**NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY**

(AN AUTONOMOUS INSTITUTION, AFFILIATED TO VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELGAUM, APPROVED BY AICTE & GOVT.OF KARNATAKA

****

**PROJECT REPORT**

On

MUSIC ANALYSIS

*Submitted in partial fulfilment of the requirement for the award of Degree of*

*Bachelor of Engineering*

*in*

*Computer Science and Engineering*

*Submitted by:*

Riya Yadav 1NT19CS159

Thota Thanmai 1NT19CS203

Varshini N 1NT19CS211

Pola Udaya Sowjanya Reddy 1NT19CS136

Under the Guidance of

DR. Vani V

IEEE SM|MACS CP|ABET PV| Professor, Dept. of CS&E, NMIT



Department of Computer Science and Engineering

**(Accredited by NBA Tier-1)** 2021-2022

**NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY**

(AN AUTONOMOUS INSTITUTION, AFFILIATED TO VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELGAUM

, APPROVED BY AICTE & GOVT.OF KARNATAKA)

Department of Computer Science and Engineering

**(Accredited by NBA Tier-1)**

****

**CERTIFICATE**

This is to certify that the **I SUPPORT-For tablet detection and reading in speech format**

is an authentic work carried out by **Riya Yadav(1NT19CS159),Thota Thanmai(1NT19CS203),Varshini N(1NT19CS189) and Pola Udaya Sowjanya Reddy(1NT19CS136)** bonafide students of **Nitte Meenakshi Institute of Technology**, Bangalore in partial fulfilment for the award of the degree of ***Bachelor of Engineering*** in COMPUTER SCIENCE AND ENGINEERING of Visvesvaraya Technological University, Belgavi during the academic year ***2021-2022.*** It is certified that all corrections and suggestions indicated during the internal assessment has been incorporated in the report. This project has been approved as it satisfies the academic requirement in respect of project work presented for the said degree.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Internal Guide** | | **Signature of the HOD** | | **Signature of Principal** | |
|  | |  | |  | |
| Dr. Vani V  Designation: IEEE SM|MACS  CP|ABET PEV|Professor-CSE | | Dr. Sarojadevi H.  Professor, Head, Dept. CSE, NMIT Bangalore | | Dr. H. C. Nagaraj  Principal,  NMIT,  Bangalore | |
| **Signature of Examiner (**Internals**)** | | **Signature of Examiner (**SEE**)** | |
|  | |  | |
|  | | 2. | |

**DECLARATION**

We hereby declare that

(i) The project work is our original work

(ii) This Project work has not been submitted for the award of any degree or examination at any other university/College/Institute.

(iii) This Project Work does not contain other persons’ data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This Project Work does not contain other persons’ writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

(v) This Project Work does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

|  |  |  |
| --- | --- | --- |
| **NAME** | **USN** | **Signature** |
| Riya Yadav | 1NT19CS159 |  |
| Thota Thanmai | 1NT19CS203 |  |
| Varshini N | 1NT19CS211 | C:\Users\RYadav\Downloads\WhatsApp Image 2022-01-16 at 10.12.56 PM.jpeg |
| Pola Udaya Sowjanya Reddy | 1NT19CS136 |  |

Date: 17/01/2022

**ACKNOWLEDGEMENT**

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crowned our effort with success. I express my sincere gratitude to our Principal **Dr. H. C. Nagaraj**, Nitte Meenakshi Institute of Technology for providing facilities.

We wish to thank our HOD**, Dr. Sarojadevi H.** for the excellent environment created to further educational growth in our college. We also thank him for the invaluable guidance provided which has helped in the creation of a better project.

I hereby like to thank our Dr. Vani V , Designation: IEEE SM|MACSCP|ABET PEV |Professor-CSE Department of Computer Science & Engineering on **her** periodic inspection, time to time evaluation of the project and help to bring the project to the present form.

Thanks to our Departmental Project coordinators. We also thank all our friends, teaching and non-teaching staff at NMIT, Bangalore, for all the direct and indirect help provided in the completion of the project.

|  |  |  |
| --- | --- | --- |
| **NAME** | **USN** | **Signature** |
| Riya Yadav | 1NT19CS159 |  |
| Thota Thanmai | 1NT19CS203 |  |
| Varshini N | 1NT19CS211 | C:\Users\RYadav\Downloads\WhatsApp Image 2022-01-16 at 10.12.56 PM.jpeg |
| Pola Udaya Sowjanya Reddy | 1NT19CS136 |  |

Date: 17/01/2022

**ABSTRACT**

People listen to different types of music daily and enjoy them irrespective of the music flow. Some will relate to the music because it gives the good vibe, and it is in sync with their mood. Some people just enjoy the music irrespective of the language and feel every part of it if it connects to their vibe. Vibe to the song plays a role in liking a music or disliking it. In this project it focuses on the music ability and nature, and it is analysed whether the song is danceable, whether it is sad song, energy of the song, artist of the song, popularity of a song etc. We use different data processing algorithms and classification techniques for the data. We plot graph for different attributes and get to know that which song is best for the attributes given in the dataset. Accuracy of the decision tree classifier and the K-NN classifier is decided.

**CONTENT**

**DECLARATION**

**ACKNOWLEDGEMENT**

**ABSTRACT**

**TABLE OF CONTENTS**

**CHAPTER 1: INTRODUCTION**

* 1. Motivation
  2. Problem Domain
  3. Aim and Objectives

**CHAPTER 2: DATA SOURCE AND DATA QUALITY**

* 1. Dataset Used
  2. Data pre-processing

**CHAPTER 3: METHODS & MODELS**

3.1Data mining questions

3.2 Data mining algorithms

3.3 Data Mining Models

**CHAPTER 4: MODEL EVALUATION & DISCUSSION (WITH VISUALIZATION)**

**CHAPTER 5: CONCLUSION &FUTURE DIRECTION**

**CHAPTER6: REFLECTION PORTFOLIO**

**REFERENCES**

**APPENDICES**

**CHAPTER 1: INTRODUCTION**

* 1. ***Motivation***

The major reason for using data mining techniques is to extract useful information and knowledge from huge amount of data. The information and knowledge gained can be used for various applications. Music has been considered as a mean of expressing and arousing emotions. So here music dataset is used to analyze its attributes and extract the useful information from it.

* 1. ***Problem Domain***

Music dataset is used to extract the useful information and knowledge where initially data pre-processing is done and then with the help of the processed data most popular songs, comparison between different attributes and its corresponding graphs are plotted. Accuracy of decision tree and KNN classifier is performed.

***1.3 Aim and Objectives***

1) To extract the useful information and knowledge from the data set.

2) Data pre-processing is done to remove the missing, noisy and inconsistent data.

3) Sampling with/without replacement is done to determine the accuracy of the research.

4) From the processed data, top 10 most popular song is determined.

5) Comparison between different attributes is done and its corresponding graphs are plotted***.***

**CHAPTER 2:DATA SOURCE AND DATAQUALITY**

***2.1 Dataset used***

Dataset used is taken from the Kaggle website and it contains many rows and columns. It contains 586673 rows and columns from A -T i.e. (20 columns).

It contains the name of the song, popularity, danceability, energy, liveness,

loudness, artist etc and we use these elements for analysis and come to conclusion whether the song matches with their vibes and also likeability of the song by the people.

***2.2 Data pre-processing***

* Data pre-processing is a data mining technique used for transforming of raw data to a useful information [1].
* Using pre-processing algorithms in the data mining,
* Number of the missing values in each row is calculated.
* Removal of rows containing missing values is done.
* Duplicate values are also calculated and after applying algorithm we came to know that there are zero duplicate values.
* Discretization and sampling techniques are performed.
* Discretization is performed to convert huge values to smaller ones [2]. Large values like energy in” music analysis” is converted to smaller one.
* Sampling is used for selection of random values from the data set because for some algorithms we just test the random values. In “music analysis” sample value is taken as 3 and values are taken accordingly. Randomly selected 1% of the data (without replacement) and displays the selected samples.

**CHAPTER 3: METHODS & MODELS**

* 1. ***Data mining Questions***

1.How to convert milliseconds to seconds?

2.How to deal with the missing values?

3.How to remove missing values?

4.How to find top 10 popular songs?

5.How to plot the graph between different attributes?

6.What algorithms are used to find the correlation between the attributes?

***3.2 Data mining Algorithms***

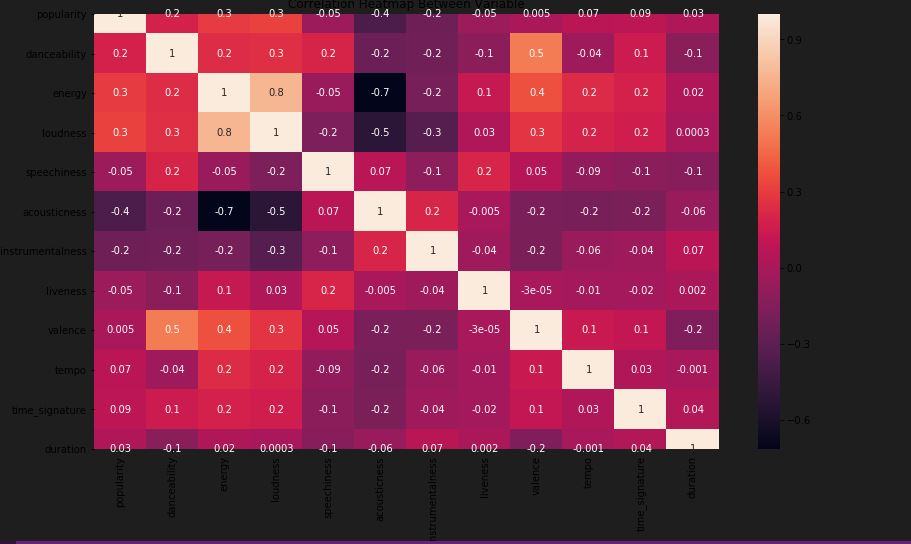
We use the algorithms for the normalization of the time where the time in in milliseconds is converted to seconds. Algorithm to find the top 10 popular songs is implemented. The release date is changed to the correct format. Correlation using the Pearson method algorithm is performed. Plot between the loudness and energy is represented in the graph format. Accuracy for the decision tree classifier and K-NN classifier is done on both test and training set.

* 1. ***Data mining Models***

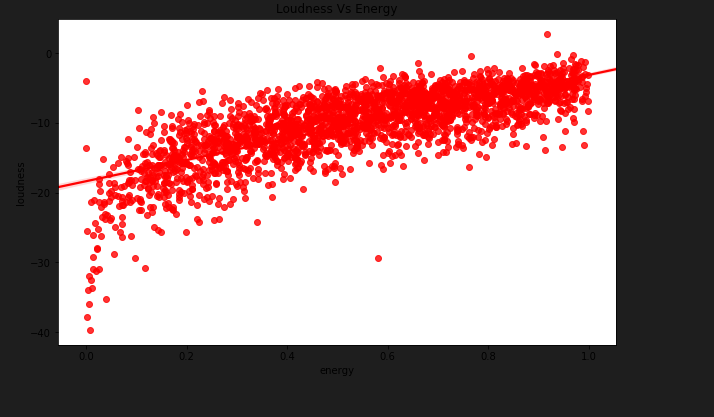
Analyzing the data based on the data mining algorithms[11]. In the music analysis decision tree classifier and K-NN classifier accuracy is tested. We test the results of the accuracy for both test and training set which is chosen randomly from available dataset.

**CHAPTER 4: MODEL EVALUATION & DISCUSSION (WITH VISUALIZATION)**

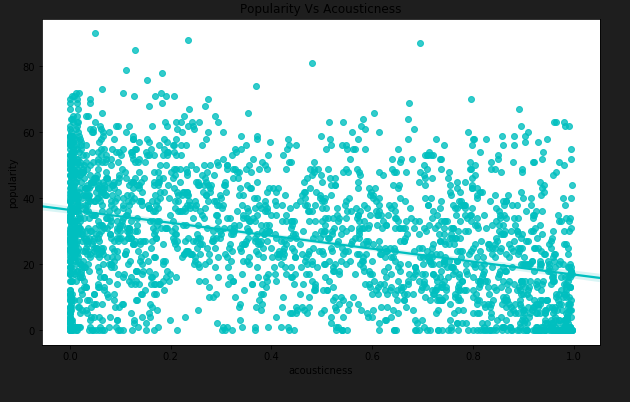
**Correlation with using Pearson method**



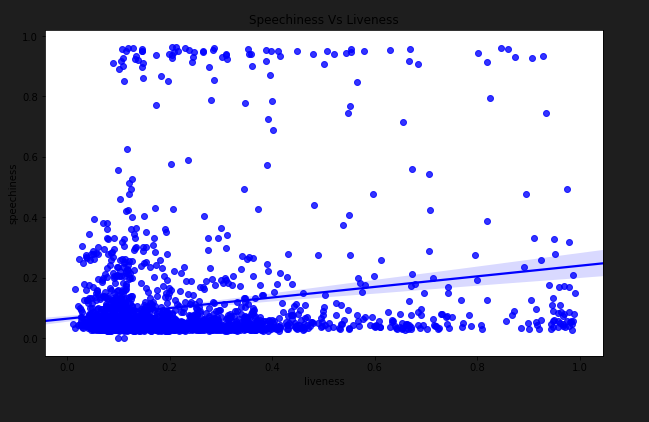
**Loudness vs energy**



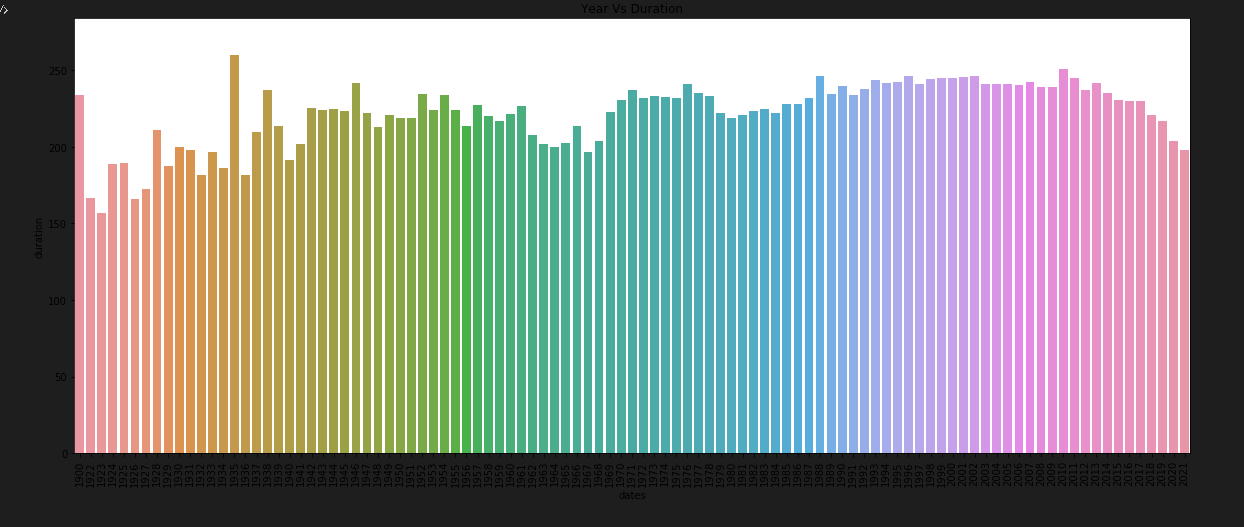
**Popularity vs Acousticness**

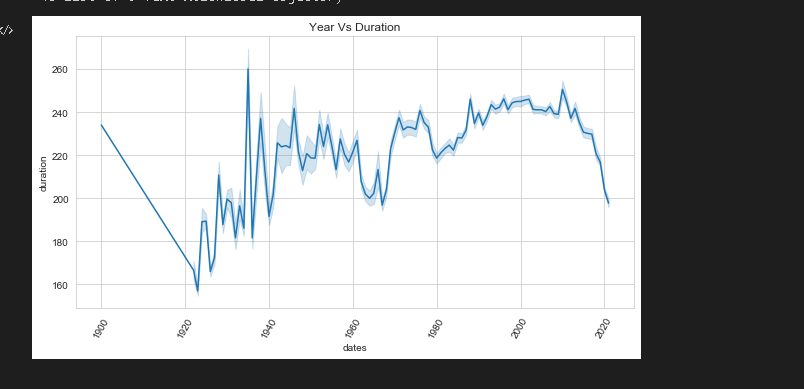


**Speechiness and liveness**



**Year vs duration**





**CHAPTER 5: CONCLUSION &FUTURE DIRECTION**

**CONCLUSION:**

Through dataset, it can be concluded that there are many songs and missing values in the dataset. The most popular songs in the dataset can be analysed. Also what is the duration of the song in each year and how it is changing every year .The energy and loudness of the song is estimated through the graph. The pre-processing techniques performed on the dataset and missing values are counted and removal of the missing values is done.

**FUTURE DIRECTION:**

There are many datasets available online and we have analysed for just one dataset .For many dataset the popular song might differ as the no of songs released per day is only more and every song might not be included in the dataset Many other algorithms can also be implemented as we have plotted the graph and the accuracy for the dataset is calculated for both test set and training set.

**CHAPTER6: REFLECTION PORTFOLIO**

We have learnt a lot from the project. We came to know how to use the algorithms and implement the algorithms to a dataset practically. We analysed the relationship between the attributes and plotted its corresponding graph and could extract useful information from the dataset.

**REFERENCES**

**[1]** [**https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/amp/**](https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/amp/)

**[2]** [**https://www.javatpoint.com/discretization-in-data-mining**](https://www.javatpoint.com/discretization-in-data-mining)

**[3]** [**https://www.kaggle.com/adisrw/spotify-data-analysis-using-python**](https://www.kaggle.com/adisrw/spotify-data-analysis-using-python)

**[4]** [**http://www.cse.msu.edu/ptan/dmbook/software/**](http://www.cse.msu.edu/ptan/dmbook/software/)

**[5]**[**https://www.kaggle.com/search?q=music+dataset**](https://www.kaggle.com/search?q=music+dataset)

**[6]**[**https://www.springboard.com/blog/data-science/data-mining-python- tutorial/**](https://www.springboard.com/blog/data-science/data-mining-python-%20%20%20tutorial/)

**[7]**[**https://www.tutorialspoint.com/data\_mining/index.htm**](https://www.tutorialspoint.com/data_mining/index.htm)

**[8]**[**https://www.youtube.com/watch?v=4rymD1Hpnho**](https://www.youtube.com/watch?v=4rymD1Hpnho)

**[9]** [**https://data-flair.training/blogs/python-project-music-genre-classification**](https://data-flair.training/blogs/python-project-music-genre-classification)

**[10]**[**https://www.kaggle.com/adisrw/spotify-data-analysis-using-python**](https://www.kaggle.com/adisrw/spotify-data-analysis-using-python)

**[11]** [**https://docs.microsoft.com/en-us/analysis-services/data-mining/mining-models-analysis-services-data-mining**](https://docs.microsoft.com/en-us/analysis-services/data-mining/mining-models-analysis-services-data-mining)

APPENDICES

APPENDIXA:<https://www.kaggle.com/adisrw/spotify-data-analysis-using-python>

APPENDIX B :jupyter notebook