e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

MARKET TREND ANALYSIS USING DEEP LEARNING

Adhyama MA*1, Alwin Babu*2, Devika S Menon*3, Hadhik Mansoor*4, Santhi P*5

*1,2,3,4Department Of Computer Science And Engineering, IES College Of Engineering, Thrissur, Kerala, India.

*5Asst. Prof. At Department Of Computer Science And Engineering, IES College Of Engineering, Thrissur, Kerala, India.

DOI: https://www.doi.org/10.56726/IRJMETS63471

ABSTRACT

The proposed system aims to leverage advanced computer vision and natural language processing (NLP) techniques to extract and interpret key elements from video ads, such as visual motifs, textual content, and sentiment. By analyzing patterns in ad content, targeting strategies, and consumer engagement, the AI system can identify emerging market trends and predict future consumer preferences. By integrating machine learning algorithms to analyze historical ad data and real-time inputs, offering actionable insights into shifting market dynamics and advertising effectiveness. This can provide businesses with a data-driven tool for anticipating market shifts, optimizing advertising strategies, and enhancing competitive positioning.

I. INTRODUCTION

In the rapidly evolving landscape of digital marketing, the ability to predict market trends and consumer preferences is crucial for maintaining a competitive edge. Traditional methods of market analysis often rely on static data and retrospective evaluations, which may fail to capture the dynamic nature of consumer behavior and emerging trends. To address this challenge, we propose a novel approach that harnesses the power of artificial intelligence (AI) to analyze video advertisements—a rich source of consumer insights.

This approach leverages advanced computer vision and natural language processing (NLP) techniques to extract and interpret key elements from video ads, including visual motifs, textual content, and sentiment. By examining patterns in ad content, targeting strategies, and consumer engagement, our AI system can identify emerging trends and predict future consumer preferences with greater accuracy.

The software we propose integrates machine learning algorithms to analyze both historical ad data and real-time inputs, offering actionable insights into shifting market dynamics and advertising effectiveness. By providing businesses with a data-driven tool for anticipating market shifts, optimizing advertising strategies, and enhancing competitive positioning, this approach represents a significant advancement in the field of market trend prediction.

This paper outlines the design and implementation of our AI-driven system, presents experimental results, and discusses the implications of this technology for the future of marketing and consumer analytics.

1.1 NATURAL LANGUAGE PROCESSING (NLP)

Natural Language Processing (NLP) is a subfield of artificial intelligence that focuses on the interaction between computers and human language. It enables machines to understand, interpret, and generate human language in a way that is both meaningful and useful. NLP combines computational linguistics with machine learning techniques to process and analyze large amounts of natural language data, such as text or speech.

Key applications of NLP include sentiment analysis, language translation, speech recognition, and text summarization. In the context of analyzing video advertisements, NLP can be used to extract and analyze the textual content, such as spoken dialogue, on-screen text, and subtitles, to understand the messaging, tone, and sentiment conveyed in the ads. This helps in identifying patterns and trends that contribute to predicting market behavior.

1.2 DEEP LEARNING

Deep learning is a subset of machine learning that involves training artificial neural networks with multiple layers to model complex patterns and representations in data. Inspired by the structure and function of the human brain, deep learning algorithms can automatically learn features and representations from raw data, making them particularly powerful for tasks like image recognition, natural language processing, and speech

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

recognition. Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are capable of processing large amounts of data and identifying intricate patterns. These models excel in tasks where traditional machine learning methods might struggle due to the complexity or volume of data. In the context of video advertisement analysis, deep learning can be used to process and interpret visual content, recognize objects, detect emotions, and analyze sequences of frames, contributing to the accurate prediction of market trends.

1.3 MULTI MODALITY

Multimodality refers to the integration and analysis of multiple types of data or inputs—such as text, images, audio, and video—within a single model or system. In multimodal systems, different data modalities are processed together to provide a richer and more comprehensive understanding of the information.

For example, in the context of video advertisement analysis, a multimodal approach would combine computer vision (to analyze visual elements), natural language processing (to interpret text and spoken language), and possibly audio processing (to assess sound and tone). By integrating these diverse data sources, multimodal systems can capture complex relationships and patterns that might be missed when analyzing a single type of data. This leads to more accurate predictions and insights, such as identifying market trends and consumer preferences from video ads.

II. LITERATURE SURVEY

Forecasting market trends and consumer behavior in the marketing sector, highlighting the difficulties posed by market volatility and numerous influencing factors. Despite the availability of analytical tools like R, creating a reliable forecasting model remains complex. The methodology includes a thorough review of stock exchanges and market indices, data collection, and comparative statistical analysis to assess the effectiveness of financial indices as economic indicators. Key findings suggest that while tools like R offer valuable insights, developing precise forecasting models requires a nuanced approach to account for the dynamic and multifaceted nature of market trends.[1]

A content-based advertising system aimed at enhancing ad relevance on video platforms by utilizing video metadata and NLP techniques. The system employs a Convolutional Neural Network (CNN) for video classification and compares its performance with a pre-trained model, using datasets created through web scraping. The proposed solution addresses the inefficiencies and irrelevance of current ad placements by implementing two classification models—one for text data and one for video data—to deliver contextually appropriate advertisements. Key findings highlight that the system's effectiveness is rooted in its ability to match ads closely with video content, thereby optimizing conversion rates, increasing user engagement, and ultimately enhancing business value through improved ad relevance and user satisfaction.[2]

Machine learning and deep learning plays a key role in modern intelligent systems and their application in electronic markets and networked businesses. It clarifies the differences between these technologies, emphasizing how they automate analytical model building while addressing challenges such as human-machine interaction and AI servitization. The methodology involves developing a framework that compares explicit programming, shallow machine learning, and deep learning approaches, focusing on data input, feature extraction, model building, and assessment. Key findings reveal that while shallow machine learning depends on manually extracted features, deep learning automated feature extraction from diverse and unstructured data types, such as images and text, enhances model performance and decision-making in complex data environments.[3]

Multimodal feature learning using deep networks is aimed at improving performance by integrating features from different modalities like audio and video. It demonstrates that training deep networks to learn shared representations across modalities can enhance classification tasks, with a focus on audio-visual speech classification. The methodology involves collecting and preprocessing diverse data types, applying feature extraction techniques, and utilizing multimodal fusion strategies and cross-attention mechanisms. Evaluation on the CUAVE and AVLetters datasets shows that the approach achieves superior results in visual speech classification and effective shared representation learning. Key findings include the effectiveness of multimodal architectures and fusion strategies in optimizing model performance and adaptability in real-world applications through techniques like supervised and self-supervised learning, contrastive learning, and model compression.[4]

A robust system for extracting and recognizing artificial text from general-purpose videos, aimed at enhancing automated content-based indexing. The system leverages temporal video features and applies an edge-detection-based text segmentation method selectively on key frames to identify text regions. It employs multiple frame integration, gray-scale filtering, entropy-based thresholding, and line adjacency graphs to enhance detected text areas. Character recognition is achieved through analysis of character side profiles, utilizing vertical and horizontal projections. Experiments on uncompressed MPEG-1 video clips demonstrate the system's effectiveness in accurately detecting and recognizing text in unconstrained video content, highlighting its potential for robust video indexing and content analysis.[5]

Multimodal sentiment analysis models focuses on video sentiment analysis. It categorizes thirty-five state-of-the-art models into eight architecture-based categories and evaluates their performance on the CMU-MOSI and CMU-MOSEI datasets. The analysis highlights that the Multi-Modal Multi-Utterance based architecture is the most effective for sentiment classification, leveraging both contextual and multimodal information. The paper also reviews popular feature extraction methods and benchmark datasets, providing valuable insights into model performance and feature integration techniques. Key findings emphasize the importance of multimodal fusion strategies, such as early and late fusion, and the use of attention mechanisms and transformers to enhance sentiment analysis. This comprehensive survey helps newcomers understand current trends and guides the development of more effective sentiment analysis models.[6]

Traditional unimodal approaches often fall short due to limited information diversity, especially in complex scenes. To tackle this, the study introduces a general MDL framework that integrates spatial information modeling with convolutional neural networks (CNNs), moving beyond pixel-wise classification. The MDL framework is validated through extensive experiments on two multimodal remote sensing datasets, focusing on multi-modality and cross-modality learning. The study explores various fusion strategies and network architectures, presenting five unified fusion architectures within the MDL framework. Key findings demonstrate that incorporating diverse data modalities, such as optical, radar, and thermal imagery, significantly enhances classification accuracy and robustness. The resulting codes and datasets will be publicly available, contributing valuable resources to the remote sensing community.[7]

A Fully Convolutional Neural Network (FCNN) designed to address issues in traditional CNN classifiers caused by linear (fully connected) layers, such as a high number of parameters leading to overfitting and the curse of dimensionality. The proposed FCNN architecture eliminates these linear layers, using only convolutional layers with a Softmax Loss for training. Additionally, a novel softmax-free loss function, POD Loss, based on Predefined Optimal-Distribution of latent features, is introduced to further enhance classification accuracy and robustness. Experiments demonstrate that this FCNN approach reduces parameter complexity and computational demands while improving recognition performance.[8]

By employing edge detection, grayscale processing, gradient detection algorithms, and speech recognition, the study aims to enhance ad matching with video attributes, potentially boosting consumer engagement and sales. The core issue is the challenge of dynamically and accurately matching ads to video content in real-time to improve viewer engagement and advertising effectiveness.[9]

There is huge difficulty in managing and responding to real-time data in financial markets. Traditional methods struggle with the rapid influx of information from various sources, making timely decision-making challenging. The proposed approach combines Natural Language Processing (NLP) and Machine Learning (ML) to analyze market sentiment in real-time, providing timely updates and enhancing decision-making for traders and investors. Ethical concerns regarding the transparency of automated sentiment analysis are also considered.[10]

Various neural network models used for NLP tasks, such as Convolutional Neural Networks (CNNs) for entity recognition, Recurrent Neural Networks (RNNs) for context-dependent language, Recursive Neural Networks for parsing, and Reinforcement Learning (RL) for dialogue generation. Key applications of NLP, such as virtual assistants, language translation, and machine transcription, are discussed.[11]

A framework for predicting the click-through rate (CTR) of video advertisements using a multimodal approach that integrates video, text, and metadata features is needed for separating and normalizing categorical and continuous metadata, and incorporating regularization layers to prevent overfitting, the proposed method improves prediction accuracy. Experimental results show a significant increase in prediction accuracy, with a correlation coefficient of 0.695 compared to a baseline of 0.487.[12]

The performance of recurrent neural networks (RNNs) can be enhanced in sequence modeling tasks. The study identifies three key areas for deepening RNNs: the input-to-hidden function, the hidden-to-hidden transition, and the hidden-to-output function. Two novel deep RNN architectures are proposed and evaluated on tasks such as polyphonic music prediction and language modeling. The results show that deeper RNNs can outperform shallow models by better capturing complex temporal dependencies.[13]

Efficient video analytics on mobile devices, which face delays due to intensive computation and high transmission times when using cloud services. The authors propose using mobile edge computing to reduce delays by offloading tasks to nearby edge servers. The problem is formulated as a contextual Multi-armed Bandit problem, and a Bayesian Optimization-based online learning algorithm is developed to adaptively select the best server and frame resolution. The approach shows significant improvements in balancing accuracy and processing rate.[14]

Convolutional Neural Networks (CNNs) and their key components, including convolutional layers, non-linearity layers, pooling layers, and fully-connected layers. CNNs can efficiently handle large datasets and perform complex pattern recognition tasks, particularly in image classification and computer vision. By managing spatially invariant features and hierarchical abstractions, CNNs overcome limitations faced by traditional Artificial Neural Networks (ANNs) and advance applications in various fields.[15]

III. METHODOLOGY

3.1. METHODOLOGICAL REVIEW

3.1.1 DATA COLLECTION

Gathered a diverse set of video advertisements. These can come from various platforms like YouTube, social media, streaming services and cover different domains like cosmetics, construction, finance, education, entertainment and technology.

Along with the video content, associated text such as advertisement titles, descriptions, and captions and metadata like ad category, duration, upload date, target audience.

3.1.2 FEATURE EXTRACTION AND INTEGRATION

Recognizing video ads requires multimodal—combining visual, textual , and metadata information and extract features from each modality. Using deep learning models such as convolutional neural networks, CNNs to extract visual features from video frames. These features capture visual patterns, objects, and scene context.

3.1.3 DEEP LEARNING MODEL

A neural network is designed that takes the combined features as input. The model of regularization Since training data may be limited, regularization layers may be needed to prevent overfitting.

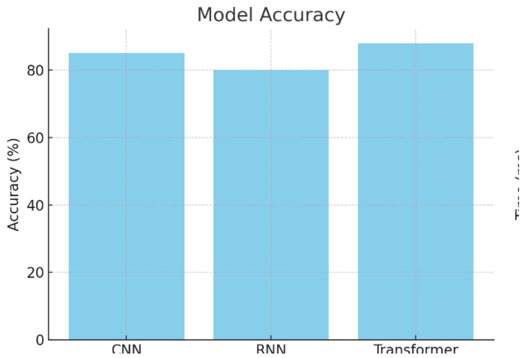
Output Layer is used to Predict the CTR (click-through rate) as a regression task.

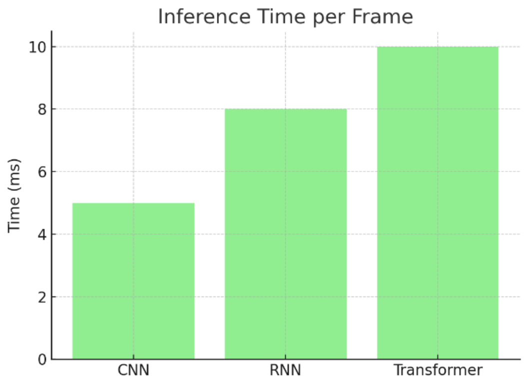
3.2 REVIEW OF DATASET

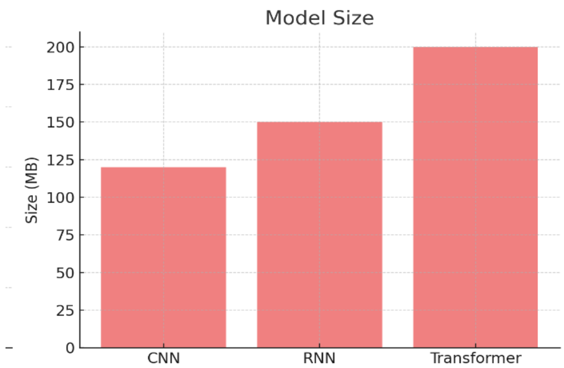
Dataset publicly available are of advertisements in newspaper. This is data set does not help our current idea to train a model to analyze video advertisements[16]. This led us to create a custom dataset with advertisements of different sectors like cosmetics, construction, finance, education, entertainment and technology. The data set consists of advertisement videos of 30 second long and 1280 x 720 pixels.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

IV. RESULTS AND DISCUSSIONS







These are graphs comparing CNN, RNN, and Transformer models for video analysis across three key metrics: Accuracy, Inference Time, and Model Size.

1. **Accuracy**: Transformers have the highest accuracy, followed by CNNs and then RNNs.
2. **Inference Time**: CNNs are the fastest, with Transformers being the slowest.
3. **Model Size**: Transformers have the largest model size, while CNNs have the smallest.

V. CONCLUSION

The software cutting-edge solution for market trend prediction by harnessing the power of artificial intelligence to analyze video advertisements. By utilizing advanced computer vision and natural language processing techniques, the system aims to effectively extract and interpret critical elements from ads, including visual motifs, textual content, and sentiment. Through the integration of machine learning algorithms, it not only processes historical ad data but also incorporates real-time inputs to deliver actionable insights. This innovative approach enables businesses to anticipate market shifts, refine their advertising strategies, and strengthen their competitive edge by providing a robust, data-driven tool for understanding and responding to evolving consumer preferences and market dynamics.

The use of CNN model is suitable for the project as it is the fastest as well as having the smallest model size along with adequate accuracy.

VI. REFERENCES

[1] C. A. Anuradha, A. Sivakumar and M. Sivaraman, "Market Analysis and Forecasting using Trending Data Analytical Tool," 2023 International Conference

[2] Conference Series, Volume 1854, International Conference on Future of Engineering Systems and Technologies (FEST) 2020 18-19 December 2020, Greater Noida, Delhi/NCR, India

[3] Janiesch, Christian, Patrick Zschech and Kai Heinrich. "Machine learning and deep learning." Electronic Markets 31 (2021): 685 - 695.

[4] Ngiam, Jiquan, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee and A. Ng. "Multimodal Deep Learning." International Conference on Machine Learning (2011).

[5] Teo Boon Chen, D. Ghosh and S. Ranganath, "Video-text extraction and recognition," 2004 IEEE Region 10 Conference TENCON 2004., Chiang Mai, Thailand, 2004, pp. 319-322 Vol. 1, doi: 10.1109/TENCON.2004.1414421

[6] Sarah A. Abdu, Ahmed H. Yousef, Ashraf Salem, "Multimodal Video Sentiment Analysis Using Deep Learning Approaches, a Survey", Information Fusion,Volume76,2021,1566-2535, https://doi.org/10.1016/j.inffus.2021.06.003.

[7] D. Hong et al., "More Diverse Means Better: Multimodal Deep Learning Meets Remote-Sensing Imagery Classification," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 5, pp. 4340-4354, May 2021, doi: 10.1109/TGRS.2020.3016820.

[8] Q. Zhu and X. Zu, "Fully Convolutional Neural Network Structure and Its Loss Function for Image Classification," in IEEE Access, vol. 10, pp. 35541-35549, 2022, doi: 10.1109/ACCESS.2022.3163849.

[9] R. V. Kaushik, R. Raghu, L. M. Reddy, A. Prasad and S. Prasanna, "Ad analysis using machine learning: Classifying and recommending advertisements for a given category of videos, using machine learning," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 2017, pp. 2434-2437, doi: 10.1109/ICECDS.2017.8389887.

[10] 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024, pp. 1585-1588.

[11] Natural Language Processing," in Understanding Artificial Intelligence: Fundamentals and Applications, IEEE, 2022, pp.29-37, doi: 10.1002/9781119858393.ch4.

[12] J. Ikeda, H. Seshime, X. Wang and T. Yamasaki, "Predicting Online Video Advertising Effects with Multimodal Deep Learning," 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, pp. 2995-3002, doi:10.1109/ICPR48806.2021.9412666

[13] Neural and Evolutionary Computing (cs.NE); Machine Learning (cs.LG); Machine Learning (stat.ML) arXiv:1312.6026 [cs.NE](or arXiv:1312.6026v5 [cs.NE] for this version) https://doi.org/10.48550/arXiv.1312.6026

[14] H. Liu and G. Cao, "Deep Learning Video Analytics Through Online Learning Based Edge Computing," in IEEE Transactions on Wireless Communications, vol. 21, no. 10, pp. 8193-8204, Oct. 2022, doi: 10.1109/TWC.2022.3164598

[15] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186

[16] Pooja Jain1 , Kavita Taneja2 , Harmunish Taneja "Convolutional Neural Network Based Advertisement Classification Models for Online English Newspapers", Turkish Journal of Computer and Mathematics Education, Vol.12 No.2 (2021), 1687-1698