USING VIDEO TITLES TO PREDICT YOUTUBE AUDIENCE BEHAVIOR BASED ON MACHINE LEARNING

CHIH-CHIEN WANG[1] , YI-FENG LIN[2], YA-CHEN HSIEH[3], YU-HAN KAO[4]

[1,2,3,4] Graduate Institute of Information Management, National Taipei University, New Taipei City, Taiwan

E-mail: [1]wangson@mail.ntpu.edu.tw, [2]jinglewind@gmail.com, [3]s711236114@gm.ntpu.edu.tw, [4]s711236116@gm.ntpu.edu.tw

ABSTRACT

As the leading video-sharing platform, YouTube has evolved into a new form of media, facilitating content dissemination by creators and fostering active audience participation. Enhanced viewer engagement offers creators increased influence and financial benefits due to heightened popularity. However, the vast quantity of videos available on YouTube means that only a select few manage to captivate significant audience attention. Within YouTube's interface, video titles play a crucial role in attracting users, making the exploration of the relationship between audience engagement and video titles an important research topic. This study aims to predict audience engagement by analyzing video titles using various machine learning methods. The study employs the Linguistic Inquiry and Word Count (LIWC) software for natural language processing, which calculates the frequency of word categories in the text, such as pronouns, emotional, and cognitive words, presenting these frequencies as relative percentages. Results indicate that different textual categories within video titles correlate significantly with audience liking and commenting behaviors. Among the evaluated machine learning techniques, Random Forest and K-Nearest Neighbors Regression models exhibit superior predictive performance. The findings provide insights into the intricate interplay between textual features of video titles and their impact on audience engagement. This research serves as a valuable reference for video creators, guiding their decision-making process concerning video titles to optimize audience interaction. Additionally, the study contributes to existing literature by elucidating the nuanced relationship between textual elements in video titles and audience responsiveness.

Keywords: *Video titles, Linguistic Inquiry and Word Count (LIWC), Machine learning, Audience behaviors*

1. INTRODUCTION

User engagement, encompassing behaviors such as views, likes, comments, and shares, is commonly used to evaluate online videos [3]. Such engagement metrics offer creators valuable insights and aid in understanding audience preferences [1]. When users first visit YouTube, video titles often represent the initial engagement point, significantly influencing their decision to view the content. Consequently, the literature has sought to predict video popularity based on the analysis of titles, with findings indicating that attributes such as character counts and emotional sentiment (positive or negative) in titles can notably affect view counts [5, 6].

Despite existing studies, research examining the influence of specific textual features within video titles on audience behavior remains limited. This study addresses this research gap by quantifying textual attributes of video titles using scalar variables derived from linguistic analysis. Specifically, we investigate how these quantized textual characteristics influence audience behaviors, including viewing, liking, and commenting. Machine learning algorithms are employed to predict these behaviors, and we compare the effectiveness of various machine learning techniques in predicting audience engagement based on video title analysis.

This research provides actionable insights for video creators, enabling informed decisions regarding title creation to enhance audience interaction. Enhanced engagement can subsequently increase video popularity and the creators' impact. Although industry practitioners have generally acknowledged the importance of video titles, most existing research has primarily emphasized superficial attributes such as title length or emotional tone. Thus, a critical gap remains regarding the impact of deeper linguistic features

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

within titles on specific aspects of audience engagement.

To address this gap, our research explores the relationship between distinct linguistic categories within video titles—quantified using the Linguistic Inquiry and Word Count (LIWC) methodology—and audience engagement behaviors. Specifically, this study aims to answer the following research questions:

- Can linguistic features derived from YouTube video titles predict audience viewing behaviors?
- How do specific textual attributes within video titles influence audience liking behavior on YouTube?
- Which linguistic characteristics of video titles significantly correlate with audience commenting behaviors?

Addressing these research questions, the study offers practical recommendations for optimizing video titles to enhance audience interaction and popularity. Additionally, the findings contribute theoretically by highlighting specific linguistic factors that influence audience responses, offering a nuanced perspective on content optimization strategies for digital media platforms.

Practically, this research provides valuable guidance for video creators seeking to formulate titles to strategically boost viewer engagement and video popularity. By understanding the linguistic elements resonating with audiences, creators can craft more appealing titles, increasing content visibility and potential revenue. Moreover, the predictive models developed in this study could serve as analytical tools to aid creators and marketers in decision-making regarding content presentation and audience targeting on YouTube and similar platforms.

The subsequent sections are organized as follows: Section two thoroughly reviews relevant literature. Section three outlines the research methodology employed in this study. Section four presents the experimental findings in detail. Finally, section five provides a comprehensive discussion, including conclusions, academic and practical implications, limitations, and recommendations for future research.

2. LITERATURE

2.1 Audience Behaviors

Audience behaviors refer to audience interactions following content consumption. Common metrics for assessing online video audience behaviors include views, likes, comments, and shares [3]. Understanding prospective audience behaviors benefits campaign planning and content creation [1]. Consequently, recent literature has shown a growing interest in video-viewing behaviors. Moldovan et al. [7] demonstrated that highly creative and informative videos significantly enhance viewer engagement, reflected in increased views and comments. Chen and Chang [8] applied machine learning techniques to predict the popularity of newly published videos during the initial release stage. Halim et al. [1] identified features, including video titles, descriptions, and duration, as influential factors for YouTube video popularity, further suggesting the use of video titles in developing prediction models.

2.2 Predicting Audience Behaviors by Video Title

In digital media, content engagement often initiates through title exposure, with attractive titles increasing the likelihood of user interaction. Studies on news article titles confirm their effectiveness in capturing audience attention and predicting article popularity [9-12]. It is thus reasonable to infer that video titles similarly impact viewer engagement.

Content creators craft titles to maximize visibility, though some resort to "clickbait," titles designed with exaggerated or misleading language to attract viewers [4]. Clickbait is frequently employed in political propaganda [13] and by news channels on social media [14], becoming a prevalent online issue. Shang et al. [4] addressed this by developing algorithms using viewer comments and interactions to detect clickbait. Varshney and Vishwakarma [15] employed viewer profiles, video content, and user comments as indicators to identify clickbait videos.

Research has specifically investigated the correlation between title characteristics and video popularity. Tafesse [5] indicated that shorter video titles might enhance viewership, as lengthy titles demand more viewer effort. Wu et al. [6] found that titles explicitly expressing emotions (positive or negative) significantly increased video dissemination. Among methods of analyzing language, the lexicon-based approach is notably effective.

Journal of Theoretical and Applied Information Technology
30th April 2025. Vol.103. No.8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.3 Lexicon-based Natural Language Processing

The proliferation of online textual content has spurred interest in Natural Language Processing (NLP), with numerous NLP methodologies developed. Lexicon-based NLP methods employ predefined dictionaries for textual analysis and categorization. Commonly used lexicon-based NLP programs include AFINN [16], BING Lexicon [17], Linguistic Inquiry and Word Count (LIWC) [18], NRC Emotion Lexicon [19], and Valence Aware Dictionary and sEntiment Reasoner [20]. LIWC, in particular, has been widely applied in marketing research [21].

2.4 Linguistic Inquiry and Word Count (LIWC)

Textual analysis aids trend detection, opinion clustering, and sentiment analysis [22]. LIWC, a prominent NLP program, has extensive applications in marketing [23], media [24], psychology [25], and linguistics [26]. Developed by Pennebaker et al. [27], LIWC quantifies the frequency of word categories such as pronouns, emotional, and cognitive terms, presenting these as relative percentages. LIWC supports multiple languages, including Chinese, Japanese, French, Italian, and Spanish [28-32].

In marketing research, LIWC has proven valuable. Syrdal et al. [33] investigated linguistic features in Instagram posts, finding that variables such as word count, complexity, emotion, and authenticity significantly influenced user engagement. Munaro et al. [3] examined linguistic style and emotional content in YouTube titles, descriptions, and transcripts, demonstrating their impact on viewer behaviors. Accordingly, this study employs LIWC to extract linguistic features from video titles to predict audience engagement.

3. METHODOLOGY

3.1 Research Framework

This study aims to predict audience engagement on YouTube videos by analyzing titles. The collected dataset includes video titles and corresponding metrics such as views, likes, and comments. The Linguistic Inquiry and Word Count (LIWC) program was employed for analyzing linguistic features present in video titles. Seven distinct machine learning techniques—Boosting Regression, Decision Tree, K-Nearest Neighbors, Neural Network, Random Forest, Regularized Linear Regression, and Support Vector Machine—were utilized to predict audience behaviors, including views, likes, and comments. Model performances were evaluated based on error metrics and R-square values. The research framework illustrating this approach is presented in Figure 1.

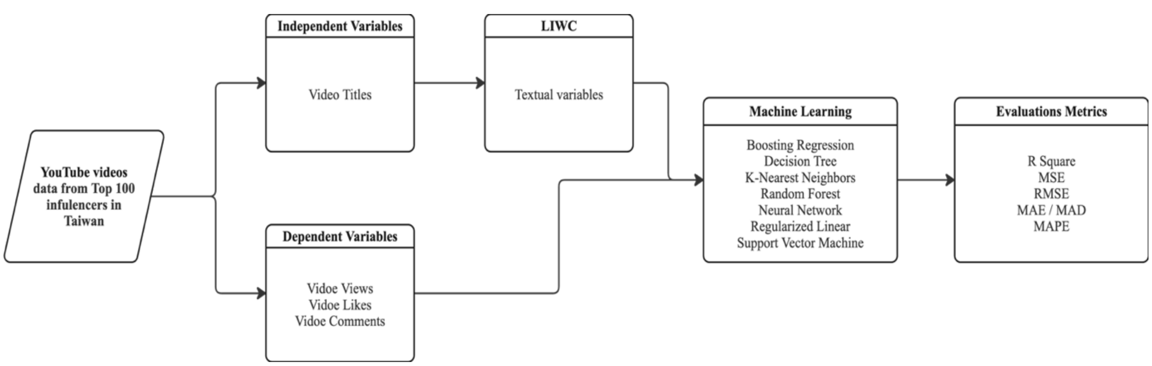

Figure 1: Research Framework

3.2 Data Collection

Python scripts were developed and executed in the Google Colaboratory environment to scrape and process video metadata, leveraging its free and robust computational capabilities. This study targeted prominent Taiwanese YouTubers listed on http://surveys.tw, a website providing comprehensive user surveys about influential social media personalities in Taiwan. Only top YouTubers actively publishing videos with substantial follower counts were included. The study collected video metadata, including titles, publish dates, views, likes, and comments, using the YouTube Data API v3 for videos released between January and September 2023. Data were obtained from 102 YouTube channels across diverse categories, such as blogs, autos, activism, technology, music, entertainment, education, comedy, news, gaming, animation, pets, sports, and lifestyle. The dataset used in this study is original and has not been employed previously in other previous research.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.3 Word Use Analysis

The Linguistic Inquiry and Word Count (LIWC) software was utilized to perform computerized textual analysis, quantifying word usage according to predefined linguistic categories. Given the focus on Taiwanese YouTubers, the Traditional Chinese version of LIWC (C-LIWC), adapted by Huang et al. [32] from the original English LIWC2007, was selected for this study. C-LIWC has been effectively applied in psychology [34], social media analysis [35], and consumer behavior studies [36], making it suitable for analyzing the linguistic features of video titles in this research.

3.4 Machine Learning Models

Machine learning methods use data-driven approaches to generate predictions and support decision-making processes [37]. Supervised learning techniques were chosen due to the nature of the collected dataset. The open-source statistical software JASP (version 0.18.1) was employed to implement seven supervised learning models: Boosted Regression, Decision Tree, K-Nearest Neighbors, Neural Network, Random Forest, Regularized Linear Regression, and Support Vector Machine, for predicting audience engagement behaviors based on video titles.

3.4.1 Boosted Regression

Boosted Regression Trees (BRT), introduced by Elith et al. [39], are advanced statistical models designed for predictive accuracy and explanatory power, capturing nonlinear relationships and variable interactions.

3.4.2 K-Nearest Neighbors

K-Nearest Neighbors (KNN) classification predicts outcomes by analyzing similarity among data points, typically using Euclidean distance. It is robust to noise but relies heavily on data quality and carefully selecting the number of neighbors (k). KNN applies to both classification and regression problems [40].

3.4.3 Neural Network

Neural Networks consist of interconnected processing units (neurons) across multiple layers, enabling them to model complex nonlinear relationships. Widely used in pattern recognition, predictions, and control systems, neural networks learn through supervised backpropagation, adjusting internal weights based on input-output relationships [41].

3.4.4 Decision Tree

Decision Trees are widely utilized supervised machine learning models that employ a hierarchical structure for classification and regression. Instances are classified by sequentially evaluating attributes at each node, moving from root to leaf nodes [40].

3.4.5 Random Forest

Random Forest models leverage parallel ensemble learning by constructing multiple decision trees from subsamples of the dataset, averaging predictions, or using majority voting. This technique reduces overfitting and significantly improves prediction accuracy compared to single-tree models [40].

3.4.6 Support Vector Machines

Support Vector Machines (SVM) construct hyperplanes in high-dimensional space to optimally separate data classes. By maximizing the margin between data points of different classes, SVMs perform effectively in various tasks, supported by kernels such as linear, polynomial, radial basis function (RBF), and sigmoid [40].

3.5 Evaluation Metrics

Five standard evaluation metrics were adopted to compare model performances: Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the explained variance (R-square, R^2). Lower error metrics and higher R-square values indicate superior model performance.

4. DATA ANALYSIS

This study collected metadata from 7,424 videos using Python scripts. Typically, the views, likes, and comments on YouTube videos significantly increase within the initial days following their release, eventually plateauing after two to three weeks. As a general practice, audience engagement metrics stabilize approximately 30 days post-release. Therefore, videos released within 30 days of data collection were excluded, resulting in a refined dataset of 6,512 videos.

We employed LIWC for linguistic analysis of video titles, which initially provided 72 textual features. A correlation analysis was conducted to identify significant relationships (p-value < 0.01) between LIWC features and engagement metrics (views, likes, comments). For viewing behavior, selected significant variables included "word count," "conjunctions," "particle," "general particle," and "insight." For liking behavior, significant variables were "word count," "big words," "we," "conjunctions," "particle," "general particle," "friend," "insight," "death," and "period." Commenting behavior variables included "word count," "words per sentence," "function," "we," "conjunctions," "quant," "negative emotion," "anger," "cognitive process," "certain," "power," "space," "religion," and "death." These variables are summarized in Table 1.

The dataset was partitioned into three subsets: 60% for training, 20% for validation, and 20% for testing. Predictive results for audience behaviors using seven machine learning algorithms are presented, with evaluation metrics including MSE, RMSE, MAE/MAD, MAPE, and R-square. The dataset processing flow is depicted in Figure 2.

Regarding prediction results (Table 2), Support Vector Machine (SVM) Regression yielded the lowest MSE (0.088), RMSE (0.297), and MAE/MAD (0.175) for predicting viewing behavior, though with limited explanatory power (R^2 = 0.063). For liking behavior (Table 3), Boosted Regression achieved the lowest MSE (0.222), RMSE (0.471), and MAE/MAD (0.293), while K-Nearest Neighbors (KNN) Regression demonstrated the highest explanatory power (R^2 = 0.116). For commenting behavior (Table 4), Regularized Linear Regression produced the lowest MSE (0.501) and RMSE (0.708), while KNN Regression showed the highest explanatory power (R^2 = 0.199).

To enhance predictive performance, variables related to likes and comments were transformed using natural logarithms. Results indicated that logarithmic transformation significantly improved the predictive capability for liking behavior (Table 5), with Random Forest Regression achieving an increased explanatory power (R^2 = 0.242). However, the logarithmic transformation did not enhance prediction accuracy for commenting behavior (Table 6), with explanatory power decreasing from R^2 = 0.199 to R^2 = 0.083.

Table 1. Independent Variables That Are Significantly Related With Dependent Variables

Dependent Variables	Significant Independent Variables	Description/Examples
View	word count	Total word count
	conjunction	and, but, so, as
	particle	particle
	general_pa	general particle
	insight	know, how, think, feel
Like	word count	Total word count
	big words	Percent words seven letters or longer
	we	we, our, us, let
	conjunctions	and, but, so, as
	particle	particle
	general_pa	general particle
	friend	friend, boyfriend, girlfriend, dude
	insight	know, how, think, feel
	death	death, dead, die, kill
	period	punctuation
Comment	word count	total word count
	word per sentence	Average words per sentence
	function	the, to, and, I
	we	we, our, us, let
	conjunctions	and, but, so, as
	quant	few, many, much
	negative emotion	bad, hate, hurt, tired
	anger	hate, mad, angry, frustrated
	cognitive process	cause, know, ought
	certain	always, never
	power	superior, bully
	space	in, out, up, there
	religion	god, hell, Christmas, church
	death	death, dead, die, kill

Source: Description/Examples are listed in the Development and Psychometric Properties of LIWC-22 [42]

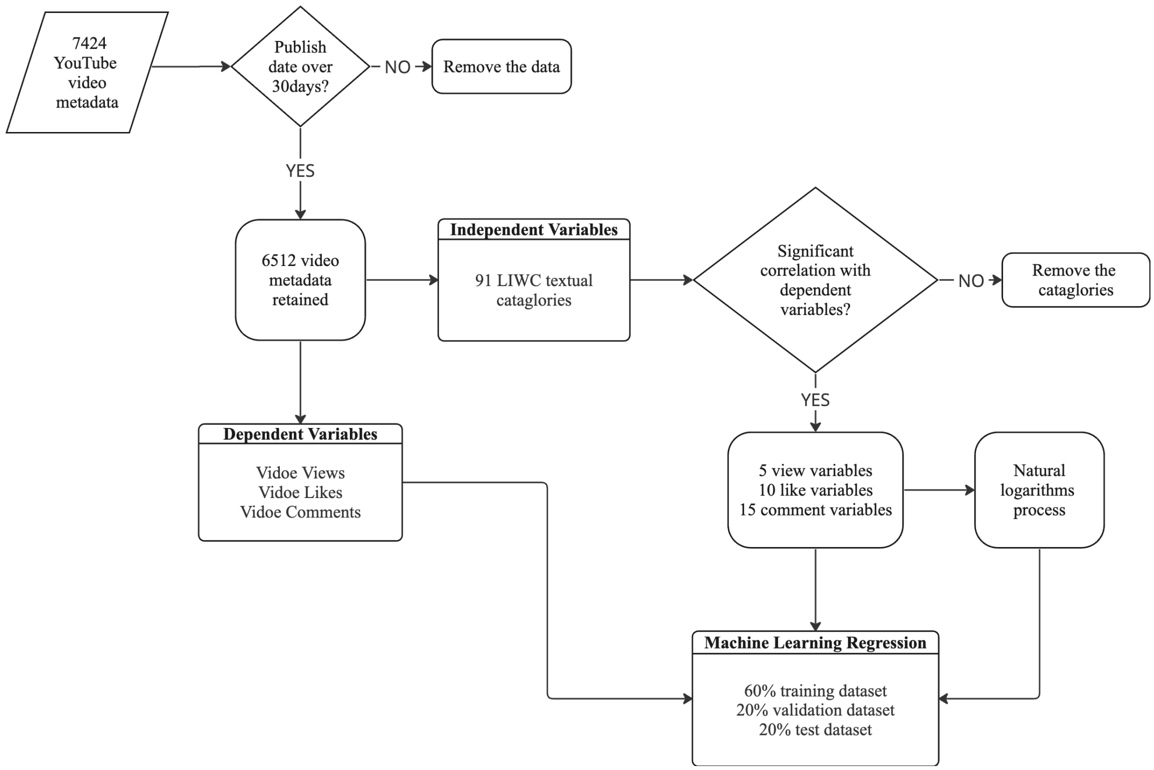

Figure 2. Datset Flow

Table 2. Machine Learning Prediction On Video Viewing Behavior

	Boosting Regression	Decision Tree Regression	K-Nearest Neighbors Regression	Neural Network Regression	Random Forest Regression	Regularized Linear Regression	Support Vector Machine Regression
MSE	0.171	0.967	0.169	0.100	0.768	0.132	0.088
RMSE	0.414	0.983	0.411	0.316	0.876	0.363	0.297
MAE / MAD	0.226	0.275	0.230	0.217	0.257	0.214	0.175
MAPE	**104.9%**	118.1%	200.4%	105.3%	221.9%	238.6%	359.9%
R^2	0	0	0.047	0.014	0.003	0.033	**0.063**

Table 3. Machine Learning Prediction On Video Liking Behavior

	Boosting Regression	Decision Tree Regression	K-Nearest Neighbors Regression	Neural Network Regression	Random Forest Regression	Regularized Linear Regression	Support Vector Machine Regression
MSE	**0.222**	0.302	0.283	0.275	0.286	0.314	0.309
RMSE	**0.471**	0.550	0.532	0.524	0.535	0.560	0.556
MAE / MAD	**0.293**	0.305	0.297	0.320	0.300	0.324	0.266
MAPE	138.5%	131.3%	165.1%	**101.9%**	236.1%	217.7%	204.4%
R^2	0.064	0.056	**0.116**	0.013	0.041	0.042	0.026

Table 4. Machine Learning Prediction On Video Commenting Behavior

	Boosting Regression	Decision Tree Regression	K-Nearest Neighbors Regression	Neural Network Regression	Random Forest Regression	Regularized Linear Regression	Support Vector Machine Regression
MSE	0.672	0.673	0.83	0.674	0.634	**0.501**	0.672
RMSE	0.820	0.820	0.911	0.821	0.796	**0.708**	0.820
MAE / MAD	0.449	0.443	0.464	0.468	0.445	0.432	**0.374**
MAPE	127.1%	209.3%	251.8%	104.9%	223.6%	197.2%	203.3%
R^2	0.030	0.185	0.199	0.008	0.121	0.086	0.051

Table 5. Machine Learning Prediction On Video Liking Behavior (Natural Logarithms)

	Boosting Regression	Decision Tree Regression	K-Nearest Neighbors Regression	Neural Network Regression	Random Forest Regression	Regularized Linear Regression	Support Vector Machine Regression
MSE	0.375	3.750	**0.244**	0.288	0.279	0.282	3.829
RMSE	0.612	1.936	**0.494**	0.537	0.528	0.531	1.957
MAE / MAD	0.344	0.354	**0.278**	0.334	0.286	0.305	0.320
MAPE	148.7%	162.1%	359.6%	**122.6%**	127.2%	191.3%	344.4%
R^2	0.001	0.012	0.122	0.046	**0.242**	0.062	0

Table 6. Machine Learning Prediction On Video Commenting Behavior (Natural Logarithms)

	Boosting Regression	Decision Tree Regression	K-Nearest Neighbors Regression	Neural Network Regression	Random Forest Regression	Regularized Linear Regression	Support Vector Machine Regression
MSE	0.926	1.054	**0.741**	1.362	0.855	1.510	0.778
RMSE	0.962	1.027	**0.861**	1.167	0.925	1.229	0.882
MAE / MAD	0.480	0.505	0.470	0.536	0.481	0.536	**0.397**
MAPE	130.0%	199.8%	256.4%	**120.44%**	261.1%	179.1%	219.3%
R^2	0.022	0.042	0.063	0.045	**0.083**	0.036	0.039

5. DISCUSSION AND CONCLUSIONS

Video titles significantly influence audience attraction and content discoverability, aligning viewers' interests with video content [43]. While prior research primarily analyzed title length and sentiment [5, 6, 44], our study explored deeper linguistic features within video titles to predict audience engagement behaviors (viewing, liking, commenting). The findings reveal that textual variables in titles partially predict audience behaviors, particularly highlighting liking behavior as the most accurately predicted, closely followed by commenting behavior.

The research reveals that Random Forest Regression most effectively predicted liking behavior among the tested machine learning models after data transformation. In contrast, using original data, KNN Regression exhibited optimal predictive performance for commenting behavior.

5.1 Theoretical and Practice Contribution

Our study highlights the significant influence of diverse textual categories within video titles, beyond basic title attributes and emotional content. These findings enrich the theoretical understanding of textual features' nuanced roles in audience engagement. Practically, this research provides valuable guidelines for video creators to strategically optimize titles strategically, potentially enhancing audience interaction, increasing content visibility, and maximizing creators' influence and profitability. Nevertheless, given the models' partial explanatory capability, creators should consider other relevant factors influencing viewer engagement.

5.2 Research Limitations

This research acknowledges several limitations. First, the causal relationship between video titles and audience engagement cannot be precisely established. Second, the Chinese LIWC dictionary (2015 version) may not fully capture contemporary internet slang, potentially limiting analytical accuracy. Third, video thumbnails and YouTube's recommendation algorithms significantly influence audience choice, factors not directly addressed in this study. Future research could integrate these elements to enhance prediction accuracy.

5.3 Future Research Suggestions

Future studies may conduct cross-cultural analyses to explore how audience responses to textual features vary across cultures. Integrating visual stimuli (e.g., thumbnails) could improve predictive accuracy. Exploring temporal dynamics in audience engagement would provide further insights. Additionally, advanced NLP methods like Transformer-based models (e.g., BERT, GPT) could capture linguistic nuances more effectively. Examining recommendation systems through controlled experiments could clarify titles' impacts. Further research might incorporate demographic-based audience segmentation and personalized engagement strategies. Comparative studies across platforms (e.g., TikTok, Instagram) could assess the findings' generalizability. Finally, evaluating the economic impact of optimized titles would offer practical implications for content creators.

5.4 Conclusions

This study concludes that linguistic features in video titles, analyzed using the LIWC method, offer valuable predictive insights into YouTube audience engagement behaviors, especially liking and commenting. Contrasting previous research focused on superficial title attributes, our work highlights subtle linguistic elements driving audience interactions. Demonstrating differential predictive power among various machine learning algorithms, particularly Random Forest and KNN, the study presents new strategies for optimizing digital content. Practically, findings equip video creators with specific linguistic guidelines to strategically enhance viewer engagement, content popularity, and potential revenue streams, underscoring the research's significant academic and practical contributions.

NOTES:

This study's initial draft was developed as a term paper for the research methods course at the Graduate Institute of Information Management, National Taipei University. The first author, Chih-Chien Wang, was the course instructor. The second author, Yi-Feng Lin, voluntarily joined this research. He earned his master's degree from National Taipei University. Both Ya-Chen Hsieh and Yu-Han Kao participated as enrolled students in the course.

REFERENCES:

[1] Z. Halim, S. Hussain, and R. Hashim Ali, "Identifying content unaware features influencing popularity of videos on YouTube: A study based on seven regions," Expert Systems with Applications, vol. 206, p. 117836, 2022/11/15/ 2022, doi: https://doi.org/10.1016/j.eswa.2022.117836.

[2] R. Wang and S. Chan-Olmsted, "Content marketing strategy of branded YouTube channels," Journal of Media Business Studies, vol. 17, no. 3-4, pp. 294-316, 2020/10/01 2020, doi: 10.1080/16522354.2020.1783130.

[3] A. C. Munaro, R. Hübner Barcelos, E. C. Francisco Maffezzolli, J. P. Santos Rodrigues, and E. Cabrera Paraiso, "To engage or not engage? The features of video content on YouTube affecting digital consumer engagement," Journal of Consumer Behaviour, vol. 20, no. 5, pp. 1336-1352, 2021, doi: https://doi.org/10.1002/cb.1939.

[4] L. Shang, D. Zhang, M. Wang, S. Lai, and D. Wang, "Towards reliable online clickbait video detection: A content-agnostic approach," Knowledge-Based Systems, vol. 182, p. 104851, 2019/10/15/ 2019, doi: https://doi.org/10.1016/j.knosys.2019.07.022.

[5] W. Tafesse, "YouTube marketing: how marketers' video optimization practices influence video views," Internet Research, vol. 30, no. 6, pp. 1689-1707, 2020, doi: 10.1108/INTR-10-2019-0406.

[6] J. Wu, Y. Yang, P. Sun, and M. Zhang, "Video Going Viral: Subjective Emotional Clash vs. Objective Emotional Assertion," presented at the E-Business. Digital Empowerment for an Intelligent Future, Cham, 2023.

[7] S. Moldovan, Y. Steinhart, and D. R. Lehmann, "Propagators, Creativity, and Informativeness: What Helps Ads Go Viral," Journal of Interactive Marketing, vol. 47, no. 1, pp. 102-114, 2019, doi: 10.1016/j.intmar.2019.02.004.

[8] Y.-L. Chen and C.-L. Chang, "Early prediction of the future popularity of uploaded videos," Expert Systems with Applications, vol. 133, pp. 59-74, 2019/11/01/ 2019, doi: https://doi.org/10.1016/j.eswa.2019.05.015.

[9] A. Voronov, Y. Shen, and P. K. Mondal, "Forecasting popularity of news article by title analyzing with BN-LSTM network," in Proceedings of the 2019 International Conference on Data Mining and Machine Learning, 2019, pp. 19-27.

[10] W. Stokowiec, T. Trzciński, K. Wołk, K. Marasek, and P. Rokita, "Shallow reading with deep learning: Predicting popularity of online content using only its title," in Foundations of Intelligent Systems: 23rd International Symposium, ISMIS 2017, Warsaw, Poland, June 26-29, 2017, Proceedings 23, 2017: Springer, pp. 136-145.

[11] Y. Keneshloo, S. Wang, E.-H. Han, and N. Ramakrishnan, "Predicting the Popularity of News Articles," in Proceedings of the 2016 SIAM International Conference on Data Mining (SDM), pp. 441-449.

[12] E. Hensinger, I. Flaounas, and N. Cristianini, "Modelling and predicting news popularity," Pattern Analysis and Applications, vol. 16, no. 4, pp. 623-635, 2013/11/01 2013, doi: 10.1007/s10044-012-0314-6.

[13] Y. Lu and J. Pan, "Capturing Clicks: How the Chinese Government Uses Clickbait to Compete for Visibility," Political Communication, vol. 38, no. 1-2, pp. 23-54,

2021/03/15 2021, doi: 10.1080/10584609.2020.1765914.

[14] A.-K. Jung, S. Stieglitz, T. Kissmer, M. Mirbabaie, and T. Kroll, "Click me…! The influence of clickbait on user engagement in social media and the role of digital nudging," PLOS ONE, vol. 17, no. 6, p. e0266743, 2022, doi: 10.1371/journal.pone.0266743.

[15] D. Varshney and D. K. Vishwakarma, "A unified approach for detection of Clickbait videos on YouTube using cognitive evidences," Applied Intelligence, vol. 51, no. 7, pp. 4214-4235, 2021/07/01 2021, doi: 10.1007/s10489-020-02057-9.

[16] F. Å. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," arXiv preprint arXiv:1103.2903, 2011.

[17] M. Hu and B. Liu, "Mining and summarizing customer reviews," presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA, 2004. [Online]. Available: https://doi.org/10.1145/1014052.1014073.

[18] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," 2015.

[19] S. Mohammad and P. Turney, "Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon," in Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, 2010, pp. 26-34.

[20] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, no. 1, pp. 216-225, 05/16 2014, doi: 10.1609/icwsm.v8i1.14550.

[21] J. Hartmann, J. Huppertz, C. Schamp, and M. Heitmann, "Comparing automated text classification methods," International Journal of Research in Marketing, vol. 36, no. 1, pp. 20-38, 2019/03/01/ 2019, doi: https://doi.org/10.1016/j.ijresmar.2018.09.009.

[22] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," Artificial Intelligence Review, vol. 55, no. 7, pp. 5731-5780, 2022/10/01 2022, doi: 10.1007/s10462-022-10144-1.

[23] S. Y. Duncan, R. Chohan, and J. J. Ferreira, "What makes the difference? Employee social media brand engagement," Journal of Business & Industrial Marketing, vol. 34, no. 7, pp. 1459-1467, 2019, doi: 10.1108/JBIM-09-2018-0279.

[24] K. Lanning, G. Wetherell, E. A. Warfel, and R. L. Boyd, "Changing channels? A comparison of Fox and MSNBC in 2012, 2016, and 2020," Analyses of Social Issues and Public Policy, vol. 21, no. 1, pp. 149-174, 2021, doi: https://doi.org/10.1111/asap.12265.

[25] H. A. Burkhardt, G. S. Alexopoulos, M. D. Pullmann, T. D. Hull, P. A. Areán, and T. Cohen, "Behavioral Activation and Depression Symptomatology: Longitudinal Assessment of Linguistic Indicators in Text-Based Therapy Sessions," J Med Internet Res, vol. 23, no. 7, p. e28244, 2021/7/14 2021, doi: 10.2196/28244.

[26] K. Brauer, R. Sendatzki, and R. T. Proyer, "Testing associations between language use in descriptions of playfulness and age, gender, and self-reported playfulness in German-speaking adults," (in English), Frontiers in Psychology, Brief Research Report vol. 13, 2022-September-01 2022, doi: 10.3389/fpsyg.2022.935009.

[27] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: LIWC 2001," Mahway: Lawrence Erlbaum Associates, vol. 71, 2001.

[28] H. Liu et al., "Prediction of Online Psychological Help-Seeking Behavior During the COVID-19 Pandemic: An Interpretable Machine Learning Method," (in English), Frontiers in Public Health, Original Research vol. 10, 2022-March-03 2022, doi: 10.3389/fpubh.2022.814366.

[29] T. Igarashi, S. Okuda, and K. Sasahara, "Development of the Japanese Version of the Linguistic Inquiry and Word Count Dictionary 2015," Frontiers in psychology, vol. 13, p. 841534, 2022.

[30] J. M. Arana, F. Gordillo, J. Darias, and L. Mestas, "Analysis of the efficacy and reliability of the Moodies app for detecting emotions through speech: Does it actually work?," Computers in Human Behavior, vol. 104, p. 106156, 2020/03/01/ 2020, doi: https://doi.org/10.1016/j.chb.2019.106156.

[31] G. Gandino et al., "Linguistic markers of processing the first months of the pandemic COVID-19: a psycholinguistic analysis of Italian university students' diaries," Current Psychology, 2023/05/19 2023, doi: 10.1007/s12144-023-04737-4.

[32] C.-L. Huang et al., "The development of the Chinese linguistic inquiry and word count dictionary," Chinese Journal of Psychology, vol. 54, no. 2, pp. 185-201, 2012.

[33] H. A. Syrdal, S. Myers, S. Sen, P. J. Woodroof, and W. C. McDowell, "Influencer marketing and the growth of affiliates: The effects of language features on engagement behavior," Journal of Business Research, vol. 163, p. 113875, 2023/08/01/ 2023, doi: https://doi.org/10.1016/j.jbusres.2023.113875.

[34] L. Sundararajan, R. S.-K. Ting, S.-K. Hsieh, and S.-H. Kim, "Religion, cognition, and emotion: What can automated text analysis tell us about culture?," The Humanistic Psychologist, vol. 50, no. 2, pp. 213-233, 2022, doi: 10.1037/hum0000201.

[35] H. Gu, J. Wang, Z. Wang, B. Zhuang, and F. Su, "Modeling of User Portrait Through Social Media," in 2018 IEEE International Conference on Multimedia and Expo (ICME), 23-27 July 2018 2018, pp. 1-6, doi: 10.1109/ICME.2018.8486595.

[36] N. Fu, "The cultural influences of narrative content on consumers' perceptions of helpfulness," International Journal of Market Research, vol. 64, no. 3, pp. 354-375, 2022, doi: 10.1177/14707853211023033.

[37] S. Ray, "A Quick Review of Machine Learning Algorithms," in 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 14-16 Feb. 2019 2019, pp. 35-39, doi: 10.1109/COMITCon.2019.8862451.

[38] R. Ahuja, A. Chug, S. Gupta, P. Ahuja, and S. Kohli, "Classification and Clustering Algorithms of Machine Learning with their Applications," in Nature-Inspired Computation in Data Mining and Machine Learning, X.-S. Yang and X.-S. He Eds. Cham: Springer International Publishing, 2020, pp. 225-248.

[39] J. Elith, J. R. Leathwick, and T. Hastie, "A working guide to boosted regression trees," Journal of Animal Ecology, vol. 77, no. 4, pp. 802-813, 2008, doi: https://doi.org/10.1111/j.1365-2656.2008.01390.x.

[40] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Computer Science, vol. 2, no. 3, 2021, doi: 10.1007/s42979-021-00592-x.

[41] A. Dongare, R. Kharde, and A. D. Kachare, "Introduction to artificial neural network," International Journal of Engineering and Innovative Technology (IJEIT), vol. 2, no. 1, pp. 189-194, 2012.

[42] R. L. Boyd, A. Ashokkumar, S. Seraj, and J. W. Pennebaker, "The development and psychometric properties of LIWC-22," Austin, TX: University of Texas at Austin, pp. 1-47, 2022.

[43] H. R. Jamali and M. Nikzad, "Article title type and its relation with the number of downloads and citations," Scientometrics, vol. 88, no. 2, pp. 653-661, 2011/08/01 2011, doi: 10.1007/s11192-011-0412-z.

[44] S. L. d. Sá, A. Paes, and A. A. d. A. Rocha, "Extracting Textual Features from Video Streaming Services Publications to Predict their Popularity," presented at the Proceedings of the Brazilian Symposium on Multimedia and the Web, Belo Horizonte, Minas Gerais, Brazil, 2021. [Online]. Available: https://doi.org/10.1145/3470482.3479624.