Video Content Based Advertisement Recommendation System

Patel Bhavya Manishkumar

Student, SCOPE

Vellore Institute of Technology, Chennai, India

Dr. Deepa Nivethika S

Assistant Professor, SCOPE

Vellore Institute of Technology,

Chennai, India

*Abstract*— In the rapidly evolving landscape of digital transformation, advertising plays a pivotal role in shaping consumer behavior and driving business growth. Among the various advertising strategies, content-based advertising stands out as an effective method for delivering targeted messages to viewers. By associating relevant ad content with specific video classes, advertisers can optimize conversion rates and enhance user engagement. Traditionally, digital advertisements are tailored based on user behavior data collected from internet interactions. However, this approach often results in irrelevant or intrusive ads, leading to a suboptimal user experience. To address this challenge, we propose a novel system that leverages natural language processing (NLP) techniques to analyze video metadata and classify videos into relevant categories. By understanding the context and content of each video, our system ensures that users are presented with advertisements directly related to the video they are watching. The key components of our proposed system include video metadata analysis, text classification, ad assignment, and user engagement optimization. By presenting contextually relevant ads, we aim to improve user satisfaction and increase the likelihood of users visiting the advertiser’s website. This personalized approach enhances the overall effectiveness of digital advertising campaigns. In summary, our content-based advertising system bridges the gap between video content and relevant advertisements, creating a win-win situation for both viewers and advertisers. As the digital advertising landscape continues to evolve, leveraging NLP techniques for video classification holds immense potential for enhancing advertising efficiency and user experience.

Keywords— Video Classification, Text Classification, Advertisement Recommendation System

# Introduction

In today’s digital marketing sphere, consumers encounter a myriad of advertisements each day, making the delivery of personalized and pertinent ads more essential than ever. Our project introduces a Content-Based Advertisement Recommendation System tailored for video streaming services. Utilizing Natural Language Processing (NLP), this system assigns videos to thematic categories such as Travel, Science, and Food, ensuring that advertisements are congruent with the video content.

The project involves gathering video metadata and corresponding advertisements. The preprocessing phase tackles issues like missing data and involves steps such as converting text to lowercase, removing numbers, punctuation, and stop words, as well as lemmatization. We employ TF-IDF vectorization to distill features from video titles and descriptions, which are then used to train our models. Our machine learning algorithms, including MultinomialNB and AdaBoostClassifier, have demonstrated high accuracy in categorizing videos, as confirmed by our evaluation metrics.

Our system significantly improves the user experience by providing personalized ads that are contextually relevant to the video content, thereby increasing user satisfaction, engagement, and the likelihood of users visiting the advertiser’s website.

In the current advertising landscape, advertisers have the opportunity to engage with consumers at every touchpoint of their online journey. Success in this domain requires a profound understanding of consumer conversion behavior and immediate access to digital decision-making tools. Modern advertisers are increasingly dependent on technology to optimize their advertising ROI and effectively reach their target audience.

With the rise of streaming platforms like YouTube and Netflix, recommendation systems have become integral to our daily online interactions, influencing everything from e-commerce product suggestions to targeted online ads based on user interests.

The social advertising market continues to expand, reaching audiences across all demographics and geographies. Social networking sites have become a preferred advertising platform due to their ability to connect potential customers with relevant products, thus achieving higher conversion rates. For instance, a study by Adobe Digital Insights highlighted that social media ads attracted three times more new visitors to retailers’ websites compared to existing customers by the end of 2017. The increasing reliance on online services for advertising simplifies consumers’ lives and amplifies the reach for advertisers. Video content, which is projected to account for 80% of web traffic, plays a significant role in the digital marketplace. Platforms like YouTube, with their vast user base, optimize ad space and memory usage, offering advertisers a unique opportunity to engage viewers during extended video sessions.

Our proposed system is designed to enhance the accuracy of ad placements within videos, thereby improving the conversion ratio for the targeted audience. By aligning advertisements with video content, we minimize the occurrence of irrelevant ads, ensuring a more pleasant viewing experience for users.

The growth of paid digital advertising across various platforms underscores the importance of meeting customer needs. Marketers leverage these platforms to connect with their core audience. Our algorithms, which are grounded in video content analysis, reduce server load and streamline data sorting, saving valuable access time. Cutting-edge technologies like machine learning and pattern matching are employed to accurately match advertisements with the right customers based on video metadata, guaranteeing that ads are displayed to users who will find them most beneficial.

# Literature survey

In the context of digital transformation, advertising plays a pivotal role in shaping consumer behavior and driving business growth. Among various advertising strategies, content-based advertising stands out as an effective method for delivering targeted messages to viewers. However, achieving relevance and personalization in ad delivery remains a challenge.

To understand the landscape, we delve into previous studies in this domain. Notably, Onur Sevli et al. [2] focused on recommendation systems within advertisements. Their approach involved leveraging Twitter to display relevant ads to users based on community interactions, language processing, and big data analysis. By identifying word patterns associated with personal interests, they tailored content delivery through a web service.

Abu Bashar et al. [3] explored consumer and advertiser experiences on a social networking platform in Punjab. Their findings emphasized the need for interesting, insightful, and interactive metrics for social media efficacy. Understanding shifting consumer preferences is crucial for effective advertising.

Thirumalaisamy Ragunathana et al. [9] proposed a customer behavior model that triggers relevant ads after a user visits a website. Leveraging Hadoop, they analyzed extensive behavioral data to display contextually appropriate advertisements

In the realm of video content, Kwangsoo et al. [10] developed a multi-label video identification system using YouTube data. As video data proliferates, automated techniques for content classification become essential. Their work utilized the YouTube-8M dataset and machine learning models to address video classification challenges.

Mariana Arantes et al. [4] investigated YouTube video advertising, focusing on revenue generation, video ad popularity, and user behavior. They emphasized the monetization potential for content creators once their channels reach specific views and subscriber thresholds.

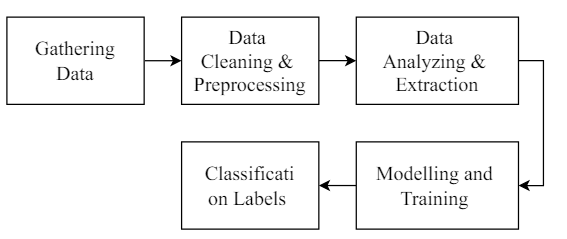
Dardis et al. [11] explored the impact of banner and video ads on brand recall. Their study, conducted in game environments (including advergames), highlighted the effectiveness of video ads, particularly mid-roll placements.

# Proposed Methodlogy

In our research on "Video Content-Based Advertisement Recommendation" using Natural Language Processing (NLP) Text Classification Techniques, we propose a comprehensive methodology aimed at enhancing advertisement suggestion systems by evaluating and categorizing videos based on their textual content. The proposed methodology comprises several sequential stages designed to preprocess data, extract relevant features, train classification models, and make predictions.

We begin by collecting video data from publicly available sources, ensuring diversity and relevance to our research subject. This data includes metadata such as video titles, descriptions, and categories. Prior to analysis, we preprocess the textual data to standardize it and remove noise. Techniques such as lowercase conversion, digit removal, punctuation removal, and stopword elimination are employed to clean the text and enhance its quality for subsequent analysis.

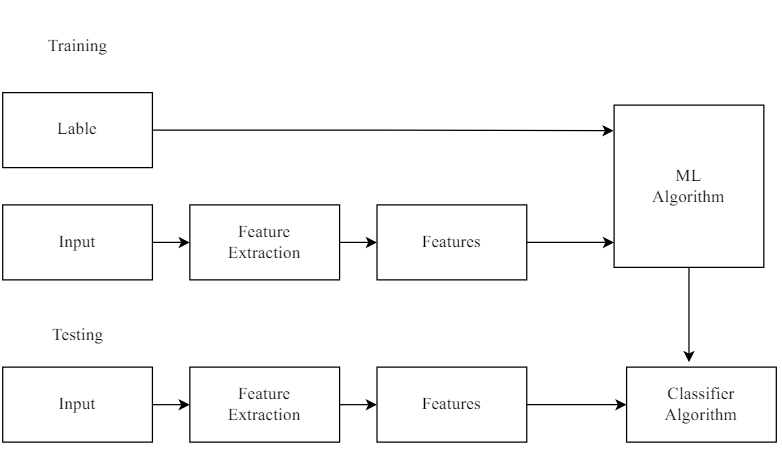
Following data preprocessing, we extract numerical features from the preprocessed textual data using techniques like Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. This process transforms the textual information into a structured format, allowing machine learning algorithms to analyze and interpret the data effectively.



*fig 1 : An overview of the text classification*

Using the extracted textual features as input, we train a variety of classification models, including Naive Bayes, Support Vector Machine (SVM), AdaBoost, and Long Short-Term Memory (LSTM) neural networks. These models are trained on labeled data to learn patterns and relationships between textual features and video categories. The performance of each model is evaluated using established classification metrics to select the most effective model for our recommendation system.

Once the models are trained and validated, we make predictions about the categories or themes of new videos based on their textual content. These predictions serve as the foundation for our recommendation engine, which delivers personalized ads to users based on their consumption habits and interests. The recommendation system utilizes the predicted video categories to match relevant advertisements to users, thereby enhancing user engagement and satisfaction.



*fig 2 : An overview of the text classification*

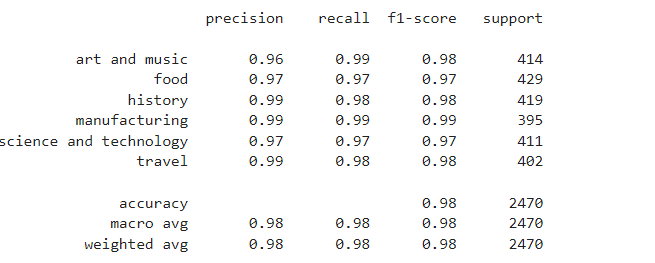
Through this detailed methodology, we aim to develop an efficient and precise framework for evaluating and categorizing videos based on their textual content, ultimately enriching advertisement recommendation systems and enhancing user engagement on digital platforms. By leveraging advanced NLP techniques and machine learning algorithms, we strive to deliver personalized ad experiences to users, thereby maximizing the effectiveness of digital advertising campaigns.

This expanded methodology provides a detailed overview of the research approach, outlining each stage in the process from data collection to advertisement recommendation. The architecture diagram serves as a visual aid to enhance understanding of the research framework, illustrating the flow of data and processes within the proposed methodology.

# Result

SVM Classifier Results

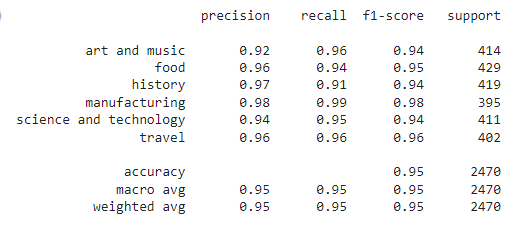
The Support Vector Machine (SVM) classifier exhibited remarkable performance across all evaluated metrics, including precision, recall, and F1-score, achieving an accuracy of 98%. Each category's precision values, which differed from 96% to 99%, showed how well the classifier could identify the relevant categories. Similarly, recall values ranging from 97% to 99% demonstrate the classifier's effectiveness in capturing relevant instances for each category. Overall, the SVM classifier demonstrated robust performance, with an average F1-score of 98% across all categories, ensuring reliable and consistent classification results. Consequently, it emerged as the top-performing model among the evaluated classifiers, providing dynamic and accurate recommendations for advertisement placement.



*Fig 3: Confusion matrix for SVM*

Naïve Bayes Classifier Results

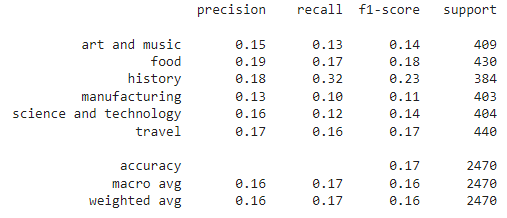
The Naïve Bayes classifier exhibited competitive performance, achieving an overall accuracy of 95%. While the precision values for most categories ranged from 92% to 98% and the recall values varied from 91% to 96%, indicating satisfactory performance, they were slightly lower when compared to the SVM classifier. Nevertheless, the Naïve Bayes classifier demonstrated a balanced performance across all categories, with an average F1-score of 95%. Despite its slightly lower accuracy when compared to SVM, the Naïve Bayes classifier still provides reliable recommendations for advertisement placement, albeit with somewhat less precision and recall.



*Fig 4: Confusion matrix for Naïve Bayes Classifier*

AdaBoost Classifier Results

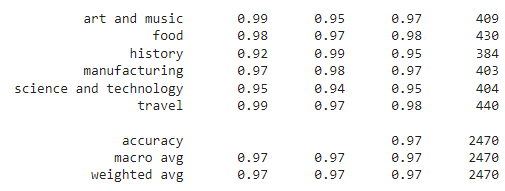
In contrast, the AdaBoost classifier's performance fell short, yielding an accuracy of only 17%. The precision, recall, and F1-score values for all categories were notably lower when compared to both SVM and Naïve Bayes classifiers. With accuracy scores ranging from 13% to 19% and recall values ranging from 10% to 32%, the AdaBoost classifier struggled with accurately identifying instances across categories. Consequently, the average F1-score remained low at 16%, indicating poor overall performance. Given its significantly lower accuracy and poor precision-recall trade-off, the AdaBoost classifier isn't appropriate for providing reliable advertisement recommendations.



*Fig 5: Confusion matrix for SVM*

LSTM Classifier Results

The Long Short-Term Memory (LSTM) classifier exhibited commendable performance across all evaluated metrics, boasting a total precision of 97%. With recall values between 94% and 99% and precision values between 92% and 99%, the classifier indicated that it could correctly identify pertinent categories. Notably, the categories for 'art and music' and 'travel' achieved precision scores of 99%, while 'history' showed a high recall score of 99%. The average F1 score of 97% across all categories underscores the consistent and reliable performance of the LSTM classifier in category identification. Overall, the LSTM classifier presents a robust and efficient approach to content categorization, ensuring accurate and precise recommendations for advertisement placement.



# Conclusion

The SVM classifier outperforms the others in terms of F1-score, accuracy, precision, and recall, according to a thorough review of the Naïve Bayes, LSTM, AdaBoost, and SVM classifiers. With its stellar accuracy of 98% and consistent performance across all metrics, the SVM classifier stands as the optimal choice for dynamic advertisement recommendations. Its strong performance guarantees accurate content classification, resulting in successful ad tactics. On the other hand, although the LSTM and Naïve Bayes classifiers both provide respectable 97% and 95% accuracy, respectively, the AdaBoost classifier performs so poorly that it is not fit for real-world use in advertisement recommendation systems. As a result, the SVM classifier becomes the model of choice for precise and effective content classification, which promotes user engagement and allows for targeted advertisement placement.

The proposed system presents significant benefits to advertisers, enhancing campaign effectiveness and fostering revenue growth. Contextual data is employed to enhance user engagement, which results in the delivery of more precisely targeted ads. Moreover, the insights gained from the research analysis have the potential to improve various aspects of online advertising, providing businesses with a competitive advantage. The algorithms employed in this study demonstrate reliable predictive capabilities, ensuring satisfactory outcomes.

# VI References

1. Virda S, Yu-Qian Z, Achmad Nizar H, Puspa I and Bo H. International Journal of Information Management 48 2019 Exploring the psychological mechanisms from personalized
2. Onur S, Ecir K. International Journal of Tehnički vjesnik 2015 Advertising recommendation system based on dynamic data analysis on Turkish speaking twitter users ISSN 1330-3651
3. Abu B, Irshad A. ELK Asia Pacific Journal of Marketing & Retail Management Effectiveness of Social media as a marketing tool: An empirical study.
4. Mariana A, Flavio F and Jussara M. The Journal of Web Science Volume 4 2018 Towards Understanding the Consumption of Video-Ads on YouTube. 1–19
5. Rushikesh K, Dr. L.M.R.J. Lobo. MAT Journals 7 Feb. 2020 Text Data Analysis for Advertisement Recommendation System Using Multi-label Classification of Machine
6. Sree Vani M. International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016 A Recommender System for Online Advertising. 599-604
7. Nhan N-T, Dana M, Kinda K, David R, Flavian V, Elena Simona L, Steven M, Dominique Q. arXiv:1909.04190v1 [cs.IR] 9 Sep 2019 Recommendation System-based Upper Confidence Bound for Online Advertising.
8. Missi Hikmatyar and Ruuhwan. Journal of Physics: Conference Series 1477 032024 2020 Book Recommendation System Development Using User-Based Collaborative Filtering.
9. Thirumalaisamy R, Sudheer K, Battulab V. International Journal of Computer Science Procedia 50 Advertisement Posting based on Consumer Behavior. 329 – 334
10. Kwangsoo S, Junhyeong J, Seungbin L, Boyoung L, Minsoo J, Jongho N. IEEE Conference on Computer Vision and Pattern Recognition Approach for video classification with multi-label
11. Dardis, F. E., M. Schmierbach, B. Sherrick, F. Waddell, J. Aviles, S. Kumble, and E. Journal of Interactive Advertising Bailey Adver-Where? Comparing the Effectiveness of Banner Ads and Video Ads in Online Video Games. 1–14
12. Arantes, M., F. Figueiredo, and J. M. Almeida. WebSci Understanding video-ad consumption on YouTube: a measurement study on user behaviour, popularity, and content properties.
13. Li, H., and H.-Y. Lo. Journal of Advertising 44(3) Do You Recognize Its Brand? The Effectiveness of Online In-Stream Video Advertisements. 208–218