# HIT SONG PREDICTION FOR INDIAN POPULAR MUSIC

Shreya Kale	Rameswhari Joshi	Aparna Dhaygude	Vaishnavi Ingole	Makarand Velankar
shreya.kale@cum- minscollege.in	rameshwari.joshi@cum- minscollege.in	aparna.dhaygude@cum- minscollege.in	vaishnavi.in- gole@cummin- scollege.in	makarand.ve- lankar@cummin- scollege.in

# ABSTRACT

The music industry attracts a large number of investors due to its high turnover. Releasing hit songs can garner profits, whereas flop songs lead to losses. Thus, predicting the popularity of a song before its release can help in promotion plans. Can we really predict hit Songs? This is the main motivation of the work. Extensive work is being done for western songs, but Indian music is relatively less explored. Hence, our work aims to predict hit songs of Indian origin using acoustic features. To analyze tracks, a data set is created from data provided by the Spotify Web API. The features are extracted using Spotify, available libraries such as Librosa and aubio, which are passed to machine learning algorithms for prediction. Along with available features, melodic features based on patterns are proposed and extracted. A comparative analysis is done for four acoustic feature sets containing timbral, pitch, rhythm and melodic features proposed. Further experimentation is performed with the combined features sets resulting in the improved performance. The results are encouraging and hit song prediction can be a reality in the near future..

# **1. INTRODUCTION**

Hit song prediction, usually referred to as 'Hit Song Science (HSS)', is defined as, "an emerging field of investigation concerned with predicting the success of songs before they are released on the market" [2]. A song is considered a hit when it has many unique streamers on music streaming services such as YouTube, Spotify, etc. or has a commendable rank in a chart dis-playing a list of hit songs such as the Billboard Hot 100, etc. [1]. In theory, all hit songs have a specific set of features which are common to them and are almost always present. These similarities can be exploited to make popularity predictions prior to market release of songs. The underlying assumption in HSS is that hit songs are similar with regard to their features. The work attempts to vali-date this hypothesis. These predictions can help artists curate and create songs that will have a greater chance of being hit and can also help artists better judge the likes and dislikes of the general population.

# 2. METHODOLOGY

During the literature survey, it was observed that the data set for Indian popular music is not readily available. Hit

song prediction is a hard problem considering various factors affecting the popularity of song. These factors can be broadly classified into 2 categories as internal and external factors. Internal factors refers to song as a product, whereas external factors refers to marketing campaign or promotion used to reach out majority of target audience. Several researches have been conducted of which many focus on predicting hit songs using internal factors such as composition of music and frequency, loudness, acoustics, dance-ability, tempo etc. Internal factors also include artist features, genre, lyrics, video, etc.Many other researches on prediction use external factors such as social media impact, view counts on various platforms, release dates, promotional campaigns and teasers conducted before the song release, release dates (such as important holidays or release of songs on themes of certain days), etc. Sometimes not so good songs may get more popularity because of social media impact or more pleasant songs can't get that much popularity because of low social media impact, less promotion campaign, wrong release date, etc. So, the effect of external factors is difficult to model and predict on the success of a song.

Feature Type	Features Included	Features Count
Timbral	accousticness, instru- mentalness, mfcc	28
Pitch	croma-mean, croma- std, zrate- mean, zrate- skew, key, mode	29
Rhythmic	danceability,loud- ness,tempo, time_sig- nature, intensity, spec- trual centroid, contrast, rolloff	31
Melodic	Valance,speechiness, ref_node, bi-gram, tri- gram	62

#### Table 1. Feature Group

Considering the need of a data set with purely Indian songs, it was necessary to create our own data set. Firstly, a data set obtained from Spotify python library which fetches data from Spotify is used. Indian songs are fetched by using API calls to Spotify and fetching these songs. Moreover, our approach requires distinguishing between hits and non-hits for musical success. The success of a song is defined based on the popularity value obtained. The initial experimentation is carried out for internal factors with acoustic features only. The methodology used is presented Figure using 1 From the entire bunch of extracted features, we had performed feature classification by classifying all 153 features into its property wise grouping such as all the features which are related to timber were included in Timbral feature classification. Similarly, the features associated with pitch info are grouped under Pitch feature classification. Features based on energy level and song's rhythmic properties, are grouped under Rhythmic feature set. Melodious features included patterns of notes represented as bi-grams and tri-grams. The feature groups with features used are as shown in Table 1.



Figure 1. Hit Song Prediction System

### 3. RESULTS AND DISCUSSIONS

Feature classification experiment results as shown in Table 2, same performance metrics were analyzed for different machine learning classifiers. Timbral features which specifically have MFCC and acousticness generated 64.87% accuracy with quite better results for F-measure that is 0.65. Pitch feature set mainly focused chroma features provided 63.51% accuracy. Similarly Rhythmic feature set were almost in line with other 2 groups, whereas melodic features results were not effective for hit song prediction. This is probably because of the inherent nature of melody present in all Indian songs. The overall observations were similar in line with the western music by other researchers with exception such as chords which are rarely used in Indian music. Another possible feature specific to Indian music could be hummable, as majority of songs are based on melody. It would be interesting to explore this feature and measure hummability of song in some way.

SVM	Tim-	Pitch	Rhyth-	Melodic
	bral		mic	
Accuracy	64.87	63.51	62.50	57.09
Precision	0.64	0.63	0.63	0.57
Recall	0.65	0.64	0.63	0.57
F-Score	0.65	0.64	0.62	0.57
RMSE	0.59	0.60	0.61	0.66
MAE	0.35	0.37	0.38	0.43

 Table 2. Results using SVM and ML classifier

## 4. FUTURE DIRECTIONS

The future directions could be identifying additional features such as lyrics or artist popularity. Further possible impacts of external factors can be analyzed and further modeled to improve prediction accuracy. The data set used while training the model has limited instances thus, expanding the data set can prove beneficial for classifying the song more accurately. Including other internal factors in the mix such as lyric-based features, video features, etc. Also, the model used does not include influence of external factors on popularity of the song such as artist popularity, current social media trends, release dates, marketing strategies, etc. These external factors can be used to improve accuracy. Songs can be hit or non - hit by using external factors. Hit song prediction can be a reality in near future by successfully modelling all possible internal and external factors.

#### 5. REFERENCES

- [1] Christopher Monterola, Cheryl Abundo, Jeric Tugaff, and Lorcel Ericka Venturina. Prediction of potential hit song and musical genre using artificial neural net works. International Journal of Modern Physics C, 20(11):1697–1718, 2009.
- [2] Franc ois Pachet and Pierre Roy. Hit song science is not yet a science. In ISMIR, pages 355–360, 2008.