**DEEP LEARNING-BASED PREDICTION OF HOTEL REVIEW VIDEO SUCCESS USING TRANSCRIBED SPEECH**

Kiet Lam VI, Thanh Huu NGUYEN, Phong Le DINH, Nghia Ngoc TRAN  
University of Economics Ho Chi Minh City (UEH), Ho Chi Minh City, Vietnam  
Corresponding author: hungngq@ueh.edu.vn

**Abstract**

Review videos have become a powerful tool in shaping consumer perceptions and purchasing decisions on platforms like YouTube, TikTok, and Instagram. However, with increasing content competition, creating engaging videos that drive high audience engagement remains a challenge. This study proposes a system to predict the success of hotel review videos based on their spoken content using Deep Learning techniques. Textual data from video transcripts is processed using NLP methods such as TF-IDF, Word2Vec, and Sentence-BERT. Engagement metrics (likes, shares, comments) are used to generate success scores, which are predicted using Bi-LSTM and Transformer-based models like BERT and RoBERTa. Evaluated on real-world TikTok datasets, the system provides insights into how spoken content impacts audience engagement, offering a data-driven approach to optimizing video scripts before publication.

**Keywords**

Video Review, Hotel Review, Natural Language Processing (NLP), Deep Learning, Engagement Prediction, Social Media Analytics

**1. Introduction**

In recent years, video content has become a powerful marketing tool, especially in the travel and hospitality industry. Platforms like YouTube, TikTok, and Instagram have provided a vast space for hotel review videos, allowing customers to share their experiences and help others make informed decisions about hotel services. These videos offer an authentic perspective on service quality, enabling potential guests to assess aspects such as comfort, amenities, and customer service before making a reservation. However, with millions of videos uploaded daily, creating a review video that stands out, attracts attention, and achieves high engagement remains a significant challenge for content creators.

Among the many factors influencing the success of a review video, spoken content plays a crucial role. The way information is conveyed—through tone, word choice, persuasive language, and calls to action—can significantly impact audience engagement. However, crafting an effective video script is not straightforward, as different audiences have distinct preferences and respond to different linguistic cues. While some content creators produce highly engaging videos, others struggle to achieve similar levels of success despite delivering valuable information. This raises the need for a data-driven approach to understanding which linguistic elements contribute to a successful video.

This study proposes a Deep Learning-based system to predict the success of hotel review videos based on their transcribed speech. The system processes textual data extracted from video transcripts and applies Natural Language Processing (NLP) techniques, including TF-IDF, Word Embeddings (Word2Vec, BERT), and Sentence-BERT [1]. To forecast video success—measured through engagement metrics such as views, likes, comments, and shares—the research employs state-of-the-art Deep Learning models, including Bi-LSTM and Transformer-based architectures such as BERT and RoBERTa [2], [3].

The dataset for this study consists of real-world TikTok hotel review videos, where transcribed speech is analyzed alongside engagement metrics. The goal is to develop a predictive model that provides actionable insights for content creators, helping them optimize their scripts before publishing. By bridging the gap between linguistic analysis and video engagement prediction, this research contributes to the growing field of social media analytics, offering valuable applications in content marketing and hospitality industries.

**2. Literature Review**

The field of video success prediction has gained increasing attention in recent years due to the rapid growth of online platforms where videos are shared and consumed by millions of users. Researchers have explored multiple factors contributing to a video's success, including content, dialogue, sentiment, and audience engagement metrics. While traditional studies focus on video metadata such as titles, descriptions, and thumbnails, recent advancements have leveraged Natural Language Processing (NLP) and Deep Learning to analyze textual data extracted from video transcripts.

One significant application of NLP in video marketing is text-based sentiment and semantic analysis, where businesses extract insights from customer reviews, video transcripts, and social media discussions. Studies by Cambria et al. [4] and Liu [5] have demonstrated that sentiment analysis can help businesses understand consumer perceptions and engagement levels, thereby improving their marketing strategies. In particular, video review sentiment analysis allows companies to assess how customers react to a product or service, offering valuable feedback for service improvement.

Beyond sentiment analysis, deep learning models have been applied to predict the success of videos based on their spoken content. For instance, Mitra et al. [6] developed a deep learning-based system to forecast video success by analyzing dialogue sentiment, engagement patterns, and linguistic features. Transformer-based architectures such as BERT and RoBERTa have also been employed in text classification tasks to improve prediction accuracy in content recommendation systems [2], [3]. LSTM-based models have been particularly effective in analyzing sequential patterns in dialogue, capturing how tone, phrasing, and word choice influence audience engagement [7].

In the hospitality sector, hotel review videos serve as a key source of consumer feedback, reflecting customer satisfaction and expectations. By analyzing the sentiment, tone, and engagement patterns within the dialogue of these videos, businesses can predict the likelihood of a video's success and gain insights into consumer preferences. Previous studies have highlighted the importance of persuasive language and storytelling techniques in influencing audience engagement, suggesting that an AI-driven evaluation system could help content creators optimize their scripts before publishing [8].

This study builds upon these existing works by integrating state-of-the-art NLP techniques and Deep Learning models (Bi-LSTM, Transformer-based models like BERT and RoBERTa) to predict the success of hotel review videos based on transcribed speech. Unlike prior research that primarily focuses on metadata or sentiment analysis alone, this approach provides a more comprehensive, linguistic-driven prediction model for video engagement analytics.

**3. The Proposed Model**

* Definition of success score: Calculated based on engagement metrics such as likes, shares, and comments.
* Text representation methods: TF-IDF, Word Embeddings (Word2Vec, BERT, Sentence-BERT).
* Deep Learning models: Bi-LSTM, Transformer-based architectures (BERT, RoBERTa).

**4. Overview of the System Architecture**

* Data flow: from speech transcription to success prediction.
* Model pipeline: text preprocessing, feature extraction, training, and prediction.
* Integration with TikTok for data collection and validation.

#### **5. Experimental Results**

**5.1 Data Sets**

* Data collection from TikTok review videos.
* Preprocessing techniques: speech-to-text conversion, text cleaning, tokenization, and feature extraction.

**5.2 Model Performance Evaluation**

* Metrics used: Accuracy, Precision, Recall, F1-score, ROC-AUC.

#### **6. Result Discussions**

#### **7. Application Development**

#### **8. Conclusion**

**References**

[1] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, “RoBERTa: A robustly optimized BERT pretraining approach,” *arXiv preprint arXiv:1907.11692*, 2019.

[2] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.

[4] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, “New avenues in opinion mining and sentiment analysis,” *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, 2013.

[5] B. Liu, *Sentiment Analysis and Opinion Mining*, 1st ed. San Rafael, CA: Morgan & Claypool Publishers, 2012.

[6] A. Mitra, R. H. Gao, and A. Gerber, “A deep learning approach to predicting video popularity,” in *Proc. IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2018, pp. 1–10.

[7] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] J. K. Min and C. Park, “Analyzing storytelling techniques in successful online videos: Implications for marketing strategies,” *Journal of Interactive Marketing*, vol. 45, pp. 72–88, 2019.