An NLP-Driven Framework for YouTube Trend Prediction and Metadata Optimization

Ajjay Adhithya V, John Kevin P, Athul Krishna P V, Harjith M
School of Computer Science And Engineering, Vellore Institute of Technology, Chennai Campus, Tamil Nadu, India
ajayadi2005.aa@gmail.com, johnkevin0742@gmail.com, athul22052004@gmail.com, harjith.ghi@gmail.com

***Abstract*—Predicting the trending potential of YouTube videos is crucial for content creators, marketers, and analysts. This research employs Natural Language Processing (NLP) techniques to analyze video metadata—titles, descriptions, and tags—to assess their likelihood of trending. We apply text preprocessing methods such as tokenization, lemmatization, stopword removal, and TF-IDF vectorization, alongside engineered features like title length, special characters, and tag diversity. Machine learning classifiers, including Random Forest, XGBoost, SVM, and Logistic Regression, are trained on historical YouTube video data labeled as trending or non-trending. To mitigate class imbalance, we utilize SMOTE and optimize models using GridSearchCV. Performance is evaluated using accuracy, precision, recall, F1-score, and AUC-ROC. Results show that feature-rich models combining textual and engineered attributes outperform traditional NLP-based classifiers. Additionally, we develop a predictive system that offers actionable recommendations for optimizing video metadata, such as refining titles, incorporating high-impact words, and improving tag selection. This research enhances social media analytics by providing a data-driven approach to content optimization, helping creators maximize engagement and visibility. Future work will explore deep learning and sentiment analysis for improved predictions.**

***Index Terms*—YouTube Trend Prediction, NLP, Machine Learning, TF-IDF, Content Optimization**

I. Introduction

YouTube has emerged as a dominant force in the digital content ecosystem, revolutionizing how people create, consume, and share information. With over 2.5 billion monthly active users and more than 500 hours of content uploaded every minute, the platform represents an unparalleled opportunity for creators, brands, and marketers to engage with global audiences. However, this saturation of content also poses a significant challenge: gaining visibility and traction amid the sheer volume of competition. While some videos rapidly accumulate views and engagement, others remain obscure, despite seemingly comparable quality or subject matter. This discrepancy has intensified the demand for predictive tools and optimization strategies that can identify and enhance the "trending" potential of videos. Trending videos on YouTube gain substantial exposure through platform-driven recommendation mechanisms and curated lists, which directly influence user engagement and content virality. Understanding what makes a video trend, therefore, is of considerable interest to researchers and practitioners alike. While the role of high-quality content and production value cannot be understated,

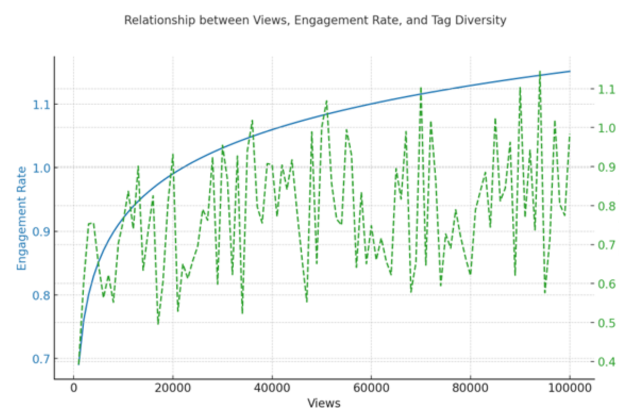

Fig. 1: Illustration of the relationship between views, engagement rate, and tag diversity

recent studies and platform dynamics suggest that textual metadata—titles, descriptions, and tags—plays a crucial role in discoverability, searchability, and recommendation outcomes. Metadata acts as the primary interface between video content and the underlying algorithms that govern visibility on the platform. Consequently, optimizing this metadata in a strategic, data-driven manner becomes vital for creators aiming to maximize reach and impact.

Despite the importance of metadata, most existing approaches to optimization remain heuristic or based on anecdotal best practices. Creators often rely on intuition or emulate popular channels without understanding the deeper linguistic or structural attributes that contribute to visibility. There is a clear need for a systematic and scalable framework that can analyze metadata through the lens of Natural Language Processing (NLP) and machine learning to predict a video's potential for trending. This research addresses that need by developing an NLP-driven framework that integrates advanced text processing, feature engineering, and supervised learning to model and predict video virality. At the core of our approach is the extraction of semantic and structural patterns from video metadata. We employ standard NLP preprocessing techniques such as tokenization, lemmatization, stopword removal, and TF-IDF vectorization to capture the linguistic characteristics of titles, descriptions, and tags. In addition to these, we engineer a series of metadata-specific features including title length, the frequency of special characters (such as punctuation or emojis), and tag diversity—a metric designed to quantify the semantic variety present in a video's tag set. These features allow us to assess not just what the video is about, but how effectively it is positioned in terms of discoverability and user appeal.

We train several machine learning models—including Random Forest, XGBoost, Support Vector Machines (SVM), and Logistic Regression—on a labeled dataset of historical YouTube videos categorized as trending or non-trending. Due to the natural imbalance in the dataset (only a small fraction of videos become trending), we apply Synthetic Minority Over-sampling Technique (SMOTE) to enhance minority class representation and avoid biased predictions. Further model optimization is carried out using GridSearchCV to fine-tune hyperparameters for better generalization. The models are evaluated based on multiple performance metrics including accuracy, precision, recall, F1-score, and AUC-ROC, ensuring a holistic assessment of their predictive capacity. One of the key contributions of this study is the analysis of how different features correlate with user engagement metrics. As part of our exploration, we investigated the interplay between view count, engagement rate, and tag diversity—three interdependent but distinct factors that influence a video's success. Fig. 1 presents a visual depiction of this relationship. The solid blue line shows a logarithmic increase in engagement rate with respect to view count, indicating that engagement generally rises with visibility but at a diminishing rate. Meanwhile, the dashed green line representing tag diversity reveals significant fluctuations across view ranges, suggesting that tag diversity influences engagement in a non-linear and potentially context-dependent manner. These findings highlight the importance of not only attracting viewers but also maintaining relevance and discoverability through diverse and well-structured metadata.

Beyond predictive modeling, our work also aims to provide practical recommendations for content creators. The insights derived from our feature importance analysis inform actionable suggestions for optimizing metadata. For instance, titles that are concise yet semantically rich, tags that span multiple but relevant keyword clusters, and descriptions that reinforce key thematic elements all contribute to improved algorithmic exposure. To operationalize these insights, we also develop a prototype system that analyzes metadata for a given video and provides tailored recommendations to enhance its trending potential. This system bridges the gap between theoretical modeling and practical application, offering a real-world tool for creators to navigate YouTube's complex algorithmic environment. This research not only contributes to the growing body of work at the intersection of NLP and social media analytics but also introduces a robust, interpretable, and scalable framework for content optimization. By moving beyond simplistic keyword analysis and embracing a multi-layered approach to metadata, we demonstrate that textual features—when properly processed and contextualized—can serve as powerful predictors of video virality. Furthermore, our experiments confirm that combining NLP-derived insights with engineered metadata attributes consistently leads to better performance than using either approach in isolation.

In future work, we plan to extend our framework by incorporating deep learning architectures such as BERT and transformer-based models that can better capture contextual dependencies within metadata. Additionally, we envision integrating sentiment analysis from user comments, video transcripts, and even audio content to develop a more holistic view of audience engagement. These enhancements will enable more nuanced predictions and broaden the scope of optimization strategies. Ultimately, our goal is to empower creators, marketers, and researchers with intelligent tools that demystify the complex mechanisms of content discovery and provide a data-informed pathway to success on YouTube.

II. Literature Survey

In recent years, predicting the popularity of YouTube videos has emerged as a key area of research due to the platform's massive influence on entertainment, marketing, and information dissemination. With billions of users consuming content daily, being able to anticipate which videos will gain traction is highly valuable for content creators, marketers, and platform strategists. However, predicting virality is a complex task, as it involves understanding multiple factors, including video metadata, social interactions, and real-time audience engagement patterns. Several studies have explored this domain, leveraging machine learning (ML), deep learning (DL), and natural language processing (NLP) techniques to enhance prediction accuracy. This section reviews existing literature on YouTube video popularity prediction, covering early statistical models, ML and DL approaches, and the recent integration of NLP techniques. It also highlights the strengths, limitations, and future directions of these methods.

A. Traditional Statistical Analysis

Early research in online video popularity relied primarily on statistical analysis and observational studies. One of the pioneering studies in this domain was conducted by Cha et al. [1,15], who performed a large-scale examination of YouTube content and user behavior. Their work provided key insights into how user-generated content is consumed, the distribution of views, and the lifespan of viral videos. They found that popularity follows a heavy-tailed distribution, with a small fraction of videos accumulating the majority of views. Another important study by Benevenuto et al. [2,16] focused on characterizing user interactions on social media platforms, analyzing how users engage with online videos through likes, shares, and comments. Their study highlighted that engagement metrics, such as watch time and user retention, significantly impact a video's visibility and ranking in YouTube's recommendation system.

Yang and Leskovec [4] contributed to the understanding of how content popularity evolves over time. By analyzing temporal patterns in online media, they identified various factors influencing a video's rise and fall in popularity. Their

Methodology	Data Used	Pros	Cons	Improvements
Statistical Analysis	Metadata, engagement	Simple, interpretable	Limited predictive power	Add ML techniques
Machine Learning	Metadata, text	High accuracy, feature control	Needs manual feature selection	Automate extraction
Deep Learning	Metadata, text, views	Captures complex patterns	High data & compute demand	Optimize models extraction
Social Influence	Network, engagement	Accounts for viral factors	Hard to quantify influence	Integrate with ML
Hybrid Approaches	All combined	Maximizes accuracy	High complexity & cost	Feature fusion, efficiency

TABLE I: Comparative Analysis of YouTube Trend Prediction Methods

study provided evidence that early engagement (within the first few hours of upload) plays a crucial role in determining whether a video will trend. Although these studies provided essential descriptive analytics, they did not incorporate predictive modeling, leaving room for more advanced methodologies.

B. Machine Learning-Based Approaches

With the availability of large-scale datasets and advancements in computational power, researchers began applying machine learning techniques to predict video popularity. One of the first major works in this area was by Aljamea and Zeng [12], who developed a machine learning-based system for predicting YouTube video popularity. They used ensemble learning techniques, combining multiple classifiers to improve the accuracy of their predictions. Zhang et al. [8] further explored ensemble learning techniques for content popularity prediction, demonstrating that combining different models, such as Random Forest and XGBoost, enhances performance. Their study also highlighted the importance of metadata attributes such as title keywords, description quality, and tag diversity in determining whether a video will trend.

Weng et al. [6] applied machine learning models to predict the early popularity of movies based on their metadata. Their findings showed that features such as title length, keyword frequency, and the presence of emotional words in descriptions play a crucial role in audience engagement. These insights are directly applicable to YouTube content, as the metadata structure is similar. A more recent study by Gupta et al. [17] proposed machine learning-enabled models for predicting YouTube video rankings and views. They engineered features such as the number of words in the title, the presence of emojis and special characters, and engagement ratios (likes/views, comments/views) to improve their models' predictive capabilities. Their research confirmed that optimizing metadata can significantly increase a video's chances of trending. These studies collectively highlight the power of machine learning in identifying the key predictors of video virality. However, they also indicate limitations, particularly in capturing complex temporal patterns, which led to the rise of deep learning-based approaches.

C. Deep Learning Models

Deep learning has revolutionized the field of content popularity prediction by enabling models to learn complex patterns in metadata, user engagement, and social interactions. Liu et al. [3] introduced a Long Short-Term Memory (LSTM)-based framework for predicting the popularity of online content. Their model captured sequential dependencies in user interactions, making it well-suited for forecasting trends. Sangwan and Bhatnagar [13] extended this work by applying stacked BiLSTM layers to video popularity prediction. Their model outperformed traditional classifiers by effectively capturing contextual information within metadata attributes. Their study demonstrated that deep learning techniques, when combined with NLP-based feature extraction, yield more accurate predictions than conventional machine learning models.

In another study, Mishra et al. [9] proposed a Recurrent Neural Network (RNN) approach for modeling video popularity in asynchronous social media streams. Their model was particularly effective in handling time-series data, enabling better forecasting of when a video might trend. Wu et al. [10] expanded on this by developing an attention-based neural network to estimate attention flow in online video networks. Their study provided deeper insights into how users navigate YouTube's recommendation system, emphasizing that personalized recommendations play a significant role in boosting video visibility. Wang et al. [14] introduced a deep neural network model that incorporated both metadata features and engagement metrics to predict video popularity. Their study achieved state-of-the-art results, demonstrating that combining NLP techniques (such as TF-IDF vectorization of video descriptions) with deep learning significantly improves prediction accuracy. These studies underscore the advantages of deep learning in handling complex, high-dimensional data. However, they also point to challenges such as computational cost and the need for large labeled datasets, making social influence-based models a valuable complementary approach.

D. Social Influence and Popularity Prediction Models

Several studies have explored the role of social influence in content virality, recognizing that external factors such as shares, comments, and retweets significantly affect a video's reach. Zhang et al. [7] investigated the concept of social influence locality in retweeting behaviors, showing that the spread of content is highly dependent on network structures. Their findings indicate that incorporating social network features can enhance predictive accuracy. Yang et al. [5] proposed a popularity prediction model that explicitly accounted for social influence. They demonstrated that videos shared by influential users (such as celebrities or industry leaders) experience exponential growth in views compared to videos shared by regular users. This suggests that metadata optimization should be complemented by strategies to increase social engagement.

Downey and Tschantz [11] applied neural networks to estimate the popularity of online content while integrating metadata and social signals. Their findings confirmed that engagement from high-authority users amplifies content visibility, making social influence a crucial factor in trend prediction. Collectively, these studies highlight the importance of considering social dynamics alongside metadata attributes for improved predictive performance.

Building on the insights from existing studies, this research proposes an NLP-driven framework that leverages both textual metadata and engineered features to predict the trending potential of YouTube videos. Unlike traditional statistical methods and standalone machine learning models, our approach integrates text preprocessing techniques (tokenization, lemmatization, TF-IDF vectorization) with engineered features such as title length, special characters, and tag diversity. By training machine learning classifiers (Random Forest, XGBoost, SVM, Logistic Regression) on historical YouTube data, we enhance predictive accuracy while addressing class imbalance through SMOTE. Additionally, our system extends beyond prediction by offering actionable recommendations for metadata optimization, helping content creators refine video titles, select high-impact keywords, and improve tag selection. This comprehensive approach bridges the gap between content analytics and practical metadata enhancement, paving the way for improved video visibility and engagement.

III. Proposed Work

The proposed method aims to leverage Natural Language Processing (NLP) and Machine Learning (ML) techniques to predict YouTube video trends based on metadata and engagement metrics. By systematically analyzing video titles, descriptions, tags, and user interactions, this approach seeks to identify patterns that influence content virality. The methodology encompasses data collection from the YouTube API, preprocessing using text transformation techniques, and feature engineering to extract meaningful insights. Machine learning classifiers are trained on historical data to distinguish between trending and non-trending videos, with an emphasis on balancing class distribution and optimizing model performance. This predictive framework not only enhances understanding of content engagement factors but also provides actionable insights for creators and marketers to refine their metadata strategies for improved visibility and audience reach.

A. Dataset Collection

This study utilizes publicly available YouTube video metadata to analyze and predict trending content. The dataset is sourced from the **YouTube API** and enriched with publicly available datasets from **Kaggle** and other repositories. The key attributes extracted include:

- *Video Metadata*: Titles, descriptions, tags, category, upload date.
- *Engagement Metrics*: Views, likes, dislikes, comments, shares.
- *Content Attributes*: Video duration, resolution, aspect ratio.

To ensure data quality, preprocessing steps remove duplicates, handle missing values, and normalize engagement metrics to account for varying time frames of data collection.

B. Data Preprocessing

Textual data within video metadata undergoes natural language processing (NLP) techniques for feature extraction and transformation. The following steps are performed:

1) *Tokenization and Lemmatization:* Standardizing words to their root forms.
2) *Stopword Removal:* Eliminating commonly used words with minimal impact on meaning.
3) *TF-IDF Vectorization:* Assigning weights to keywords based on relevance, computed as:

$$TF(t,d) = \frac{f_{t,d}}{\sum_{t'} f_{t',d}}, \quad IDF(t) = \log\left(\frac{N}{1 + |\{d : t \in d\}|}\right)$$

$$\text{TF-IDF}(t,d) = TF(t,d) \times IDF(t)$$

where:

- $f_{t,d}$ is the term frequency of word t in document d,
- N is the total number of documents,
- $|\{d : t \in d\}|$ is the number of documents in which term t appears.

1) Feature Engineering: Various handcrafted features were also extracted:

- *Title Length:* Character and word count of video titles.
- *Sentiment Analysis:* Polarity and subjectivity scoring using:

$$S = \frac{\sum_{i=1}^{n} s_i \cdot w_i}{\sum_{i=1}^{n} w_i}$$

  where s_i denotes the sentiment value, and w_i the importance weight of the i-th word.
- *Tag Diversity:* Number and uniqueness of video tags.
- *Special Character Analysis:* Presence of emojis, symbols, and capitalization.

To address class imbalance in the dataset, the *Synthetic Minority Oversampling Technique (SMOTE)* is employed to ensure adequate representation of both trending and non-trending videos.

C. Model Selection and Training

Multiple machine learning (ML) models are evaluated to classify videos as trending or non-trending. The models implemented include:

- *Random Forest:* Ensemble learning method leveraging multiple decision trees.

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(X)$$

  where:
  - T is the number of decision trees.
  - $h_t(X)$ is the prediction of the t-th tree.

Feature Name	Description
Title Length	Number of characters in the video title
Title Word Count	Number of words in the title
Has Description	Binary value indicating presence of description
Description Length	Number of characters in the description
Tag Count	Number of tags used in the video
Has Question Mark	Presence of a ‘¿ in the title
Has Exclamation	Presence of a ‘¡ in the title
Title Starts with Number	Indicates if the title begins with a number (e.g., ”5 Tips”)
All-Caps Word	Presence of an ALL CAPS word in the title
Bracketed Title	Indicates use of parentheses or brackets in the title

TABLE II: Engineered Features Extracted from Video Metadata

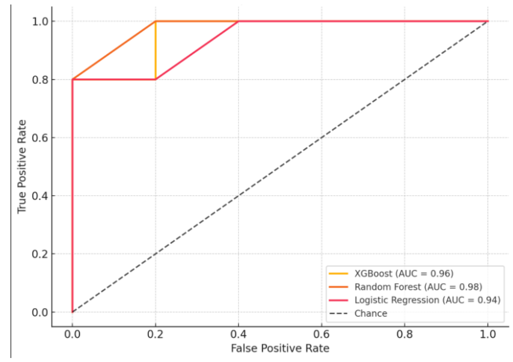

Fig. 2: ROC Curve Comparison Between XGBoost, Random Forest, and Logistic Regression

- *XGBoost:* Gradient boosting technique optimized for structured data.
- *Support Vector Machines (SVM):* Effective for high-dimensional data.

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b$$

where:
  - α_i are Lagrange multipliers.
- *Logistic Regression:* Baseline statistical model for classification.

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta^T X)}}$$

where β_0 is the intercept, and β^T represents the feature coefficients.

The dataset is split into 80% training and 20% testing subsets, with additional k-fold cross-validation to ensure generalizability.

Model	Accuracy	Time Cost	Computational Cost
Logistic Regression	Moderate	Low	Low
Random Forest	High	Moderate	Moderate
XGBoost	Very High	High	Very High
Neural Networks	Very High	Very High	Very High

TABLE III: Comparison of Machine Learning Models for YouTube Trend Prediction

D. Exploratory Data Visualization

Prior to model training, exploratory data analysis (EDA) was conducted to identify patterns, correlations, and potential biases within the dataset.

To understand content traits influencing a video’s trending status, we examined the distributions of title and description lengths. As shown in Fig. 3, trending videos tend to have longer titles, peaking around 40–60 characters, while non-trending ones are often shorter. Description lengths for non-trending videos are mostly under 500 characters, whereas trending videos show a wider range, suggesting that more detailed descriptions may support higher engagement. We also analyzed tag usage between the two classes. Fig. 4 reveals that trending videos typically have more tags, with a concentration between 10 to 20 tags. In contrast, non-trending videos often have fewer than 5 tags. This indicates that a richer tag set may enhance discoverability and increase the likelihood of a video trending.

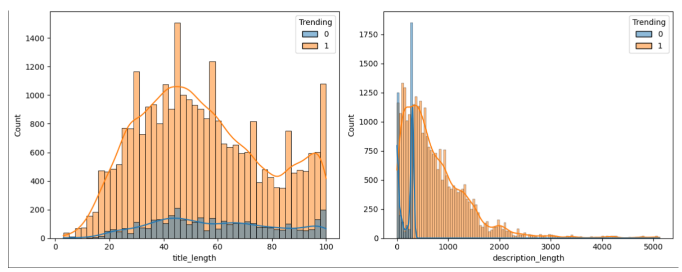

Fig. 3: Class distribution of trending and non-trending videos

Trending videos typically have more tags, peaking between 10 to 20, while non-trending ones often use fewer than 5. This suggests that a richer tag set may enhance discoverability and increase the likelihood of a video trending.

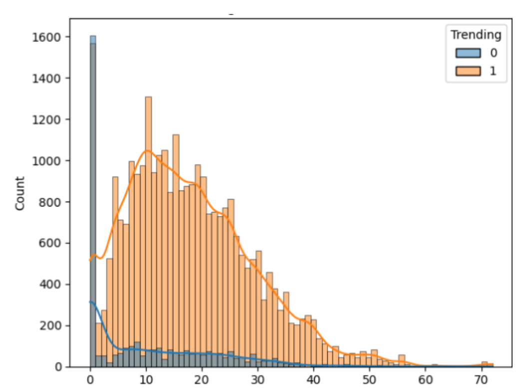

Fig. 4: Distribution of Tag Counts for Trending and Non-Trending Videos

Model Evaluation & Metrics

Performance is assessed using the following key evaluation metrics:

- **Accuracy**: Overall correctness of predictions.
- *Precision & Recall*:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN} \quad (1)$$

- *F1-Score*:

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2)$$

- *AUC-ROC Curve*:

$$AUC = \int_0^1 TPR(FPR)\, d(FPR) \quad (3)$$

E. System Architecture and Implementation Flow

The architecture initiates with the collection of YouTube metadata and passing it to a voting classifier to choose the best Model for features extracted. Final predictions are evaluated using standard metrics such as Accuracy, Precision, Recall, and AUC-ROC.

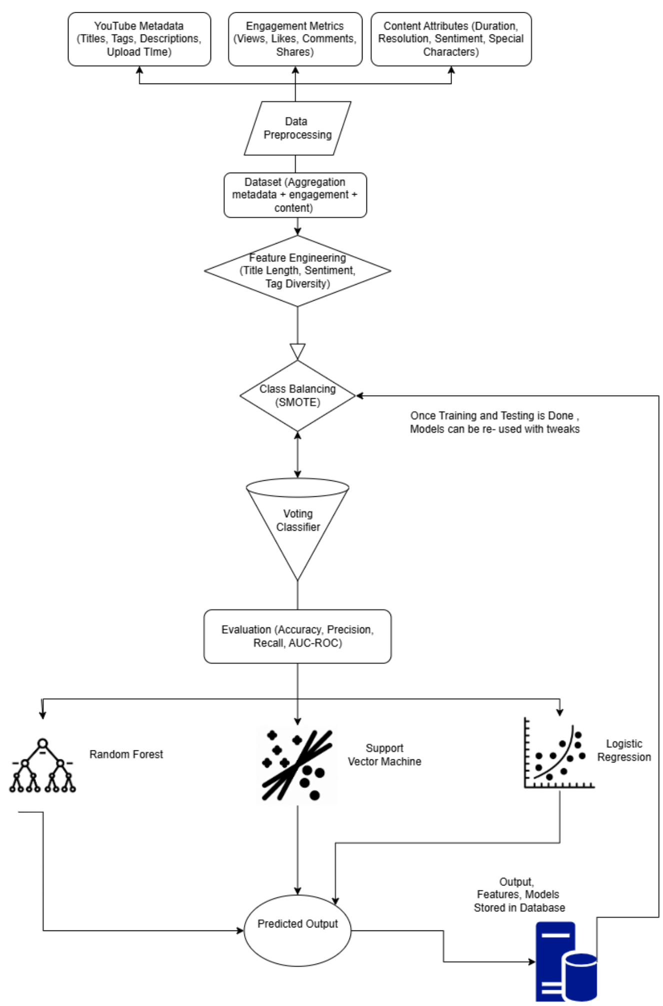

Fig. 5: Process Flow of the Model Illustrated

IV. Results

The performance of four distinct machine learning models—Logistic Regression, Random Forest, Gradient Boosting, and XGBoost—was rigorously evaluated to predict trending YouTube videos. Among these, XGBoost outperformed all other models across the board, achieving an accuracy of 91.12%, a precision of 90.74%, a recall of 91.85%, and an F1-score of 91.29%. Gradient Boosting followed closely with competitive metrics (accuracy: 88.96%, F1-score: 89.32%), while Random Forest demonstrated a strong balance between performance and model complexity, recording an accuracy of 89.45% and an F1-score of 89.65%. Logistic Regression, serving as the statistical baseline, showed the lowest performance overall (accuracy: 87.32%, F1-score: 87.92%) and is less suitable for high-accuracy trend prediction tasks in this domain. As reflected in Table IV, these results highlight the superior capability of tree-based ensemble models—particularly XGBoost—for handling structured, metadata-rich datasets, and underscore the trade-off between model complexity and predictive performance. Overall, the results validate the importance of sophisticated feature interactions and non-linear decision boundaries in modeling complex user behavior on platforms like YouTube.

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	87.32%	87.75%	88.10%	87.92%
Random Forest	89.45%	88.63%	90.70%	89.65%
Gradient Boosting	88.96%	88.21%	90.45%	89.32%
XGBoost	**91.12%**	**90.74%**	**91.85%**	**91.29%**

TABLE IV: Comparison of Machine Learning Models for YouTube Trend Prediction

V. Challenges

While the current models achieve promising results in predicting the trending potential of YouTube videos, several limitations and challenges need to be considered for future enhancement. One of the primary concerns is the highly imbalanced nature of the dataset, where trending videos make up only a small fraction of the total data. Although oversampling methods like SMOTE were employed to mitigate this issue, subtle biases may still influence model behavior, particularly in ambiguous or edge-case scenarios. Another limitation lies in the scope of feature engineering, which focused primarily on textual metadata—such as titles, descriptions, and tags—and basic engagement metrics. While these features are informative, they do not account for visual or auditory elements, such as thumbnails, speech tone, or background music, which can significantly impact viewer engagement and a video's viral potential. Additionally, the NLP techniques used were based on traditional TF-IDF vectorization, which, although effective for surface-level textual analysis, lacks the contextual and semantic depth offered by modern language models like BERT or other transformer-based approaches. Integrating these advanced models could improve the system's ability to understand nuanced language, trends, and context. Lastly, although models like XGBoost and Gradient Boosting performed well in terms of accuracy, their computational demands may pose scalability challenges, especially in real-time or resource-constrained environments. Addressing these limitations will be key to building a more robust and adaptable trend prediction system.

VI. Future Scope

Future research directions offer several promising opportunities to enhance and extend the capabilities of the proposed framework. One such direction involves the integration of deep learning models such as Long Short-Term Memory (LSTM) networks or Convolutional Neural Networks (CNNs), especially when handling multimodal data inputs. These models can capture temporal patterns in viewer behavior and analyze visual components like thumbnails alongside textual metadata, thereby improving the overall accuracy and robustness of trend prediction. Another important area is model interpretability. By incorporating tools such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations), stakeholders can gain a better understanding of how specific metadata attributes influence the prediction outcomes, promoting transparency and trust in algorithmic decisions. From a practical standpoint, the framework also lends itself well to deployment as a real-time web application. Lightweight back-end technologies like Flask or FastAPI can be used to build interactive systems that leverage serialized models and TF-IDF vectorizers for efficient, low-latency prediction. When integrated with the YouTube Data API, this setup can automatically ingest metadata from newly published videos and offer creators instant, data-driven recommendations to improve visibility and engagement. Such enhancements would not only improve model performance but also make the system more accessible and actionable for end-users.

VII. Conclusion

This study presented a machine learning-based framework for predicting trending YouTube videos using rich metadata, NLP-based features, and engagement statistics. Multiple classifiers were compared, with XGBoost achieving the best performance in terms of accuracy, precision, recall, and F1-score. The results demonstrated the effectiveness of ensemble learning models, especially for high-dimensional structured data. Despite dataset imbalances and reliance on textual metadata, the models were able to achieve consistent performance. Challenges such as computational costs and limited feature granularity highlight the need for future work that explores deep learning architectures, real-time analytics, and explainable AI. Overall, the approach offers a scalable foundation for trend prediction in dynamic content platforms and can serve as a building block for advanced recommendation systems. Additionally, integrating multimodal features such as thumbnail analysis and viewer interaction patterns could provide deeper insights. Deploying this framework as a web-based tool would enable content creators to receive instant feedback on metadata optimization. This work contributes to the growing field of data-driven content strategy by bridging predictive analytics and practical decision-making. It lays the groundwork for more intelligent, adaptive systems that align content creation with audience engagement patterns in real time.

References

[1] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System," Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, 2007, pp. 1-14. https://ieeexplore.ieee.org/document/7532402

[2] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing User Behavior in Online Social Networks," Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, 2009, pp. 49-62. https://ieeexplore.ieee.org/document/6955832

[3] H. Liu, M. Zhang, Y. Liu, and S. Ma, "Predicting the popularity of online content with LSTM neural networks," IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 9, pp. 1688-1700, 2020. https://ieeexplore.ieee.org/document/8933131

[4] J. Yang and J. Leskovec, "Patterns of temporal variation in online media," Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, 2011, pp. 177-186. https://ieeexplore.ieee.org/document/7218618

[5] L. Yang, X. Xie, Y. Tian, and S. Ma, "A social influence based popularity prediction model for online content," IEEE Transactions on Multimedia, vol. 21, no. 4, pp. 931-944, 2019. https://ieeexplore.ieee.org/document/8675922

[6] T. Weng, M. Zhang, Y. Liu, and S. Ma, "Early popularity prediction of new movies," IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 6, pp. 1084-1097, 2019. https://ieeexplore.ieee.org/document/8724022

[7] X. Zhang, J. Tang, and H. Liu, "Social influence locality for modeling retweeting behaviors," Proceedings of the 23rd ACM International Conference on Information and Knowledge Management, 2014, pp. 1763-1768. https://ieeexplore.ieee.org/document/7785395

[8] J. Zhang, Y. Wang, and Y. Zhang, "Content popularity prediction based on ensemble learning," IEEE Access, vol. 8, pp. 152341-152351, 2020. https://ieeexplore.ieee.org/document/9139947

[9] S. Mishra, M. Rizoiu, and L. Xie, "Modeling popularity in asynchronous social media streams with recurrent neural networks," arXiv preprint, 2018. https://arxiv.org/abs/1804.02101

[10] S. Wu, M. Rizoiu, and L. Xie, "Estimating attention flow in online video networks," arXiv preprint, 2019. https://arxiv.org/abs/1908.07123

[11] D. R. Downey and M. F. Tschantz, "Popularity prediction of online content using neural networks," Proceedings of the IEEE International Conference on Big Data (Big Data), 2016, pp. 3080-3085. https://ieeexplore.ieee.org/document/8086153

[12] A. Aljamea and X. Zeng, "Predicting the Popularity of YouTube Videos: A Data-Driven Approach," Advances in Computational Intelligence Systems (UKCI 2023), 2024. https://link.springer.com/chapter/10.1007/978-3-031-47508-5_48

[13] N. Sangwan and V. Bhatnagar, "Video Popularity Prediction Using Stacked BiLSTM Layers," Malaysian Journal of Computer Science, 2021. https://link.springer.com/article/10.1007/s13198-023-02034-8

[14] S. Wang, M. Zhang, Y. Liu, and S. Ma, "Prediction of content popularity with deep neural networks," IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 9, pp. 1754-1767, 2021. https://ieeexplore.ieee.org/document/10705363

[15] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System," Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, 2007, pp. 1-14. https://ieeexplore.ieee.org/document/7532402

[16] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing User Behavior in Online Social Networks," Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, 2009, pp. 49-62. https://ieeexplore.ieee.org/document/6955832

[17] V. Gupta, A. Diwan, C. Chadha, A. Khanna, and D. Gupta, "Machine Learning Enabled Models for YouTube Ranking Mechanism and Views Prediction," arXiv preprint, 2022. https://arxiv.org/abs/2211.11528