

|  |
| --- |
| AMPBA class of 2020 winter  Term IV |
|  |
| Team: Group-1  Unsupervised Machine learning-II  Project Proposal  Karthik Shankar S (PGID: 11920008)  Pundareek Chandrashekhar (PGID: 11920049)  Rajesh Rangaswamy (PGID: 11920072)  Praveenkumar Agoorkaisetty (PGID: 11920096) |



Project Title:

Song recommendation system

Project Statement:

From an offline repository having details of listening history of English track songs, we aim to create a recommendation system to a music listener based on his /her music listening history or preference. The recommendation system is being devised to produce the desired output through Unsupervised Machine learning algorithm.

Dataset:

The core of the dataset is the feature analysis and metadata for one million songs, provided by The Echo Nest. The dataset does not include any audio, only the derived features. The Million Song Dataset is a cluster of complementary datasets contributed by the community, where in we are considering only the following.

1. User Data: Data compromise of about 1.1K number of unique users in the form of user ids.
2. Song Metadata: This data file contains more than 3,80,000 songs.
3. Taste Profile Subset: This is a subset of bigger data that has about more than million user listening histories with features involving user, song and play count.

Source: <https://www.kaggle.com/c/msdchallenge>

Research Question:

1. How to suggest or recommend a song based on user listening history?
2. What would be the song recommendation for a new user based on the model being developed?
3. What type of modelling technique would yield a better song recommendation or suggestion?

Solution Implication:

This is expected to inspire and develop a one stop platform for music lovers, where the enthusiast listens to the recommended track without much fuss. This also would make the engine refine its suggestion over time and improve the prediction accuracy. The implication of such recommender system is also expected to increase Subscriber Retention, improved engagement and responsive to changing dynamics of subscriber listening preference.

A revenue model can be built on top of this channel, if it impresses the advertisement agencies of the subscriber base which our system will be aggregating due to increased patronage for prediction accuracy or rightly suggesting a song in the first instance. This business proposition may take-off similar to popular entertainment portals and join the league of successful entertainment channels.

Citation:

Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The Million Song Dataset. In Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011), 2011.

Acknowledgement:

The Million Song Dataset was created under a grant from the National Science Foundation, project IIS-0713334. The original data was contributed by The Echo Nest, as part of an NSF-sponsored GOALI collaboration. Subsequent donations from SecondHandSongs.com, musiXmatch.com, and last.fm, as well as further donations from The Echo Nest, are gratefully acknowledged.