Spotify Music Popularity Prediction

Munukutla D. V. Kashyap  
 Department of Computer Science  
University of Illinois at Chicago  
 Chicago, Illinois, United States  
 dmunuk2@uic.edu

Sai Mahesh Vemulapalli  
 Department of Computer Science  
 University of Illinois at Chicago  
 Chicago, Illinois, United States  
 [svemul20@uic.edu](mailto:svemul20@uic.edu)

Naga Sumanth Choudary M.  
Department of Computer Science  
 University of Illinois at Chicago  
 Chicago, Illinois, United States  
 nmovva2@uic.edu

ABSTRACT

Identifying and understanding the estimated popularity of a track before it is shown to the public based on its features is an integral part of music production. Musicians depend on this to best plan their releases and tune their tracks better, and music platforms use this estimation to better cater to the trends on their platform. Spotify is one of these platforms, and it is also one of the leading platforms in the entire industry.

Our project aims to improve this part of the music industry, which involves estimating or predicting whether a music track would be popular or not in the near future on Spotify’s platform using the features of the track as the predictors. Existing works surrounding this topic often concentrate on the exact popularity index prediction, which may be obtuse to a newcomer. This project instead is focused on being more accessible and approaching the task from a classification standpoint using the acoustic features of the track. The Random Forest classifier was picked as the top choice from the tested models, with an accuracy of 78% when predicting whether the song would be popular.

INTRODUCTION

Music has existed since ancient times and has become an integral part of almost everyone’s life. Due to this immense demand and the internalization efforts in the last few centuries, the music industry has reached a staggering 28.29 billion USD in the year 2023 and estimated to grow to 42.62 billion USD by the year 2028, according to a report from Globe Newswire. And the people who create music are, by extension, numerous. Most of these artists rely on this craft to make a living. This means that it is crucial for them to understand the trends of highs and lows in the industry to make sure that they are ahead of the curve and would not release a track at the wrong time and risk a considerable amount of their work going unnoticed.

Out of all the platforms that listeners use to listen to music in all formats, Spotify occupies the top spot, with a 30.5% share of the entire music streaming industry, followed by Apple Music at a much lower 13.7%. Hence it is imperative to make sure that the artists are able to obtain an estimation of whether their track would be popular in the near future on Spotify’s platform, as it would account for the majority of their revenue on their work. Spotify uses a proprietary algorithm to assign a popularity index to all the tracks on their platform. This algorithm is highly complex and obscure, using attributes like the number of listens for the track and how recent the listens are in order to generate this popularity index. For an outside viewer, it is generally not possible to gauge the popularity trends on the platform without proper analysis. Here, our project enters the scene to bridge the gap by building and training a model that can understand the popularity trends and provide the artist with a reliable estimate of their track’s popularity on the platform in the upcoming weeks. On the other hand, music streaming platforms rely on popular tracks being shown to viewers in order to generate more and more traffic on their platforms. A report by MusicAlly estimates that 15% of the most popular tracks on Spotify account for 95% of the platform’s revenue in the year 2022. So it is crucial for the platform to identify these undiscovered diamonds in the rough as soon as possible to provide quality recommendations for the users.

CCS CONCEPTS

• Insert CCS text here • Insert CCS text here   • Insert CCS text here

KEYWORDS

XGBoost – Extreme Gradient Boosting

ACM Reference format:

FirstName Surname, FirstName Surname and FirstName Surname. 2018. Insert Your Title Here: Insert Subtitle Here. In *Proceedings of ACM Woodstock conference (WOODSTOCK’18). ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/1234567890

1 Methodology

1.1 Data Collection

The updated template, user manuals, samples, and required fonts, all are available at the URL <https://www.acm.org/publications/proceedings-template>. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in different clips. **Display Formula with Number**

 (1)

**Continuation part of Paragraph Text** The user must style this paragraph in **ParaContinue** style, which follows immediately after the **DisplayFormula** (numbered equation). The **DisplayFormula** style is applied only in case of a numbered equation. A numbered equation always has a number to its right. Insert paragraph text here. **Display Formula without Number**



The **DisplayFormulaUnnum** style is applied only in case of an unnumbered equation. An unnumbered display equation never contains an equation number to its right, and this unique property distinguishes it from a numbered equation.



Figure 1: Figure Caption and Image above the caption [In draft mode, Image will not appear on the screen]

**Theorem/Proof/Lemma.** Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement.

....Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract.

1.2 Data Cleaning

In the below paragraph, it is explained how alt-txt value is placed in **MS Word 2010**. To add alternative text to a picture in Word 2010, follow these steps:

1.3 Data Analysis

1.4 Feature Selection

1.5 Model Training

1.6 Model Tuning

1.7 Model Evaluation

2 Conclusion

1. In a Word 2010 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. Select the **Alt Txt** option from the left-side panel options.
4. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

Below are steps to place alt-txt value in **MS Word 2013/2016**. To add alternative text to a picture in Word 2013/2016, follow these steps:

1. In a Word 2013/2016 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. In the settings at the right side of the window, click on the "Layout & Properties" icon (3rd option).
4. Expand **Alt Txt** option.
5. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

*1.1.1 Heading Level 3.* Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

*1.1.1.1 Heading Level 4.*Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

ACKNOWLEDGMENTS

Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

REFERENCES

[1] Patricia S. Abril and Robert Plant, 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan, 2007), 36-44. DOI: <https://doi.org/>10.1145/1188913.1188915.

[2] Sten Andler. 1979. Predicate path expressions. In *Proceedings of the 6th. ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL '79)*. ACM Press, New York, NY, 226-236. DOI:https://doi.org/10.1145/567752.567774

[3] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. DOI:https://doi.org/10.1007/3-540-09237-4.

[4] David Kosiur. 2001. *Understanding Policy-Based Networking* (2nd. ed.). Wiley, New York, NY..

Conference Name:ACM Woodstock conference

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00