CSE3505

FOUNDATION OF DATA ANALYTICS

ABSTRACT

Prepared By

Sri HarshaVardhan Palla -20BCE1308

Rukksana A -20BCE1968

Siddharth M -20BPS2007

Background:-

Listening to music has become one of the most frequently resorted to pastimes of people ranging from the youth to the elder. While there are umpteen songs of different genres and artists from yesteryears in the podcast, it becomes essential that there is a recommendation System that analyzes the liking of a specific user with the help of the datasets genre and artists of the songs that he/she listened to in the past three days. This project aims to envision a music recommendation system with such a function

Objective:-

Music service providers need an efficient way to manage songs and help their costumers to discover music by giving quality recommendation. Thus, there is a strong need of a good recommendation system. The objective of this project is to develop a music recommendation system which will determine the musical preferences of the users based on the analysis of their interaction during use . Our system learns from the users past listening history and recommends them songs which they would probably like to hear in future. Currently music service providers have generic, mood-based playlists, that are the same for all users. Here, we suggest improvements to these playlists by offering custom playlists for each user based on user input .

Methods:-

k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster.It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training.It is a centroid-based algorithm .The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

Findings and Conclusion:-

From our correlation plot, we observed that variables have strong correlation with each other, indicating that this dataset has multicollinearity. With further deep diving into this matter, we learnt that such types of dataset are not suitable for various classification algorithms. So we dropped our plan for CART (Classification Tree) and Random Forest.