Predicting and Analysing the Viral Fragment of Songs

*CSC 440: Data Mining Project

1st Shubham Shailesh Tamhane
MS in Data Science
University of Rochester
Rochester, USA
stamhane@ur.rochester.edu , UR ID: 32372250

2nd Ajeesh Ajayan Nayaruparambil
MS in Data Science
University of Rochester
Rochester, USA
anayarup@ur.rochester.edu , UR ID: 32381214

***Abstract*—By following an systematic approach, this study analyses the viral fragments of songs. The study observed significant features which were present in the viral fragments which were otherwise absent in the rest of the song. The study also includes comparison of various machine learning models which were implemented to predict the viral fragment of unseen new songs.**

***Index Terms*—short form content, dynamic time warping, class imbalance, data mining**

I. Introduction

Nowadays, short-form video content ranging between 15-60 seconds are getting popular. Platforms such as TikTok, Instagram Reels and YouTube Shorts are being extensively used to consume this content. Production and consumption of this content has become part and parcel of everyone's life, especially teenagers and young adults. Previous studies have shown that short form content platforms are successfully captivating user's attention. Engagement with the application such as scrolling the Reels, etc increases the amount of time spent on the application which is desired by the platforms. Another thing to notice is that all these platforms are vertical video viewing platforms unlike traditional long length content like YouTube which is mostly viewed in horizontal. This can be further elaborated by saying that the content which was once professionally shot in set backgrounds has changed to in-the-moment shoots usually done with mobile cameras. As a result, the platforms providing the ease of producing content and enabling users to consume content easily opened up a new way to produce and consume video content

II. Problem Statement

The aim of this project is to understand what makes some song fragments more socially popular than others, and to predict popular fragments for newer songs. Song clips to be analyzed are the ones which are present in the short-form content applications stated above. Analysis can be done by understanding various components of the audio file such as bass, beats, rhythm, etc. Alternatively, the lyrics can also be used to derive the reason behind the popularity of the clip.

III. Related work

Fine-grained Video Attractiveness Prediction Using Multi-modal Deep Learning on a Large Real-world Dataset[1]. This paper focuses on finding how much does each video segment attract views and finding the video segments of several second length which are attractive. This is similar to our idea to find the viral song fragment.

Factors influencing Instagram Reels usage behaviours[2]. An examination of motives, contextual age and narcissism. This paper focused on the use of Instagram reels for India only. The drawback of this paper is that the analysis done may not align with other cultural backgrounds. The researchers recommend using original eighteen item contextual age scale instead of the three which they have used.

The paper titled Predicting Music Popularity on Streaming Platforms[3] focuses on assessing the popularity of a song in the forms of scores and rankings. Then, predict whether an already popular song may attract more public interest and become viral. It also tries using acoustic features such as MFCC, Spectral Centroid, Spectral Flatness, Zero Crossings and Tempo to make predictions. Adding the acoustic information did not improve the model. The finalized model was trained only using the data from song popularity scores and rankings.

The paper titled "nowplaying the Future Billboard: Mining Music Listening Behaviors of Twitter Users for Hit Song Prediction"[4] aims to understand general music trends. Research was conducted which focuses on using Twitter hashtags and music-related tweets to understand general music trends. Ultimately, it was used to forecast the Billboard rankings and hit music.

The report [5] focuses on how users react to different kind of reels. The literature dated previous to the paper focused on how a singular emotional state influences consumer behavior. However, while scrolling through different reels, the user is exposed to random variety of emotional valences in a short period of time. The report concludes that positive videos which trigger the happy emotions increase the willingness-to-pay

(WTP) for users.

The paper [6] analyses how the YouTube algorithm works. As the data fetched in our project extracts data from YouTube, this paper has been referred. The paper focuses on finding out which way rating given by YouTube influence it to be on the trending list. The researchers have used ”Rapidminer” software to perform the data mining tasks, namely classification, association and clustering. The paper concludes that that the likes, dislikes, views and comments of the audience influences the trending videos for YouTube.

IV. Methodology

A. Data Acquisition

Initially, a list was created comprising of the songs which are popular in TikTok, Instagram Reels and YouTube Shorts. From these applications, the most viral or viewed soundtrack was extracted corresponding to the above-mentioned list. It is possible that some videos may not have a soundtrack, so they need to be removed as no information can be gained without soundtrack for this project. After the videos are collected, the audio files can be separated from them, and features were extracted from these soundtracks. pytube and pafy library in python language are used for acquiring data.

B. Analysis

Different features of the audio files are explored. Trends are observed in the clips which are used in the social media platforms. The data is initially transformed from .mp3 format to amplitude and sampling rate. For this project, the sampling rate used is 22050. Analysis is performed after we get the amplitude and sampling rate. The different tasks that are performed are creating and plotting the spectral centroid and spectral rolloff. The zero crossing rate is also taken into consideration. It tells us how many times does the signal cross 0, i.e, positive to negative or negative to positive. We also plot spectrogram and chromagram. The chromagram plots the time with respect to the audio file against the pitch classes which are B, A, G, F, E, D, C in descending order. Librosa library was primarily used to perform analysis.

C. Modelling

A dataset was created by dividing the full entire song into windows of the short duration. Then, the short was compared with each of these windows for a song and the most similar one was chosen. This was done by using dynamic time warping (DTW). Dynamic time warping is an algorithm for measuring similarity between two temporal sequences. These g sequences were acquired by calculating the mfcc values for the audio signals. Several machine learning models were used. Some of them are Logistic Regression, Weighted Logistic Regression, Naive Bayes, Support Vector Machines(SVM), Weighted Support Vector Machines(SVM), Random forest, SVM bagging classifier.

V. Performance Evaluation

Multiple models were created and used to predict the clips of the song which should get popular. This model was tested on the test data and the best was chosen. Models predicting the clips having the most overlap with the actual viral clip is considered as the best model. Other evaluation metrics such as precision, recall, F1 score and more were also considered while selecting a model.

VI. Results

Figures 1 to 10 are the analysis of one of the YouTube shorts. Similar analysis was done on the rest of the dataset to obtain these plots and for exploring the features. Figure 8 shows the comparison of different algorithms. Figures 9 to 12 are plots of various metrics which were used to evaluate the machine learning models.

Analysis Results:

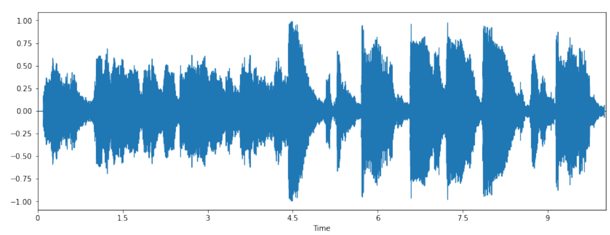

Fig. 1. Waveplot

The waveplot shows how the audio signal looks like. It is normalized so that all audio signals can follow the same scale

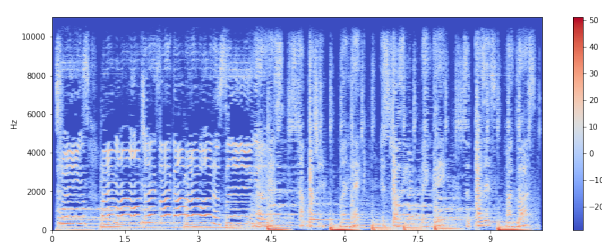

Fig. 2. Spectrogram without taking log

A spectrogram is a visual representation of the spectrum of frequencies of a signal as it varies with time. If the spectrogram is plotted directly, not much information can be gained as the orange-reddish values lie on the bottom of the graph. The spectrogram is plotted in the form of heatmap where the vertical axis shows frequency and horizontal axis shows time.

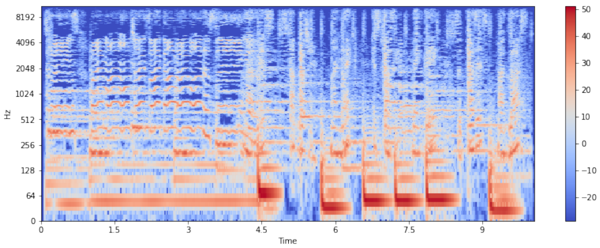

Fig. 3. Spectrogram after taking log

After taking the logarithm, the values get scaled and we get a better representation of the audio file by looking at this spectrogram.

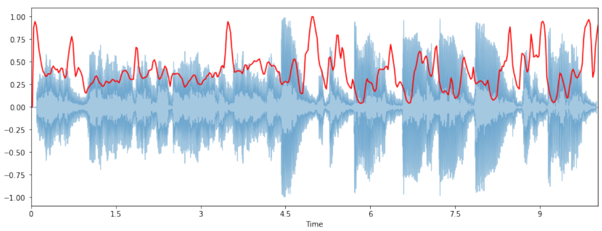

Fig. 4. Spectral Centroid

Spectral Centroid is used to indicate the frequency at which energy of a spectrum is centered upon. It can also be considered as center of mass for sound. In the above plot we can see that there is a rise in spectral centroid after half of the audio signal.

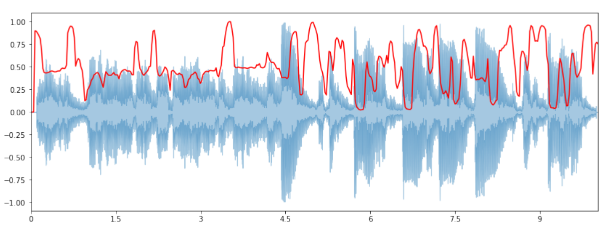

Fig. 5. Spectral Rolloff

Spectral rolloff is used to measure the shape of the signal.

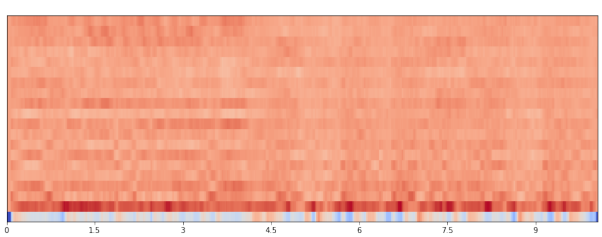

Fig. 6. Spectrogram of MFCC values

MFCC which stands for Mel frequency cepstral coefficients represent the small set of features(20 in our case) which concisely describe the overall shape of the spectral envelope.

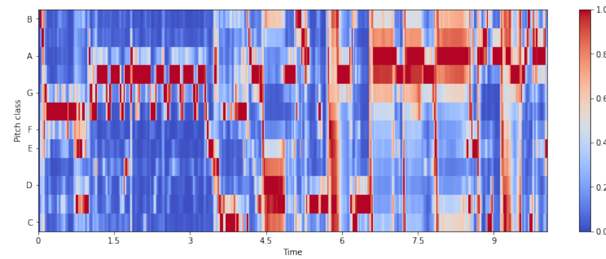

Fig. 7. Chromagram

Chromagram represents the chroma features which typically is a 12 element feature vector indicating how much energy of each pitch class(y-axis) is present in the signal. This is often used while classifying music in different genres.

Modelling Results:

Our main focus was on maximizing both Recall and accuracy since the data was highly imbalanced. We have plotted cross validation accuracy and Recall of 7 different models as well as their test accuracy and Recall. The Undersampling rate should also be kept as minimum as possible (preferably less than or equal to 0.5).

	sampling_rate	Algorithm_detailed_name	Algorithm	Accuracy_cv	Recall_cv	Accuracy	Recall
0	0.5	Weighted Logistic Regression {0: 0.5, 1: 1}	Weighted Logistic Regression	0.344697	1.000000	0.300000	1.000000
1	0.5	Weighted SVM Classifier {0: 0.5, 1: 1}	Weighted SVM Classifier	0.344697	1.000000	0.300000	1.000000
2	0.4	Weighted Logistic Regression {0: 0.4, 1: 1}	Weighted Logistic Regression	0.304396	1.000000	0.234043	1.000000
3	0.4	Weighted SVM Classifier {0: 0.4, 1: 1}	Weighted SVM Classifier	0.304396	1.000000	0.234043	1.000000
4	0.3	Weighted Logistic Regression {0: 0.3, 1: 1}	Weighted Logistic Regression	0.233987	1.000000	0.224138	1.000000
5	0.3	Weighted SVM Classifier {0: 0.3, 1: 1}	Weighted SVM Classifier	0.233987	1.000000	0.224138	1.000000
6	0.2	Weighted Logistic Regression {0: 0.2, 1: 1}	Weighted Logistic Regression	0.172283	1.000000	0.150000	1.000000
7	0.2	Weighted SVM Classifier {0: 0.2, 1: 1}	Weighted SVM Classifier	0.172283	1.000000	0.150000	1.000000
8	0.5	Naive Bayes Classifier	Naive Bayes Classifier	0.574242	0.270000	0.625000	0.166667
9	0.4	Naive Bayes Classifier	Naive Bayes Classifier	0.654396	0.213333	0.680851	0.090909

Fig. 8. Comparision of algorithms

These are the top 10 algorithms ordered in a descending manner by Recall and Accuracy and with lower Undersampling rate. We have to give importance to Recall as the data is imbalanced. We can observe that Weighted Logistic Regression and Weighted SVM Classifier are the preferred models due to high Recall. But the accuracy, as you can observe, is not much favorable. Let's visualize the plots.

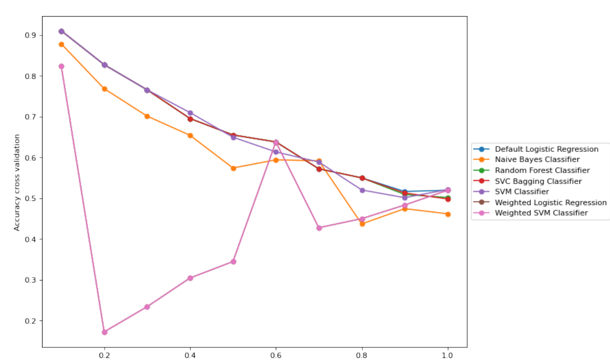

Fig. 9. Cross validation Accuracy at various Undersampling rates

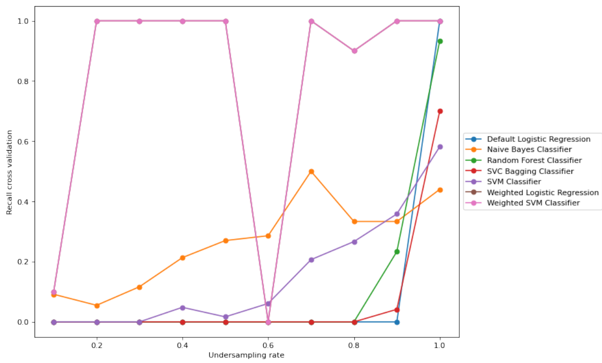

Fig. 10. Cross validation Recall at various Undersampling rates

The cross validation accuracy of SVC Bagging Classifier, SVM Classifier and Naive Bayes Classifier are decent enough. But their corresponding cross validation Recall values are very low. Hence, the next best models are Weighted Logistic Regression and Weighted SVM Classifier having Recall = 1 at many Undersampling rates. The optimum Undersampling rate to be choosen would be 0.5.

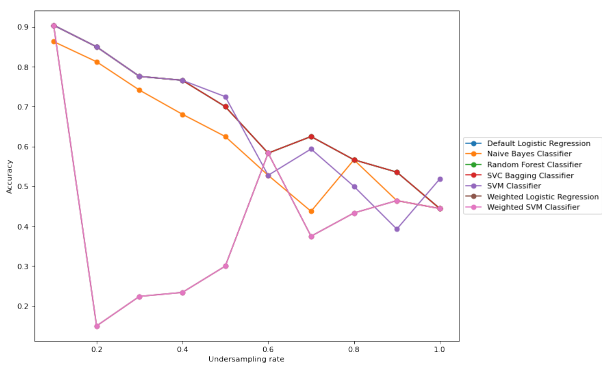

Fig. 11. Test Accuracy at various Undersampling rates

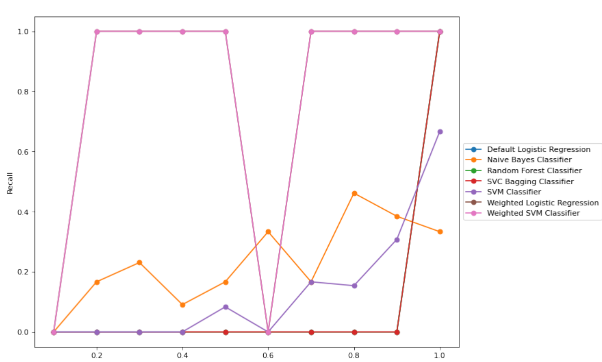

Fig. 12. Test Recall at various Undersampling rates

Similarly, the test accuracy of SVC Bagging Classifier, SVM Classifier and Naive Bayes Classifier are decent enough. But their corresponding test Recall values are very low. Hence, the next best models are Weighted Logistic Regression and Weighted SVM Classifier having Recall = 1 at many Undersampling rates. The optimum Undersampling rate to be choosen would be again 0.5.

VII. Conclusion

In this report we have discussed an approach to predict and analyze the viral fragment of songs. There are several works in the literature that deal with this problem. Our approach is unique in the way that it completely focuses on the acoustic features to isolate the viral fragment.

After extracting MFCC features from the song as well as the short and labeling them, we built seven classification models to predict the viral song fragment. Among the proposed models, Weighted Logistic Regression and Weighted SVM Classifier gives the highest recall value (1.0) with a mediocre accuracy. Since the dataset is highly imbalanced, the recall had to be prioritized and hence the weighted algorithms worked the best.

VIII. Future Work

As future work, we plan to perform deeper analysis by exploring the spectogram and chromagram features. In addition, we are interested in analyzing and using the lyrics of the music for the modeling since lyrics can also be one of the reasons for the virality of a song fragment. We also plan to use better pre-trained models in the future which could fit significantly better on imbalanced dataset.

References

[1] Xinpeng Chen, Jingyuan Chen, Lin Ma, Jiebo Luo and Tong Zhang, "Finegrained Video Attractiveness Prediction Using Multimodal Deep Learning on a Large Realworld Dataset," The Web Conference (WWW), Lyon, France, April 2018.

[2] Devadas Menon, Factors influencing Instagram Reels usage behaviours: An examination of motives, contextual age and narcissism, Telematics and Informatics Reports, Volume 5, 2022, 100007, ISSN 2772-5030, https://doi.org/10.1016/j.teler.2022.100007.

[3] Araujo, Carlos & Cristo, Marco & Giusti, Rafael. (2019). Predicting Music Popularity on Streaming Platforms. 141-148. 10.5753/sbcm.2019.10436.

[4] Yekyung Kim, Bongwon Suh, and Kyogu Lee. 2014. #nowplaying the future billboard: mining music listening behaviors of twitter users for hit song prediction. In Proceedings of the first international workshop on Social media retrieval and analysis (SoMeRA '14). Association for Computing Machinery, New York, NY, USA, 51–56. https://doi.org/10.1145/2632188.2632206

[5] https://deepblue.lib.umich.edu/bitstream/handle/2027.42/172883/Stephanie

[6] J. F. Andry, S. A. Reynaldo, K. Christianto, F. S. Lee, J. Loisa and A. B. Manduro, "Algorithm of Trending Videos on YouTube Analysis using Classification, Association and Clustering," 2021 International Conference on Data and Software Engineering (ICoDSE), 2021, pp. 1-6, doi: 10.1109/ICoDSE53690.2021.9648486.

[7] Li Sun*, Haoqi Zhang*, Songyang Zhang, and Jiebo Luo, "Content-based Analysis of the Cultural Differences between TikTok and Douyin," Special Session on Intelligent Data Mining, IEEE International Conference on Big Data, Atlanta, GA, December 2020.

[8] Songyang Zhang, Tolga Aktas, and Jiebo Luo, "Mi YouTube es Su YouTube? Analyzing the Cultures using YouTube Thumbnails of Popular Videos," Special Session on Intelligent Data Mining, IEEE International Conference on Big Data, December 2021.

[9] Charles Parker, Dhiraj Joshi, Phoury Lei, and Jiebo Luo. 2011. Finding geographically representative music via social media. In Proceedings of the 1st international ACM workshop on Music information retrieval with user-centered and multimodal strategies (MIRUM '11). Association for Computing Machinery, New York, NY, USA, 27–32. https://doi.org/10.1145/2072529.2072538

[10] Montag Christian, Yang Haibo, Elhai Jon D., ”On the Psychology of TikTok Use: A First Glimpse From Empirical Findings”, Frontiers in Public Health, Volume 9, 2021, ISSN 2296-2565
[11] https://en.wikipedia.org/wiki/Dynamic_time_warping
[12] https://towardsdatascience.com/oversampling-and-undersampling-5e2bbaf56dcf .
[13] https://bootcamp.uxdesign.cc/youtube-shorts-efc97a02888e .