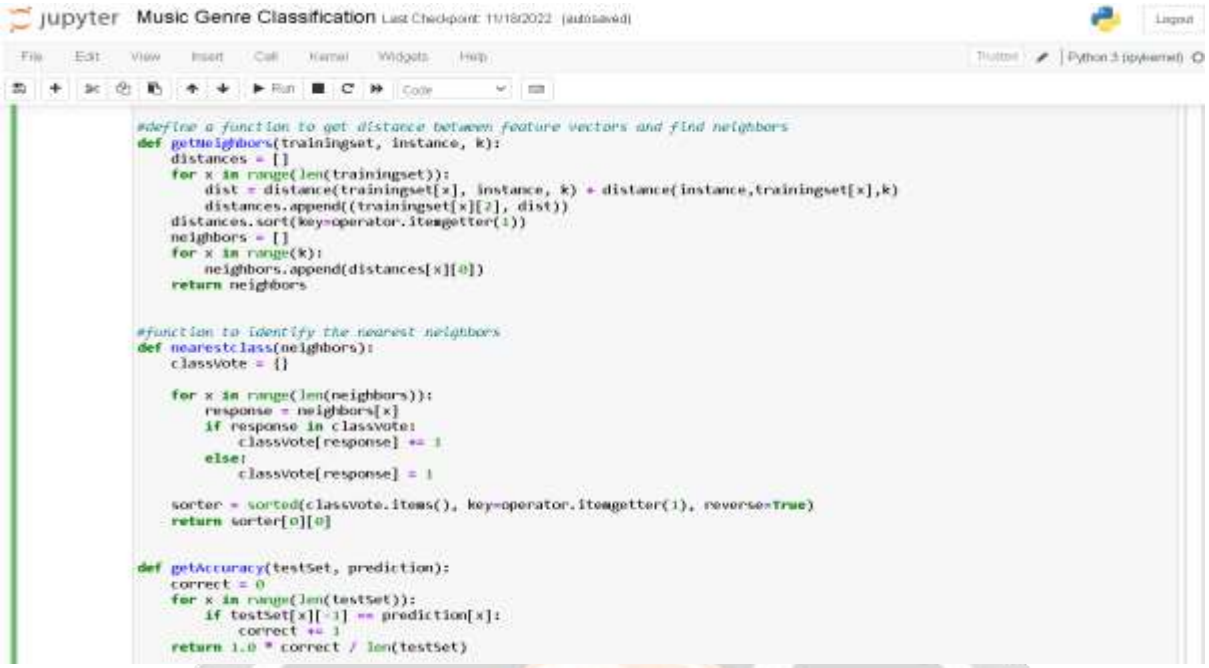
- Quick calculation time
- Simple algorithm – to interpret
- Versatile – useful for regression and classification
- High accuracy – you do not need to compare with better-supervised learning models
- No assumptions about data – no need to make additional assumptions, tune severalparameters, or build a model. This makes it crucial in nonlinear data cases.

3. RESULTS

Step-1&2: Importing Libraries and Data



Step-3&4: Distance between featurevectors and find neighbors, identify nearestneighbors



Step-5&6: Loading Data set

```
dataset = []

def loadDataset(filename, split, trset, teset):
    with open('mydataset.dat','rb') as f:
        while True:
            try:
                dataset.append(pickle.load(f))
            except EOFError:
                f.close()
                break
    for x in range(len(dataset)):
        if random.random() < split:
            trset.append(dataset[x])
        else:
            teset.append(dataset[x])

trainingSet = []
testSet = []
loadDataset('mydataset.dat', 0.68, trainingSet, testSet)
```

Step 7: Prediction using KNN

```
def distance(instance1, instance2, k):
    distance = 0
    mm1 = instance1[0]
    cm1 = instance1[1]
    mm2 = instance2[0]
    cm2 = instance2[1]
    distance = np.trace(np.dot(np.linalg.inv(cm2), cm1))
    distance += (np.dot(np.dot((mm2-mm1).transpose(), np.linalg.inv(cm2)), mm2-mm1))
    distance += np.log(np.linalg.det(cm2)) - np.log(np.linalg.det(cm1))
    distance -= k
    return distance


# Make the prediction using KNN(K nearest Neighbors)
length = len(testSet)
predictions = []
for x in range(length):
    predictions.append(nearestclass(getNeighbors(trainingSet, testSet[x], 5)))
```

Step 8: Results

```
from collections import defaultdict
results = defaultdict(int)

directory = "Data/genres_original"

i = 1
for folder in os.listdir(directory):
    results[i] = folder
    i += 1


pred = nearestclass(getNeighbors(dataset, feature, 5))
print(results[pred])
```

```
Got an exception:  File format b'\xcb\x15\x1e\x16' not understood. Only 'RIFF' and 'RIFX' supported. in folder:  jazz  fil
ename:  jazz.00054.wav
0.7169230769230769
rock
```

```
In [ ]:
```

```
In [ ]:
```

4. CONCLUSIONS

- We started the project with the initial setup and used MFCC to extract features from audiofiles. After that, we have built a KNN classifier from scratch that finds the K number of nearest neighbor based on features and maximum neighbor belonging to class gives as an output. We got approximately 70 per cent accuracy on the model.

5. REFERENCES

- **https://www.analyticsvidhya.com/blog/2022/03/music-genre- classification-project-using-machine-learning-techniques/**
- **http://ijics.com/gallery/21-may-1186.pdf**
- *https://www.irjmets.com/uploadedfiles/paper/volume3/issue_7_july_2 021/14720/1628083571.pdf*