IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124211

YouTube Trending Videos' Prediction & Analysis

Vibhas Meshram[1], Vishal Gaikwad[2], Vaishali Pathak[3], Ankita Mohite[4], Prof. Rama Barwal[5]

Computer Science Department, Savitribai Phule Pune University, Genba Sopanrao Moze College of Engineering, Balewadi, Pune, Maharashtra, India[1-5]

Abstract: Social media platforms have become a crucial part of our daily lives and play significant roles in various aspects such as business, entertainment, marketing, education, media, and communication. Among these platforms, YouTube has gained massive popularity as the most widely used platform for sharing videos due to its unique behavior. The platform allows anyone to create an account and upload videos of their choice, which can be viewed by millions of people. This has become a trend in the entertainment industry, making it easy for videos to reach users and gain popularity online. However, not all YouTube videos become popular, and many channel owners take various actions to make their videos popular.

This research study uses sentiment analysis and feature extraction methods to derive the set of features required to consider in the development of YouTube videos. By analyzing user comments, the study aims to discover the most important trending videos related to user video types and the most trending videos that users will want to see.
The study uses machine learning methods to analyze the trending features and identify key recommendations for the users. The results of the study will enable YouTubers to create videos that resonate with their audience and increase their chances of being popular.

The study on YouTube trending videos and Support Vector Machine (SVM) algorithm has revealed the importance of views, likes, and dislikes in determining a video's trend. The SVM algorithm uses these factors to identify and predict which videos will become popular on the platform. This research study provides recommendations to YouTubers on how to create videos that can become trending by analyzing user comments, identifying their requirements, and using machine learning methods to derive the necessary features.

Keywords: KNN (K-Nearest Neighbors), SVM (Support Vector Machine), HDFS (Hadoop Distributed File System), CNN (Convolutional Neural Network), Feature Extraction methods, Machine Learning, MLP (Multi-Layer Perception), popularity

I. INTRODUCTION

YouTube is a widely popular video-sharing platform where users can upload, watch, and share videos. The platform features a section called "Trending" that displays videos that are currently popular among users. The videos in the Trending section are selected based on several factors such as view count, likes, comments, and shares, and the section is updated periodically.

Being featured in the Trending section can provide significant exposure for the videos and the creators' channels, potentially resulting in more views, likes, and subscribers. Thus, predicting trends on YouTube has become crucial for content creators and marketers to optimize their content to reach a broader audience. By identifying the topics and formats that are likely to become popular, creators can produce videos that have a higher chance of being featured in the Trending section. Similarly, marketers can use trend prediction to identify opportunities for brand integration and sponsorship. By partnering with creators whose content aligns with their brand, marketers can leverage the creators' reach and influence to promote their products or services.

For viewers, trend prediction can help them discover new content that aligns with their interests and preferences. By exploring the videos in the Trending section, viewers can discover popular and engaging content that they might have otherwise missed. Students, in particular, use YouTube as a resource to supplement their coursework and locate additional content. The popularity of videos on YouTube is determined by various factors, including the views, likes, dislikes, and comments received.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023
DOI: 10.17148/IJARCCE.2023.1242

The YouTube algorithm uses these factors to determine which videos can be trending and featured in the Trending section. Researchers have used data mining software to analyse these factors and determine their impact on a video's popularity. By understanding how the algorithm works and which factors are important, content creators can optimize their videos to increase their chances of being featured in the Trending section and attracting more views and engagement.

Overall, YouTube has become a valuable resource for users seeking a wide range of video content, and understanding the factors that contribute to a video's popularity can help creators and viewers alike.

II. LITERATURE REVIEW

Various studies have delved into the utilization of the morphological sentence pattern model to perform sentiment analysis on social media platforms. The significance of sentiment analysis in comprehending the general public's opinion and responses on platforms like YouTube has been emphasized in works by Han and Kim [1] and Pozzi et al. [4].

In [2], Chen investigates the concept of personal branding on YouTube and its potential impact on user engagement and viewership. The research offers valuable insights into the key factors that contribute to the success of personal branding on YouTube and how it affects user behavior. Several studies, including those by Li [3] and Mohana Prabha et al. [6], have examined the factors that impact the popularity of online videos on platforms like YouTube. To predict the popularity of these videos, sentiment analysis is utilized to evaluate the comments and feedback received.

In their research, Soleymani and colleagues [7] conducted an in-depth exploration of multimodal sentiment analysis. This approach involves combining multiple types of data, including text, audio, and visual cues, to gain a more comprehensive understanding of user sentiment. The researchers identified various potential applications for this technique, including online advertising and improving user experience.

Rangaswami and colleagues [8] investigate the application of sentiment analysis in extracting metadata and categorizing YouTube videos based on their sentiment. Their proposed framework involves utilizing sentiment analysis for extracting metadata from YouTube videos, and subsequently classifying them into appropriate categories based on their sentiment.

In another study, Mulholland et al. [9] utilize sentiment analysis techniques to examine emotional sentiment expressed in user comments on YouTube channels. This study highlights the significance of sentiment analysis in comprehending users' emotional reactions and their impact on user engagement.

Furthermore, Chelaru and colleagues [10] explore the efficacy of incorporating social feedback into the learning process for ranking YouTube videos. Their proposed method involves integrating social feedback and demonstrating its effectiveness in enhancing the accuracy of YouTube video ranking.

In summary, the literature review suggests that sentiment analysis is a crucial tool for analyzing user behavior and feedback on social media platforms such as YouTube. It has been utilized in predicting video popularity, examining personal branding strategies, and enhancing user experience. The integration of diverse data sources and the implementation of machine learning techniques have expanded the potential applications of sentiment analysis in various domains.

TABLE I. Literature Review

Year	Author	Paper Name	Algorithms Used	Acc. %
2022	S.O. Olukumoro; C. A Adenus; E I Ofoegbunam	Prognosticate Trending Days of YouTube Videos Tags Using KNN Algorithm	KNN, Random Forest, Grid Search	81.7 86.5 78.6
2021	H. M. Caldera; Suresha Perera; G. S. Meedin; Indika Perera	Classification of Trending Videos in YouTube	Decision Tree, Random Forest, Navies Bayes, SVM	84.1, 85.6, 82.7, 83.6
2022	Arushi Pathik; S. Patni; Vaibhav Patel; Jash Patel; Artika Sing	YouTube Trend Analysis	Linear Regression, Decision Tree, Random Forest	82-88.8

DOI: 10.17148/IJARCCE.2023.1242

2018	Farzana Shaik; Danish Pawaskar; Abutalib Siddiqui;	YouTube Data Analysis using MapReduce on Hadoop	Hive, Apache Mahout, HDFS, MapReduce Algorithms	82.5 83.5, 86.8, 84.5
2022	RSangeetha,R.Vidhya	Prediction of YouTube View Count using Supervised and Ensemble ML Techniques	SVM, Linear Regression, Gradient Boosting, KNN	86.4 81.3 79.7, 86.9
2018	Nafis Irtiza Tripto; Mohammed Ali	Detecting Multilabel Sentiment and Emotions from Bangla YouTube Comments	SVM, CNN, MLP, Random Forest	83 87.4 72.8, 86.5
2018	Jihyeon Lee; Hayoung Oh	YouTube aware Personalized Ranking System for Future ICT Education	Decision trees, Random forests, Neural Network	84.7 92.5 82.4

III. PROPOSED METHODS

The dataset used in YouTube trend prediction studies typically includes information on a large number of YouTube videos, such as their titles, descriptions, view counts, likes and dislikes, comments, and other engagement metrics. The dataset may also include information on the videos' upload dates, categories, and tags.

Some studies also incorporate data on external factors that may impact video trends, such as news events or social media trends. The dataset may be obtained from YouTube's API or web scraping tools and may be pre-processed to remove duplicate or irrelevant data points, as well as to transform the data into a format suitable for analysis using machine learning algorithms.

The size of the dataset can vary depending on the scope of the study, but it is typically quite large, often containing tens of thousands or even millions of video data points. The dataset may also be divided into training and testing sets, with the former used to train the machine learning models and the latter used to evaluate their performance

DATA PREPROCESSING AND FEATURE ENGINEERING:
Data preprocessing involves cleaning and transforming raw data to make it suitable for analysis. In the context of YouTube trend prediction, this may involve removing irrelevant data points or filling in missing values.

Feature engineering refers to the process of selecting or creating variables, or features, that will be used to train a machine learning model. In the context of YouTube trend prediction, some common features that may be used include video title, description, tags, view count, likes and dislikes, and comments.

Data preprocessing and feature engineering methods vary depending on the specific research question and data set being used. Some common techniques used in previous studies on YouTube trend prediction include:

1. Data collection: Collect information on employee attrition, demographics, job satisfaction, pay, and other pertinent variables. This can be accomplished using surveys, interviews, or information already in the company.
2. Data analysis: Analyze the data using Machine learning models such as Random Forest, SVC, K Nearest Neighbors, and Decision tree.
3. Model development: Develop a model to predict employee attrition based on the findings from the data analysis.
4. Validation: Validate the model by using a holdout sample or cross-validation techniques.
5. Recommendations: Based on the results of the study, provide recommendations for reducing employee attrition and improving employee retention in the organization.

It's important to note that the specific methods used in an employee attrition research study may vary based on the research design, data sources, and specific objectives of the study. However, the above steps provide a general overview of the process involved in conducting employee attrition research.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023
DOI: 10.17148/IJARCCE.2023.1242

IV. RESULTS

The result of this model, YouTube Trending Videos' Prediction & Analysis would be the ability to predict which videos are likely to trend on YouTube and analyze the factors contributing to a video's success. This could be achieved by developing a machine learning model that analyzes various features of a video, such as a title, description, tags, category, view count, like count, dislike count, comment count, and publishing time, to predict its likelihood of trending. The model could also be used to analyze the impact of these features on a video's success and identify patterns and trends that can inform content creators and marketers in creating and promoting their videos. Overall, the result of this model would be to provide valuable insights and predictions that can help content creators and marketers optimize their YouTube video content and increase their chances of success on the platform.

V. CONCLUSION

The YouTube trending videos Support Vector Machine (SVM) algorithm constantly includes the quantity of likes, audience comments, and views when identifying a trending video. The impact of these factors on a video's ability to trend has been studied extensively. Research has shown that videos with more views, likes, and comments are more likely to be featured in the Trending section, which in turn increases their exposure to a wider audience. This phenomenon has made it essential for content creators to understand the factors that contribute to a video's popularity and optimize their content accordingly. However, the most important factors that determine a video's popularity on YouTube are the number of views, likes, and comments. By analyzing these metrics, content creators can determine the type of content that resonates with their audience and optimize their videos to increase their chances of being featured in the Trending section.

The SVM algorithm used by YouTube is constantly evolving to ensure that the most popular and engaging videos are featured in the Trending section. This algorithm considers not only the number of views, likes, and comments but also the engagement rate, watch time, and audience retention. By understanding how the algorithm works and which factors are important, content creators can optimize their videos to increase their chances of being featured in the Trending section and attracting more views and engagement. In conclusion, understanding the factors that contribute to a video's popularity on YouTube is crucial for content creators looking to increase their exposure and reach a wider audience. The number of views, likes, and comments are the most important factors that determine a video's popularity, and optimizing these metrics can increase a video's chances of being featured in the Trending section. The SVM algorithm used by YouTube is constantly evolving to ensure that the most engaging videos are featured in the Trending section, making it essential for content creators to stay up-to-date with the latest trends and optimization techniques.

REFERENCES

[1] Han and Kim (2017) proposed a morphological sentence pattern model for sentiment analysis on social media, which can be applied to analyze sentiment in YouTube comments.
[2] Chen (2013) studied personal branding on YouTube, providing insights on how content creators can build and maintain their brand image to attract and retain viewers, 12:332-347.
[3] Li (2016) characterized and predicted the popularity of online videos, including YouTube videos, using various features such as social network metrics and video content attributes, 1630-1641.
[4] Pozzi et al. (2016) conducted sentiment analysis on social networks, including YouTube, and investigated the effectiveness of different sentiment analysis techniques, pp. 1-6.
[5] Tripto and Ali (2018) presented a method for detecting multilabel sentiment and emotions from Bangla YouTube comments, which can be applied to other languages as well.
[6] Prabha et al. (2019) proposed a sentimental analysis-based approach for predicting the popularity of trending YouTube videos, achieving high accuracy in their experiments, ISSN: 2278-3075, Volume-8, Issue6S3.
[7] Soleymani et al. (2017) conducted a survey of multimodal sentiment analysis techniques, which can be applied to analyze various types of data including text, image, and audio, 3-14.
[8] Rangaswamy et al. (2016) proposed a metadata extraction and classification approach for YouTube videos, which uses sentiment analysis as one of the classification features.
[9] Mulholland et al. (2016) analyzed emotional sentiment in YouTube channel comments, investigating how emotions are expressed and perceived by viewers, 181-188.
[10] Chelaru et al. (2013) investigated the usefulness of social feedback for learning to rank YouTube videos, which can be used to improve the ranking and recommendation of videos on the platform, pp. 1-29.
[11] Table I in this literature review summarizes the accuracy and performance of different models proposed in the previous studies, providing a comparison between the different approaches.