Song Recommendation System

Team K-Pop Clusterers

Martand Javia (AU1841064)
Namit Shah (AU1841067)
Suhanee Patel (AU1841113)
Devarsh Patel (AU1841146)

MACHINE LEARNING

Introduction

- There are more than 100 million songs currently available on the internet.
- Classifying songs based on different attributes has become difficult over time.
- Every song have some similarity index with other songs and thus can be grouped together.
- Our end goal here is to create a recommendation system which groups different songs based on their features, and suggest songs which belongs to the same group as user's preference.

Problem Statement

- Design a model which can recommend songs based on certain attributes and features.
- Create an unsupervised learning model to classify the song dataset into various different clusters.
- Create a precise recommendation model including multiple features of songs.

Existing Body of Work

A collaborative recommendation system by H. Chen wherein the music samples are grouped based on similarity index and user's preferences.

Many system computes the mean of the songs preferred by the user and then recommend songs based on these values.

Reference: https://dl.acm.org/doi/10.1145/502585.502625

MACHINE LEARNING

Approach

Processing Dataset

Inorder to use the dataset for building a recommendation system, it is necessary to clean and process data.

Finding optimal K-value

It is important to find an optimal K-value in order to build a legitimate clustering model.

Clustering Dataset by k-means

Dataset is clustered into several different groups to differentiate among each other.

Fuzzy C Means Clustering

Clustering using another algorithm knows as fuzzy c clustering algorithm.

Recommendation system

We create a system, that finds the songs similar to user's preference using the clustered dataset obtained by manual, inbuilt and also by the fuzzy-c implementation of clustering algorithm..

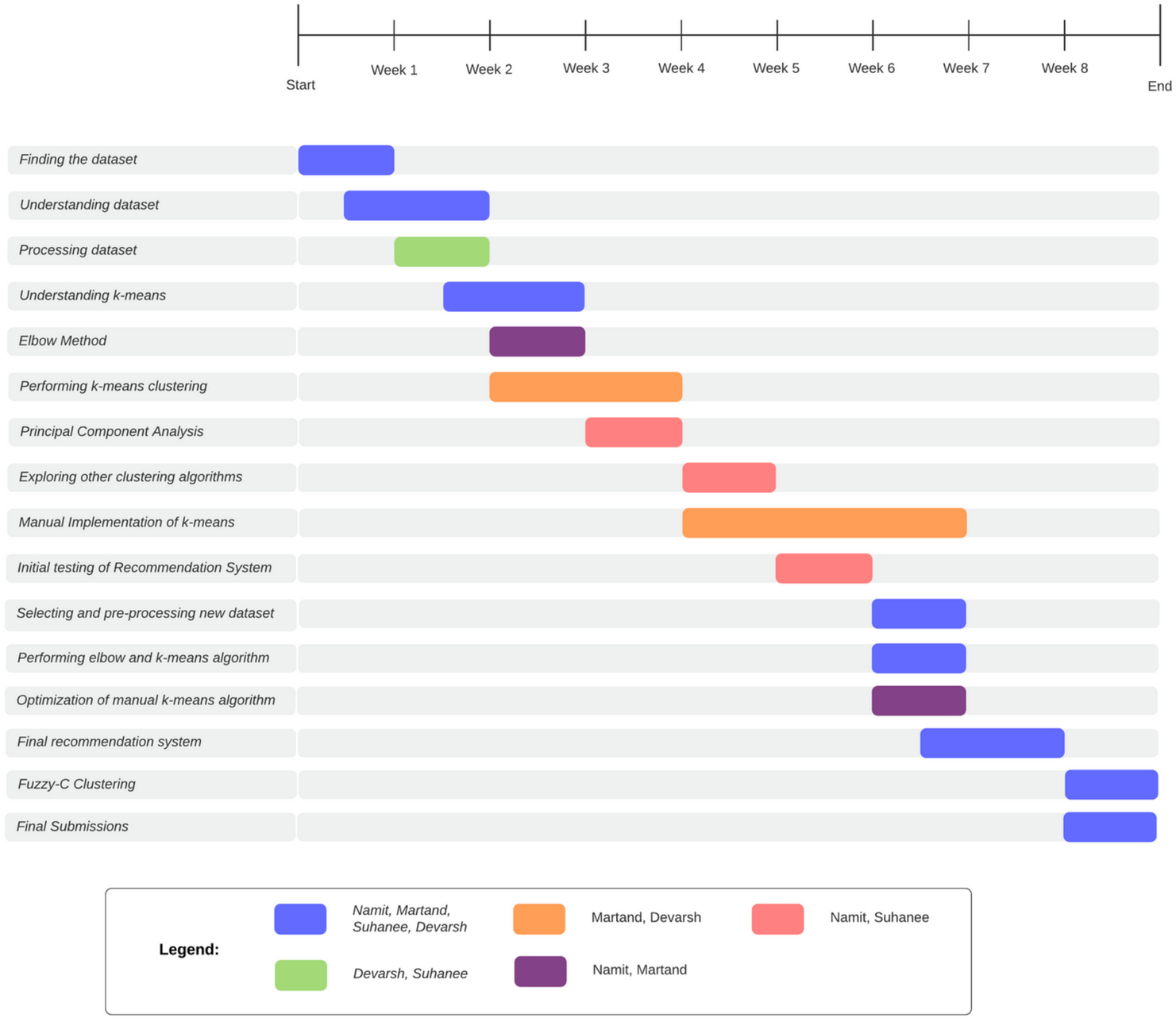
Fuzzy-C Reference: https://www.researchgate.net/publication/222868331_FCM-the_Fuzzy_C-Means_clustering-algorithm

Dataset Reference: https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks

K-means Reference: https://ieeexplore.ieee.org/document/5453745

MACHINE LEARNING

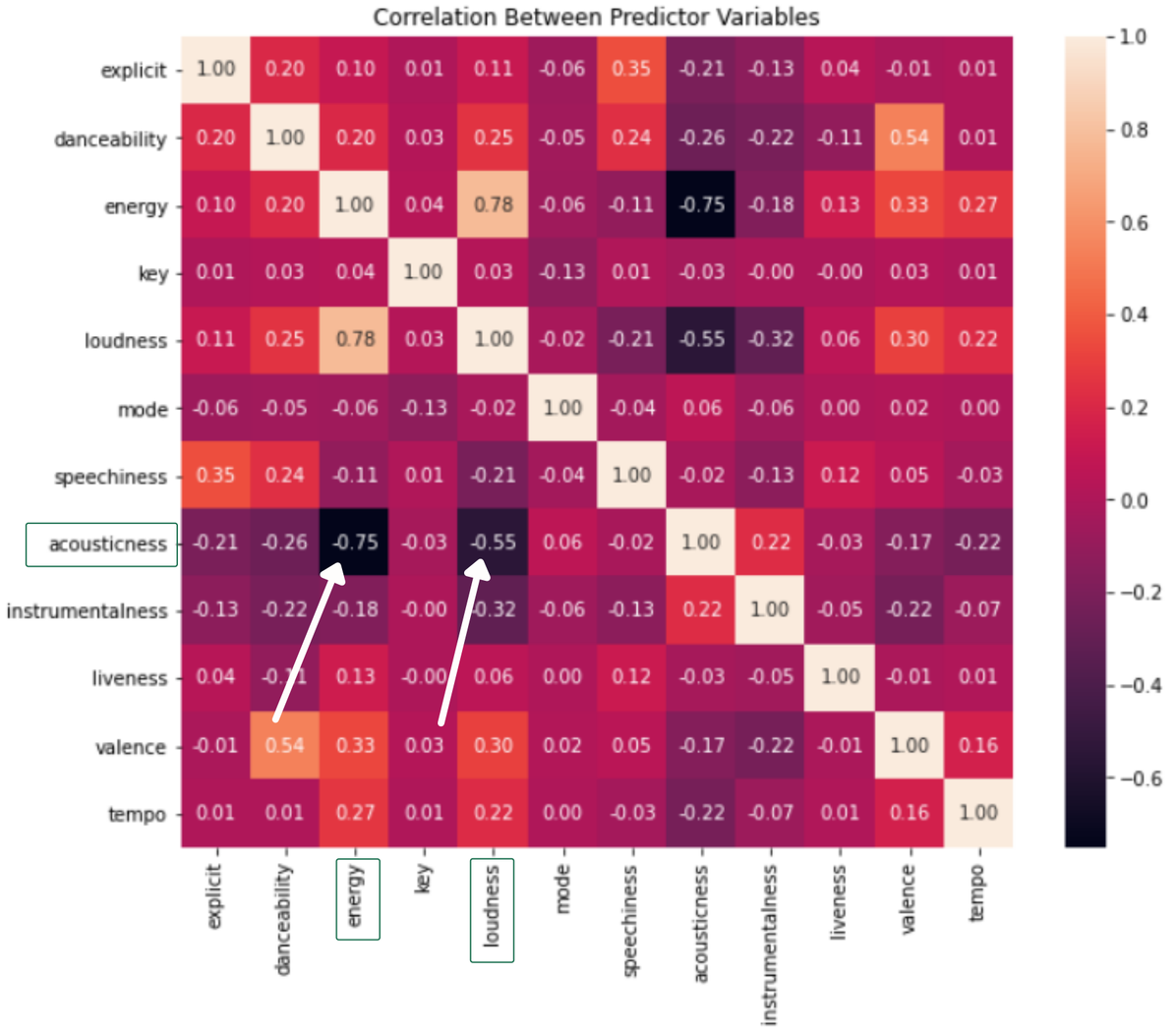
GANTT Chart



Final Results

Data Correlation

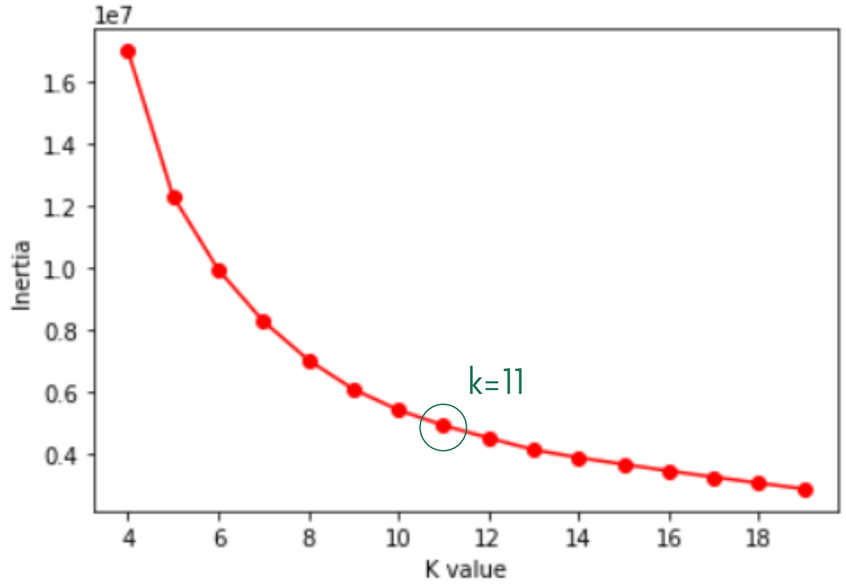
Correlation matrix of all the audio features of the song

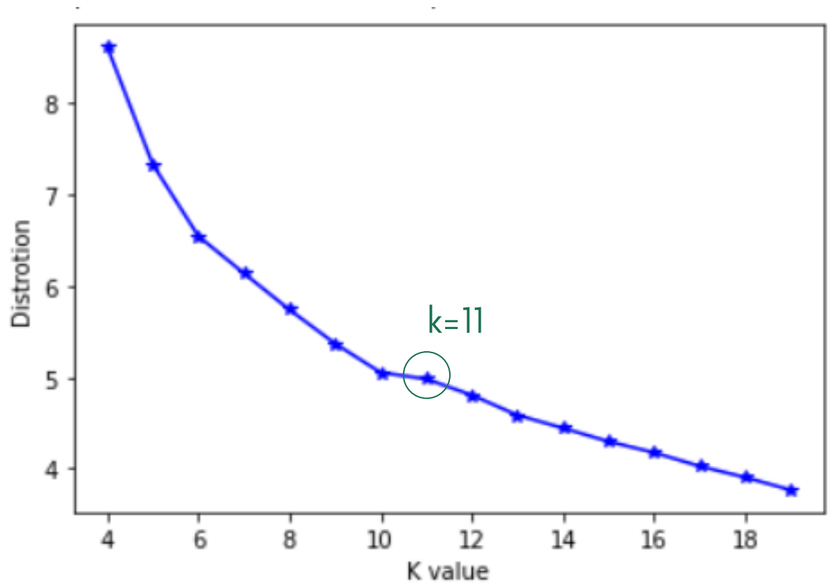


Due to high correlation between Acousticness, Energy, and Loudness, two of these features were dropped.

Elbow Method

Performing elbow method to obtain the optimum number of clusters for K-Means clustering.





MACHINE LEARNING

Inbuilt K-Means Recommendation System

Songs Recommended based on the clusters formed by Inbuilt K-Means Implementation

```
recommend_songs()
```

```
Enter a song name : Perfect
Enter artist names of the song entered (',' seperated if > 1) : Ed Sheeran
Add More ? (Y/N) : N
Number of songs to be recommended for each song : 3
Cluster #8
Avg. cosine similarity : [0.99999825]
```

	name	artists	explicit	danceability	key	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo
18808	Perfect	['Ed Sheeran']	0	0.599	8	1	0.0232	0.1630	0.000000	0.1060	0.168	95.050
88401	Wasted	['Carrie Underwood']	0	0.548	8	1	0.0306	0.0936	0.000000	0.0721	0.311	94.998
99923	All Through the Night	['Cyndi Lauper']	0	0.576	8	1	0.0252	0.2130	0.000001	0.0619	0.340	95.100

MACHINE LEARNING

Manual K-Means Recommendation System

Songs Recommended based on the clusters formed by Manual K-Means Implementation

```
recommend_songs()
```

```
Enter a song name : Perfect
Enter artist names of the song entered (',' seperated if > 1) : Ed Sheeran
Add More ? (Y/N) : N
Number of songs to be recommended for each song : 3
Cluster #8
Avg. cosine similarity : [0.99999825]
```

	name	artists	explicit	danceability	key	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo
18808	Perfect	['Ed Sheeran']	0	0.599	8	1	0.0232	0.1630	0.000000	0.1060	0.168	95.050
88401	Wasted	['Carrie Underwood']	0	0.548	8	1	0.0306	0.0936	0.000000	0.0721	0.311	94.998
99923	All Through the Night	['Cyndi Lauper']	0	0.576	8	1	0.0252	0.2130	0.000001	0.0619	0.340	95.100

MACHINE LEARNING

Multiple Songs Recommendation System

Songs Recommended based on the clusters formed by Manual Implementation for multiple number of songs which are of similar type.

```
recommend_songs()
```

```
Enter a song name : Often
Enter artist names of the song entered (',' seperated if > 1) :
Add More ? (Y/N) : Y
Enter a song name : sobeautiful
Enter artist names of the song entered (',' seperated if > 1) :
Add More ? (Y/N) : N
Number of songs to be recommended for each song : 2
Cluster #1
Cluster #1
Avg. cosine similarity : [0.9999843]
```

	name	artists	explicit	danceability	key	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo
18462	Often	['The Weeknd']	1	0.572	7	0	0.0476	0.2220	0.0	0.1350	0.0713	134.078
54380	sobeautiful	['Musiq Soulchild']	0	0.689	7	0	0.0637	0.0656	0.0	0.0885	0.2100	133.987
34114	Un Siglo Sin Ti	['Chayanne']	0	0.611	7	0	0.0279	0.0961	0.0	0.1290	0.2610	133.906
15880	Brand New Year 2021	['Nikky Philip']	0	0.556	7	0	0.0306	0.0026	0.0	0.3800	0.3650	134.024

MACHINE LEARNING

Multiple Songs Recommendation System

Songs Recommended based on the clusters formed by Manual Implementation for multiple number of songs of different types

```
recommend_songs()
```

```
Enter a song name : 7 Rings
Enter artist names of the song entered (',' seperated if > 1) :
Add More ? (Y/N) : Y
Enter a song name : Perfect
Enter artist names of the song entered (',' seperated if > 1) : Ed Sheeran
Add More ? (Y/N) : N
Number of songs to be recommended for each song : 3
Cluster #1
Cluster #8
Avg. cosine similarity : [0.99999742]
```

	name	artists	explicit	danceability	key	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo
19201	7 rings	['Ariana Grande']	1	0.778	1	0	0.3340	0.5920	0.000000	0.0881	0.327	140.048
106400	idfc - Tarro Remix	['blackbear', 'Tarro']	1	0.580	1	0	0.2210	0.4810	0.000001	0.1090	0.389	139.751
19350	Lemonade	['Internet Money', 'Gunna', 'Don Toliver', 'NAV']	1	0.799	1	0	0.0790	0.2560	0.000000	0.1110	0.471	140.040
18808	Perfect	['Ed Sheeran']	0	0.599	8	1	0.0232	0.1630	0.000000	0.1060	0.168	95.050
88401	Wasted	['Carrie Underwood']	0	0.548	8	1	0.0306	0.0936	0.000000	0.0721	0.311	94.998
99923	All Through the Night	['Cyndi Lauper']	0	0.576	8	1	0.0252	0.2130	0.000001	0.0619	0.340	95.100

Fuzzy-C Recommendation System

Songs Recommended based on the clusters formed by Fuzzy-C Implementation

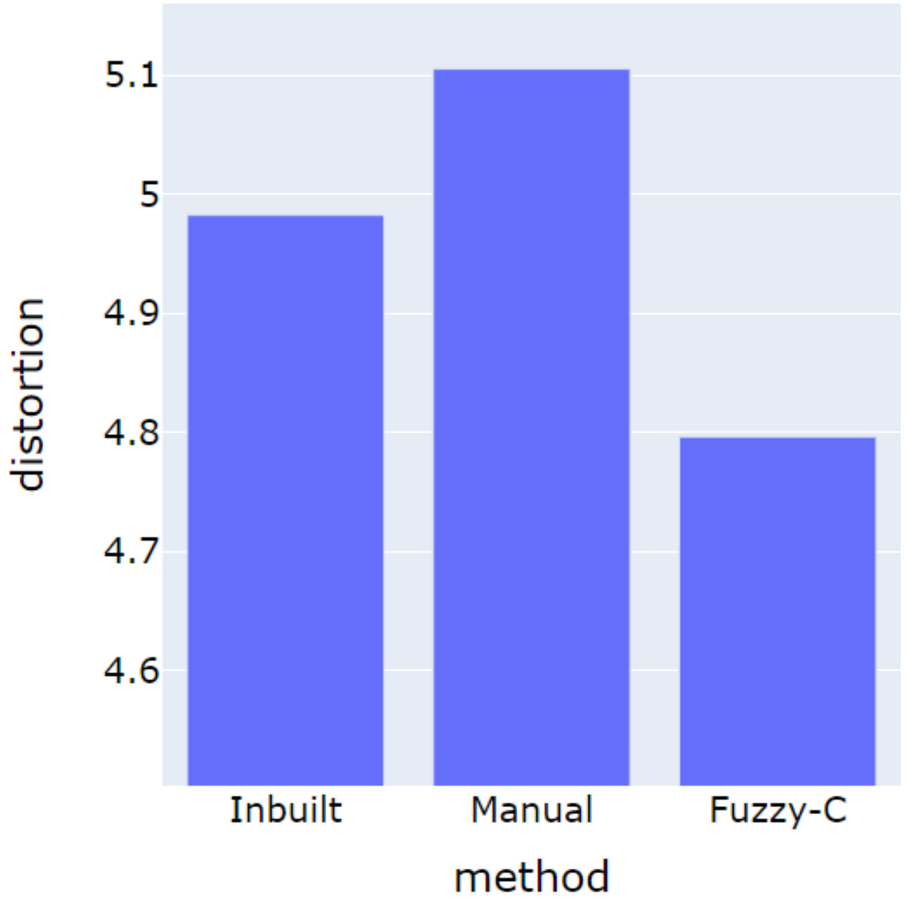
```
recommend_songs(sys="Fuzzy-C")
```

```
Enter a song name : Perfect
Enter artist names of the song entered (',' seperated if > 1) : Ed Sheeran
Add More ? (Y/N) : N
Number of songs to be recommended for each song : 3
Cluster #4
Avg. cosine similarity : [0.99999825]
```

	name	artists	explicit	danceability	key	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo
18808	Perfect	['Ed Sheeran']	0	0.599	8	1	0.0232	0.1630	0.000000	0.1060	0.168	95.050
88401	Wasted	['Carrie Underwood']	0	0.548	8	1	0.0306	0.0936	0.000000	0.0721	0.311	94.998
99923	All Through the Night	['Cyndi Lauper']	0	0.576	8	1	0.0252	0.2130	0.000001	0.0619	0.340	95.100

Clustering Algorithm Comparison

Distortion values for clusters formed by each clustering algorithms to compare the spread of each clusters.



Conclusion

In the proposed recommendation system created using the manual implementation of K-Means is almost in line with the performance of inbuilt K-Means and fuzzy-c algorithm. Moreover, the system can recommend songs for diverse user inputs and preferences. The proposed work can be improved using larger and varied dataset as well as using another machine learning approaches.

Thank You!