



Survey of TV Show Popularity Prediction Using Sentiment Analysis in Social Network

Monali Pawar¹, Dr. Avinash S. Kapse²

¹Post-Graduate Student, CSE, Anuradha Engineering College, Chikhli, India

²Associate Professor, CSE, Anuradha Engineering College, Chikhli, India

Abstract: Predicting the popularity of a TV show using sentiment analysis is a highly intriguing and demanding task. One important requirement is to forecast the popularity of online serials, which can facilitate several applications, including online advertising and serial suggestion. The problem at hand is that the success or failure of a program lies solely with the viewers. By identifying the most prevalent characteristics that viewers desire in a program and incorporating these requirements during the production phase using effective scientific methods, we can increase the likelihood of the proposed TV program being well-received. The objective of this study is to assess the efficacy of a TV show and determine the level of audience affinity towards specific programs or actors within the show. Additionally, the study aims to anticipate the popularity of these shows based on textual reviews. We are receiving feedback on social media platforms such as Twitter.

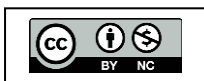
Keywords: Tasks, Application Prediction, Methodology, Popularity, etc.

I. INTRODUCTION

The topic of research that focuses on predicting the popularity of TV shows using sentiment analysis in social networks shows great promise. The process entails evaluating the emotional tone of social media posts pertaining to a specific television program and utilizing the findings to forecast its level of popularity. Sentiment analysis is a methodology that employs natural language processing (NLP) and machine learning algorithms to detect and extract subjective information from textual data. Within the realm of television programs, sentiment analysis can be employed to extract the thoughts, feelings, and attitudes conveyed by social media users regarding a specific show. Sentiment analysis can be employed to assess the sentiment of social media posts pertaining to a TV show, so enabling the prediction of its popularity. Sentiment can be categorized as positive, negative, or neutral. Subsequently, the quantity of affirmative and adverse attitudes might be juxtaposed to ascertain the overarching sentiment towards the television program.

Sentiment analysis can be integrated with social network data, including metrics like follower count, shares, and comments, to develop a predictive model for determining the popularity of a TV show. Training this model using past data can enhance its accuracy. In summary, the utilization of sentiment analysis in social networks to anticipate the popularity of TV shows has the capacity to offer significant insights to TV creators, advertisers, and broadcasters. It can assist individuals in comprehending the preferences of the audience and customizing their content to align with their expectations.

Reality TV has become the prevailing focus of television producers and channel executives. The word "the" is a definite article used in English grammar to indicate a specific noun or noun phrase. The





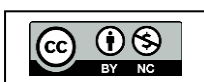
primary objective is to maximize TRP ratings. In the present day The majority of television programs consist of reality shows that focus on the genres of dance, singing, and acting. We propose the development of a system that can accurately identify and categorize individuals' emotional comments regarding television programs. The tweets pertaining to the specific show will be retrieved. The comments will be collected from several sources such as social networking platforms like Twitter. The rating of the TV show will be determined based on people's comments and its popularity. The system enables administrators to add text views, together with their corresponding likes-dislikes and sentiment comments.

II. RELATED WORK

Cammarano et. al. states that the use of social media has grown exponentially in recent years up to become a reflection of human social attitudes and to represent today the main channel for conducting discussions and sharing opinions. For this reason, the vast amount of information generated is often used for predicting outcomes of real-world events in different fields, including business, politics, and health, as well as in the entertainment industry. In this research, we focus on how data from Twitter can be used to predict ratings of a large set of TV shows regardless of their specific genre. Given a show, the idea is to exploit features concerning the prerelease hype on Twitter for rating predictions. We propose a novel machine learning-based approach to the genre-independent TV show popularity prediction problem.

We compared the performance of several well-known predictive methods, and as a result, we discovered that LSTM and Random Forest can predict the ratings in the USA entertainment market, with a low mean squared error of 0.058. Furthermore, we tested our model by using data of "never seen" shows, by deriving interesting results in terms of error rates. Finally, we compared performance against relevant solutions available in the literature, with discussions about challenges arising from the analysis of shows in different languages. Social media is changing the way and the pace with which people share their opinions, resulting in the creation of a huge data in a scale of seconds. [1].

Singh et. al. states that the television industry is a constantly evolving multi-billion dollar industry. With online streaming services such as Netflix and Amazon Prime, people have access to thousands of TV shows. The rating and reviews that the audience provides is the biggest indication of whether the show is successful or not. With such data available, we can find out what features the most successful shows have in common and the shows of which genre are likely to be more successful with the help of various Machine Learning techniques such as classification and clustering. Algorithms such as k-NN, SVM, Naïve Bayes, Decision Trees and Gradient Descent can be employed to build a model with high accuracy. With the worded reviews provided by the audience, we can also perform sentiment analysis using natural language processing to find out what the audience thinks about any particular show. Based on the predictions made by the model we can also make favorable recommendations to different demographics based on their interests. The disadvantage of the existing system architecture is that it explores only ratings and metadata but does not analyze what users have to say about particular media programs. Here, we argue that text comments are excellent indicators of user satisfaction. [2]





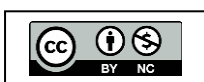
Phanshikar et. al. states that the activity of Social-TV viewers has considerably grown in the last few years. The web TV Show. Recommendation based on analysis is quite interesting and extremely challenging. Reality shows are growing day by day within today's generations. There are many exceptional ways to find the Television Rating Point (TRP). Users can be able to add ratings, suggestions, comment on them and review on TV shows. The web has been socially empowering many audiences in unique ways. Nowadays the rapid development of sharing websites, so many people would like to be audiences in their daily entertainments such as TV shows and websites, to attract major audiences many efforts have been taken for popularity prediction. Distinct episodes are released on specific days to are expecting the popularity of TV suggests to be tough so we introduce this scheme that predicts the TV shows primarily based on sentiment evaluation. Sentiment analysis algorithms provide social media surveillance as they give us an overview of the general public opinion behind some topics. The use of assumption examination can separate bits of knowledge from social information is a training that is in effect generally embraced by associations across the world. This work aims to gauge the performance of TV programs and conjointly calculate what number of folks are likable to a specific show or actors of that show and predict the quality of that show, supported by the text reviews. we tend to be becoming reviews on social networking websites. The main purpose is to find the ratings. Nowadays most of the TV shows are reality based on dancing, singing, and acting. [3]

Akula et. al. states that Television is an ever-evolving multi-billion-dollar industry. The success of a television show in an increasingly technological society is a vast multi-variable formula. The art of success is not just something that happens, but is studied, replicated, and applied. Hollywood can be unpredictable regarding success, as many movies and sitcoms that are hyped up and promise to be a hit end up being box office failures and complete disappointments. In current studies, linguistic exploration is being performed on the relationship between Television series and target community of viewers. Having a decision support system that can display sound and predictable results would be needed to build confidence in the investment of a new TV series. The models presented in this study use data to study and determine what makes a sitcom successful. In this research, we use descriptive and predictive modeling techniques to assess the continuing success of television comedies: The Office, Big Bang Theory, Arrested Development, Scrubs, and South Park. The factors that are tested for statistical significance on episode ratings are character presence, director, and writer. [4]

III. PROPOSED SYSTEM

A. Hashtag Datasets

A hashtag dataset refers to a compilation of social media posts that have been labeled with particular hashtags. These datasets have versatile applications, including sentiment analysis, trend analysis, and social network analysis. Hashtags are employed on popular social media platforms such as Twitter, Instagram, and Facebook to classify material and facilitate the discovery of postings pertaining to a specific subject. Hashtags are commonly indicated by the "#" symbol followed by a word or phrase, such as #travel, #foodie, or #fitness. In order to generate a collection of hashtags, social media posts are initially gathered through the application programming interface (API) of the site. Subsequently,





the posts are refined to solely encompass those that incorporate a particular hashtag or group of hashtags. Automating this process is possible by utilizing data collection programs such as Hootsuite, Brandwatch, or Socialbakers.

Hashtag datasets can be utilized to acquire valuable information about current patterns on social media, track references to a brand or product, and pinpoint influential individuals in a specific field. Additionally, they can be utilized for training machine learning models in tasks such as sentiment analysis or other natural language processing activities. It is crucial to acknowledge that hashtag datasets may not accurately reflect the entire population, as they solely consist of publicly published postings that have been tagged with a certain hashtag. Moreover, the data's quality may differ based on the origin and the applied filtering criteria. Hence, it is crucial to meticulously assess and authenticate the data prior to utilizing it for research or analytical objectives.

B. Tokenization

Tokenization is the act of dividing text into smaller pieces known as tokens, typically consisting of words or subwords. Tokenization is a crucial procedure in natural language processing (NLP) that transforms unprocessed text into a machine-readable format for analysis. Tokenization involves the division of text into separate tokens according to specific principles. These rules may encompass whitespace, punctuation marks, or other requirements. In English language tokenization, a sentence can be divided into distinct tokens by using a period, question mark, or exclamation mark. Tokenization can be performed at several levels, including word-level, subword-level, or character-level. Word-level tokenization involves dividing the text into separate words, whereas subword-level tokenization breaks down words into smaller units based on how often they appear.

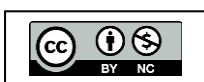
Character-level tokenization involves dividing the text into separate units based on specific characters. Tokenization is a crucial preprocessing stage in numerous NLP applications, including text classification, machine translation, and information retrieval. Segmenting the text into smaller pieces simplifies the data and enhances the efficiency of machine learning algorithms in analyzing the language. Various programming languages offer a range of tokenization tools, including NLTK and SpaCy in Python, Stanford CoreNLP in Java, and Apache OpenNLP in Java and .NET.

C. Normalization

Normalization in natural language processing (NLP) is a method employed to convert text input into a standardized format. The procedure entails the application of a set of regulations to transform text into a standardized format that can be efficiently handled and examined by machines.

Normalization generally consists of the following procedures:

- Case normalization involves converting all the letters in the text to either lowercase or uppercase. This is done to maintain consistency across the text. To prevent treating words with identical letters but varying cases as distinct tokens, this approach can be employed.
- Remove punctuation marks, such as periods, commas, and question marks, from the text as they are not meaningful in NLP analysis.





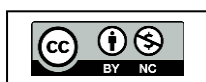
- Stop word removal involves eliminating frequent terms, known as stop words, that lack significant significance in a sentence, such as "the", "a", "an", and so on. These words can be omitted to decrease the amount of the textual material and concentrate on more significant words.
- Normalization might encompass stemming or lemmatization. Stemming involves reducing words to their root form, whereas lemmatization involves reducing words to their base form. In stemming, words like "running", "ran", and "runs" are reduced to their root form "run". In lemmatization, they are reduced to their base form "run".
- POS tagging: The system determines whether the word token is a noun, verb, or adjective. POS Tagging is the process of assigning a word a label based on its syntactic functions.

The objective of normalization is to guarantee uniformity in the textual data, minimize the data's size, and prioritize the most significant words. This can be beneficial for a range of Natural Language Processing (NLP) activities, including sentiment analysis, text categorization, and topic modeling. Various computer languages offer a wide range of libraries and tools, such as NLTK and SpaCy in Python, and CoreNLP in Java, which provide diverse normalization strategies.

IV. APPLY NLP

NLP, short for natural language processing, has diverse applications for analyzing, comprehending, and producing natural language text. Here are few typical scenarios where Natural Language Processing (NLP) might be utilized:

- Sentiment Analysis:** Natural Language Processing (NLP) can be utilized to ascertain the sentiment or emotional state conveyed in a given text, such as a review or a post on social media. Businesses can utilize this tool to oversee client input, while researchers can employ it to scrutinize public opinion.
- Named Entity Recognition (NER):** It is a technique in Natural Language Processing (NLP) that can be employed to detect and extract specific entities, such as individuals, locations, companies, and dates, from a given text. These techniques can be valuable for extracting information, classifying documents, and indexing search results.
- Text Classification:** NLP can be employed to categorize text into distinct classes or subjects, such as news articles or consumer comments. This can be advantageous for the automation of document categorization or for the surveillance of social media trends.
- Machine Translation:** NLP enables the automatic translation of text from one language to another. This feature can be advantageous for multinational corporations or for facilitating interlingual communication.
- Question Answering:** Natural Language Processing (NLP) can be employed to automatically provide answers to questions asked in a natural language format. This can be advantageous for chatbots, virtual assistants, or search engines.
- Content Generation:** NLP can be employed to produce coherent and fluent content in natural language, such as automatically created product descriptions or tailored recommendations.





This has practical applications in e-commerce, content generation, and marketing. Various computer languages, including Python, Java, and R, offer a wide range of NLP frameworks and tools that can be utilized to achieve various use cases. Notable NLP libraries comprise NLTK, SpaCy, CoreNLP, and Gensim.

V. SENTIMENT ANALYSIS

Sentiment analysis is an NLP technique that entails examining a text to ascertain the underlying sentiment or mood. Sentiment analysis is a technique that can be employed to determine the general sentiment of a text, categorizing it as either good, negative, or neutral. Sentiment analysis can be approached using several methods, such as rule-based and machine learning-based approaches. Sentiment analysis in rule-based systems relies on a predetermined set of rules and dictionaries to detect words and phrases that convey sentiment. Sentiment analysis in machine learning relies on a trained model that uses labeled data to predict the sentiment of fresh text.

Sentiment analysis has broad applications across several businesses and scenarios, such as:

- Social media monitoring is the use of sentiment analysis to assess public sentiment, evaluate brand perception, and analyze customer input on various social media platforms.
- Customer service: Sentiment analysis can be utilized to automatically detect and rank customer complaints and comments according to their emotional tone.

Sentiment analysis can be employed in product research and development to examine client feedback regarding products, thereby identifying potential areas for enhancement and new product prospects.

Political analysis: It involves the application of sentiment analysis to examine political speeches, news stories, and social media posts in order to get insights into public opinion and political attitude.

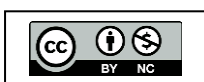
Market research: Sentiment analysis can be employed to examine customer feedback and reviews in order to comprehend consumer preferences and behavior.

Various sentiment analysis tools and packages are accessible in diverse programming languages, including Python, Java, and R. Notable sentiment analysis tools include NLTK, SpaCy, TextBlob, and VADER.

VI. ANALYSIS OF TV POPULARITY

TV popularity prediction by sentiment analysis in social networks entails the examination of social media data to forecast the level of popularity of a television program. Here is a potential methodology for examining this issue:

- **Data Collection:** Gather social media data pertaining to television programs, including tweets, comments, and likes. The collection of this data can be accomplished through the utilization of site scraping tools or social media application programming interfaces (APIs).
- **Data Pre-Processing:** The acquired data should be pre-processed to eliminate any unwanted noise, such as spam or irrelevant data. Additionally, normalization techniques should be applied to standardize the text data. This process may entail the elimination of stop words, punctuation marks, and special characters, as well as the conversion of the text to lowercase.



- **Feature Extraction:** Extract aspects from the social media data that are indicative of the popularity of a TV show. This task may entail identifying keywords or subjects that are linked to famous television programs.
- **Model Training:** Utilize the retrieved features and sentiment data to train a machine learning model that can accurately predict the popularity of TV shows. This can entail utilizing regression or classification algorithms to forecast the popularity of a television show by considering the sentiment and other characteristics.
- **Model Evaluation:** Assess the effectiveness of the machine learning model by measuring measures such as accuracy, precision, recall, and F1-score. This can entail employing cross-validation procedures to confirm the model's generalizability and prevent overfitting of the data.
- **Prediction and Visualization:** Utilize the trained model to forecast the popularity of upcoming TV shows and employ charts or graphs to visually represent the outcomes, facilitating informed decision-making.

In summary, the process of predicting TV popularity using sentiment analysis in social networks includes gathering and analyzing social media data, conducting sentiment analysis, extracting relevant characteristics, training a machine learning model, and assessing the model's effectiveness. The efficacy and utility of the model will be contingent upon the caliber and volume of the data, the extracted features, and the selection of the machine learning algorithm.

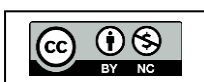


Figure 4.4: Popularity Percentage on the Basis of Social Tweets

The result showing when we select the TV shows and it depends upon the recent tweets that was generated when the reviews was given



Figure 4.5: Another Popularity Percentage on the Basis of Social Tweets





The result showing when we select the TV shows and it depends upon the recent tweets that was generated when the reviews was given and predicted the percentage of its polarity whether it is negative or positive

VII. CONCLUSIONS

Here we present a system TV Show Popularity Prediction using Sentiment Analysis in Social Network for users, which predict the popularity of the show among several shows, actors, reality shows and serials based on based on the text reviews which are getting from social networking websites like Twitter. The advantages of using this system are that it helps in analyzing TV Show details and helps to rate prediction based on Twitter tweets. To estimate the popularity of TV series that premiere on different days, we developed a sentiment analysis-based prediction technique. Sentiment Analysis algorithms can monitor social media and provide insights into public opinion on certain issues. Uses of assumption Examining knowledge to distinguish it from social information is a common training practice across organizations worldwide.

ACKNOWLEDGMENT

We are thankful for the institutions for allowing us to submit a paper and interested to contribute and to evaluate the result of students which helps for the performance of the students in many ways.

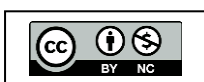
DECLARATION

We hereby declared that we cited the suthors work and try to improvised the work done by many researcher in the field of education for its betterment.

| Decalarion | Suggestions |
|--|---|
| Availability of Data and Material / Data Access Statement | We make the use of existing data of review from various TV and social viewwers. |
| Authors Contributions | A new filtering system for tv publkcicity was designed to understand the viewers reaction and making the prediction of popularity of shows. |

REFERENCES

- [1] Maria Elena Cammarano, Alfonso Guarino, Delfina Malandrino, Rocco Zaccagnino1, "TV shows popularity prediction of genre-independent TV series through machine learning-based approaches", Multimedia Tools and Applications <https://doi.org/10.1007/s11042-024-18518-z>, 2024
- [2] Rashmi Singh, Srushti Nemade, Atharv Pillai, Binoy Vijaykumar, and Prof. Gayatri Hegde, " TV Show Popularity Analysis Rashmi Singh, Srushti Nemade, Atharv Pillai, Binoy Vijaykumar, and Prof. Gayatri Hegde", 2021 IJCRT, Volume 9, Issue 4 April 2021, ISSN: 2320-2882
- [3] Ms. Medha Pradeep Phanshikar, Mr.D.A.Patil, "Research Paper on TV Show Popularity Analysis", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 08 Issue: 12 | Dec 2021 p-ISSN: 2395-0072





- [4] Ramya Akula, Zachary Wieselthier, Laura Martin, Ivan Garibay, "Forecasting the Success of Television Series using Machine Learning", arXiv:1910.12589v1 [cs.LG] 18 Oct 2019
- [5] Joshua M. Mutisya, "Sentiment analysis for TV show popularity prediction: case of Nation Media Group's NTV", 2019, Strathmore University Library
- [6] T. Jeon, 1. Cho, S. Lee, G. Baek, and S. Kim, "A movie rating prediction system of user propensity analysis based on collaborative filtering and fuzzy system," in Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on. IEEE, 2009, pp. 507-511.
- [7] O. Bora Fikir, İ lker O. Yaz, Tansel Özyer 2010 International Conference on Advances in Social Networks Analysis and Mining "A Movie Rating Prediction Algorithm with Collaborative Filtering" on. IEEE, 2009, pp. 507-511.
- [8] Jun Ai, Linzhi Li, Zhan Su, Chunxue Wu "Online-rating prediction based on an improved opinion spreading approach" 2017 29th Chinese Control And Decision Conference (CCDC)
- [9] G. Adomavicius and A. Tuzhilin, "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Transactions on Knowledge and Data Engineering 17 (2005), 634-749.
- [10] Jun Ai, Linzi Li, Zhan Su, Chunxue Wu "Online-rating prediction based on an improved opinion spreading approach" 2017 29th Chinese Control and Decision Conference (CCDC)
- [11] Fanglin Wang, Daguang Li, Mingliang Xu, "A location-aware TV show recommendation with localized semantic analysis", DOI 10.1007/s00530-015-0451-z, Springer-Verlag Berlin Heidelberg 2015.
- [12] Jinhoh Oh a, Sungchul Kim b, Jinha Kim b, Hwanjo Yu, "When to recommend: A new issue on TV show recommendation", 2014 Published by Elsevier Inc.
- [13] Li Bian, "Online Friend Recommendation through Personality Matching and Collaborative Filtering", Copyright (c) IARIA, 2011. ISBN: 978-1-61208-171-7
- [14] Tomoko Murakami, Koichiro Mori, and Ryohei Orihara, "Metrics for Evaluating the Serendipity of Recommendation Lists", c Springer-Verlag Berlin Heidelberg 2008.
- [15] L. Lü, D.-B. Chen, and T. Zhou, "The small world yields the most effective information spreading," New Journal of Physics, vol. 13, no. 12, p. 123005, 2011.

