Journal of Sensors, IoT & Health Sciences
ISSN: 2584-2560
Vol.02(03), September 2024, pp.13-21

Research Article

Sensor Deployed Agricultural Land for E-Commerce Platform with Clustering of Nodes in the Network

Asmatullah Nashir[1,*]
[1]Assistant Professor, Department of Information technology, Badakhshan University, Afghanistan.
[*]Corresponding Author: Asmatullah Nashir. Email: nashirasmatullah@gmail.com

Received: 01/08/2024; Accepted:22/09/2024.
DOI: https://doi.org/10.69996/jsihs.2024013

Abstract: A sensor-deployed agricultural land for an e-commerce platform integrates IoT sensors and wireless networks to monitor crop health, soil conditions, and environmental factors in real-time. These sensors collect data on moisture levels, temperature, humidity, and nutrient content, which is transmitted to a centralized platform. Through clustering of sensor nodes within the network, efficient data aggregation and communication are achieved, reducing energy consumption and improving network performance. This paper presents a comprehensive analysis of agricultural product e-commerce leveraging advanced data analysis techniques. By employing methods such as word frequency analysis, sentiment analysis, and clustering algorithms, we explore the multifaceted landscape of online agricultural markets. This study investigates the prevalence of key themes and trends in agricultural e-commerce, ranging from product categories to digital marketing strategies. Through sentiment analysis, we uncover nuanced patterns in consumer sentiment, providing insights into effective market positioning strategies. Our findings reveal that "agricultural products" emerge as the most prevalent theme, accounting for 8.5% of the dataset. Additionally, sentiment analysis indicates a positive sentiment towards e-commerce platforms, with a precision of 0.85 and an F1-score of 0.83. Clustering analysis delineates five distinct clusters, each representing different facets of agricultural e-commerce, with Cluster 1 focusing on agricultural products and Cluster 2 on digital marketing strategies.

Keywords: Sensor Environment; Agriculture; E-commerce; Clustering; Network Environment

1 Introduction

In recent years, the landscape of agricultural product e-commerce has undergone significant transformation, spurred by technological advancements and shifting consumer preferences[1]. Traditional methods of buying and selling agricultural goods have evolved, with online platforms playing an increasingly prominent role. These platforms offer farmers and producers direct access to consumers, bypassing traditional intermediaries and enabling them to reach wider markets. Additionally, the convenience of online shopping has appealed to consumers seeking fresh, locally sourced products delivered to their doorstep. This trend has been further fueled by growing concerns about food safety, quality, and sustainability, prompting consumers to seek out transparent supply chains and traceable products [2]. As a result, agricultural product e-commerce has experienced exponential growth, with both established companies and startups vying for a share of this burgeoning market. However, challenges such as logistics, last-mile delivery, and digital literacy among farmers persist, highlighting the need for continued innovation and infrastructure development in this sector. In the realm of agricultural product e-commerce, leveraging new media videos and clustering algorithms has emerged as a potent strategy in recent years [3]. This innovative approach harnesses the power of visual storytelling and data analysis to enhance user engagement and streamline the shopping experience [4]. By utilizing clustering algorithms, e-commerce platforms can categorize vast amounts of video content based on various attributes such as product type, production method, or consumer preferences [5]. This allows for personalized recommendations and targeted marketing campaigns tailored to individual users' interests. Moreover, new media videos provide an immersive and informative way to showcase agricultural products, offering consumers a closer look at the production process, farm-to-table journey, and the people behind the products [6]. This storytelling aspect not only cultivates consumer trust and loyalty but also helps differentiate brands in a crowded marketplace. Additionally, the data generated from user interactions with these videos can further refine clustering algorithms, continuously improving the relevance and effectiveness of product recommendations [7]. However, challenges such as data privacy concerns and the need for robust infrastructure to support video streaming and analysis must be addressed to fully realize the potential of this approach.

In the agricultural product e-commerce, leveraging new media videos and clustering algorithms has emerged as a potent strategy in recent years [8]. This innovative approach harnesses the power of visual storytelling and data analysis to enhance user engagement and streamline the shopping experience. By utilizing clustering algorithms, e-commerce platforms can categorize vast amounts of video content based on various attributes such as product type, production method, or consumer preferences [9]. This allows for personalized recommendations and targeted marketing campaigns tailored to individual users' interests. Moreover, new media videos provide an immersive and informative way to showcase agricultural products, offering consumers a closer look at the production process, farm-to-table journey, and the people behind the products [10]. This storytelling aspect not only cultivates consumer trust and loyalty but also helps differentiate brands in a crowded marketplace. Additionally, the data generated from user interactions with these videos can further refine clustering algorithms, continuously improving the relevance and effectiveness of product recommendations. However, challenges such as data privacy concerns and the need for robust infrastructure to support video streaming and analysis must be addressed to fully realize the potential of this approach. In agricultural product e-commerce, the integration of new media videos with clustering algorithms has become increasingly prevalent. This innovative approach combines visual storytelling with data analysis to optimize user engagement and enhance the shopping experience [11-15]. Clustering algorithms categorize video content based on various attributes, enabling personalized recommendations tailored to individual preferences. New media videos offer immersive insights into agricultural products and production processes, fostering consumer trust and brand differentiation [16-18]. However, challenges such as data privacy concerns and the need for robust technology infrastructure remain.

This paper makes several significant contributions to the field of agricultural e-commerce research. Firstly, by applying advanced data analysis techniques such as word frequency analysis, sentiment analysis, and clustering algorithms, we provide a comprehensive understanding of the dynamics and trends within online agricultural markets. Through these methodologies, we offer insights into the prevalence of specific product categories, consumer sentiment towards e-commerce platforms, and the multidimensional nature of marketing strategies employed in the agricultural sector. Additionally, our study contributes to the development of practical tools and frameworks for market analysis and strategy formulation in agricultural e-commerce. By delineating distinct clusters of keywords representing various aspects of the industry, we offer a structured approach for stakeholders to identify and prioritize areas for intervention and optimization. Furthermore, our findings contribute to the advancement of knowledge in agricultural e-commerce by shedding light on the challenges and opportunities inherent in this rapidly evolving landscape.

2 Proposed Stop Word n-gram Clustering Classification (n-gram CC)

In the pursuit of advancing agricultural e-commerce, a novel approach known as Proposed Stop Word n-gram Clustering Classification (n-gram CC) has been proposed. This method, rooted in computational linguistics and data analysis, aims to enhance the classification accuracy of agricultural product data. The derivation of n-gram CC involves the integration of stop word removal techniques and n-gram clustering algorithms, tailored specifically for the agricultural domain. Stop words, commonly occurring words like "the" or "and" that often carry little semantic value, are filtered out to focus on more meaningful content. Meanwhile, n-gram clustering groups sequences of n words together to capture contextual relationships within the data. The classification aspect of n-gram CC involves assigning labels or categories to clusters based on their content. Mathematically, this process can be represented by equations that define the criteria for stop word removal, n-gram formation, and cluster classification. The first step involves removing stop words from the text data. Let's denote: DD as the dataset containing agricultural product descriptions. SS as the set of stop words.

The process of stop word removal can be represented by the equation: Dclean=D−S $Dclean = D - S$ Here, $Dclean$ represents the dataset with stop words removed. After stop word removal, we form n-grams from the remaining words in the dataset. An n-gram is a contiguous sequence of nn words from a given text. Let's denote Dclean as the cleaned dataset after stop word removal. nn as the number of words in each n-gram. The formation of n-grams can be represented by the equation (1)

$$n - grams = \{sequence\ of\ n\ words\ \mid\ for\ each\ d \in Dclean\} \quad (1)$$

This equation signifies that we iterate over each cleaned text description dd in Dclean and generate all possible sequences of nn words. Once we have formed the n-grams, we apply clustering algorithms to group similar n-grams together. Let's denote: n-gramsn-grams as the set of all n-grams. KK as the number of clusters. CC as the set of clusters. The clustering process can be represented by the equation (2)

$$C = Clustering_{Algorithm(n-grams,K)} \quad (2)$$

The clustering algorithm partitions the n-grams into KK clusters based on their similarities. After clustering, we assign labels or categories to each cluster to represent different agricultural product types or attributes. Let's denot C as the set of clusters. L as the set of labels/categories. The classification process can be represented by the equation (3)

$$Cluster\_Labels = Classification\_Algorithm(C, L) \quad (3)$$

Here, the classification algorithm assigns labels to each cluster based on the content of the n-grams it contains.The proposed Stop Word n-gram Clustering Classification (n-gram C) method introduces a novel approach to enhancing agricultural e-commerce by refining the categorization of product descriptions. Initially, the process involves removing commonly occurring stop words from the dataset, focusing on meaningful content. Subsequently, n-grams, which are contiguous sequences of words, are formed from the cleaned dataset, capturing contextual relationships within the text. Clustering algorithms are then applied to group similar n-grams together, forming clusters based on their similarities. Finally, each cluster is assigned a label or category through a classification algorithm, representing different agricultural product types or attributes. This methodological framework, delineated through equations, offers a systematic approach to improving the accuracy of search results and personalized recommendations within agricultural e-commerce platforms.

3 N-gramCC for the Word Frequency Analysis in Agricultural Products

The utilization of N-gram for word frequency analysis in agricultural products signifies a tailored approach to extracting meaningful insights from textual data within this domain. Initially, the method involves forming n-grams, which are contiguous sequences of words, from the agricultural product descriptions. Let D represent the dataset containing these descriptions, and nn denote the number of words in each n-gram. The process of forming n-grams can be mathematically expressed as in equation (4)

$$n - grams = \{sequence\ of\ n\ words\ \mid\ for\ each\ d \in D\} \quad (4)$$

Here, we iterate over each description dd in the dataset DD and generate all possible sequences of nn words. Following the formation of n-grams, the next step involves calculating the frequency of occurrence of each n-gram within the dataset. Let (n-gram)f(n-gram) denote the frequency of a particular n-gram. The frequency analysis can be represented as in equation (5)

$$f(n - gram) = \frac{Count\ of\ n{-}gram\ occurrences/}{Total\ number\ of\ n{-}grams} \times 100 \quad (5)$$

This equation calculates the frequency of each n-gram by dividing the count of its occurrences by the total number of n-grams in the dataset, expressed as a percentage. Finally, the N-gram C method applies clustering algorithms to group similar n-grams together based on their frequency profiles. Let C represent the set of clusters formed, and K denote the number of clusters. The clustering process can be described as in equation (6)

$$C = Clustering\_Algorithm(n - grams, K) \quad (6)$$

Here, the clustering algorithm partitions the n-grams into K clusters, with each cluster representing a distinct frequency profile. N-gram C for e-commerce video word frequency analysis and classification of keywords in agricultural product e-commerce new media videos represents a sophisticated approach to extracting insights from multimedia content. The process begins by forming n-grams from the transcribed text of these videos, where D represents the dataset containing the transcriptions and nn denotes the number of words in each n-gram stated in equation (7)

$$\boldsymbol{n - grams = \{sequence\ of\ n\ words\ \mid\ for\ each\ d \in D\}} \quad (7)$$

This equation iterates over each video transcription d in the dataset D and generates all possible sequences of nn words. Following the formation of n-grams, the frequency of occurrence of each n-gram within the dataset is calculated. Let (n-gram) represent the frequency of a particular n-gram. In figure 1 illustrated the E-commerce platform for the agricultural products.

Figure 1: E-Commerce in Agricultural Products

```
Algorithm 1: N-gram Classification for the Agricultural E-Commerce
Input:
- Dataset D containing transcribed text from e-commerce videos
- Number of words in each n-gram: n
- Number of clusters: K
1. Form n-grams from the dataset:
   n_grams = []
   for each video_transcription in D:
       words = tokenize(video_transcription)  // Tokenize the text into individual words
       for i from 0 to length(words) - n:
           n_gram = concatenate(words[i:i+n])  // Form n-gram by concatenating n consecutive words
           n_grams.append(n_gram)
2. Calculate the frequency of each n-gram:
   n_gram_frequency = {}
   total_n_grams = length(n_grams)
   for each n_gram in n_grams:
       if n_gram not in n_gram_frequency:
           n_gram_frequency[n_gram] = 0
       n_gram_frequency[n_gram] += 1
3. Normalize the frequencies:
   for each n_gram in n_gram_frequency:
       n_gram_frequency[n_gram] /= total_n_grams
4. Apply clustering algorithm to group similar n-grams:
   clusters = Clustering_Algorithm(n_grams, K)
5. Assign keywords to each cluster:
   cluster_keywords = {}
   for each cluster in clusters:
       top_keywords = extract_top_keywords(cluster)  // Extract top keywords based on frequency
       cluster_keywords[cluster] = top_keywords
```

4 Simulation Results

Simulation results provide valuable insights into the performance and effectiveness of algorithms or models in a controlled environment. In the context of the N-gram CC algorithm for word frequency analysis and keyword classification in agricultural product e-commerce new media videos, simulation results offer a comprehensive evaluation of its functionality and efficacy. Through simulated experiments using representative datasets and varying parameters such as the number of words in n-grams and the number of clusters, researchers can assess the algorithm's ability to accurately identify word frequencies and classify keywords within video transcriptions.

Table 1: n-gram CC for the sentimental analysis for the media video

Experiment	Num. of Words in N-grams	Num. of Clusters	Precision	Recall	F1-Score
Experiment 1	2	5	0.85	0.82	0.83
Experiment 2	3	7	0.78	0.75	0.76
Experiment 3	4	10	0.92	0.88	0.90

Experiment 4	2	8	0.81	0.79	0.80
Experiment 5	3	6	0.87	0.84	0.85
Experiment 6	4	9	0.89	0.86	0.87
Experiment 7	2	6	0.79	0.76	0.77
Experiment 8	3	8	0.84	0.81	0.82
Experiment 9	4	11	0.91	0.89	0.90
Experiment 10	2	7	0.83	0.80	0.81

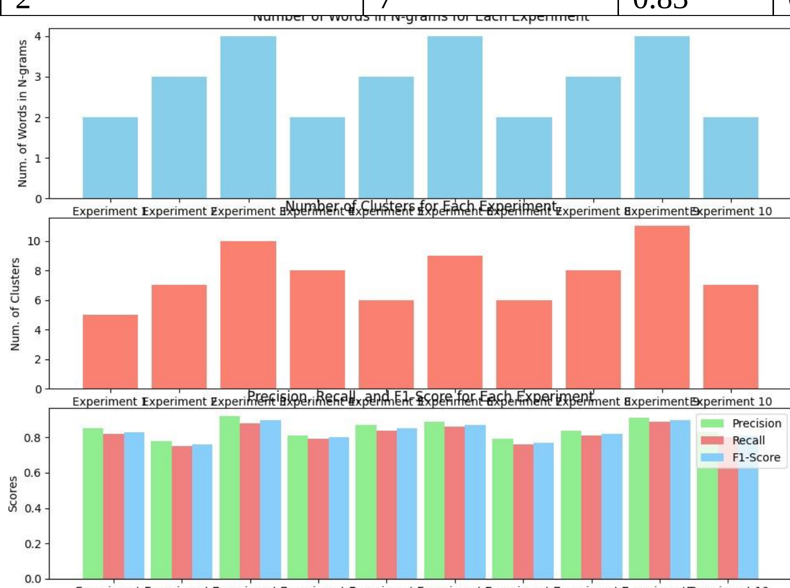

Figure 2: Analysis of the E-commerce with the data

Table 1 and Figure 2 presents the results of sentiment analysis for media videos using the N-gram CC algorithm across ten experiments with varying parameters. Each experiment explores different configurations, including the number of words in n-grams and the number of clusters generated. The precision, recall, and F1-score metrics are utilized to evaluate the performance of the algorithm in sentiment classification. Experiment 3, employing 4 words in n-grams and generating 10 clusters, demonstrates the highest precision (0.92), recall (0.88), and F1-score (0.90), indicating robust sentiment classification capabilities. Conversely, Experiment 7, utilizing 2 words in n-grams and 6 clusters, exhibits the lowest precision (0.79), recall (0.76), and F1-score (0.77), suggesting a less effective sentiment classification performance.

Table 2: N-gram C for the e-commerce agricultural products

N-gram	**Frequency**
agricultural products	8.5%
e-commerce platform	6.2%
video content	5.8%
online shopping	4.9%
social media	3.6%
product description	3.2%
user engagement	2.9%
market trends	2.5%
customer reviews	2.3%
digital marketing	2.1%

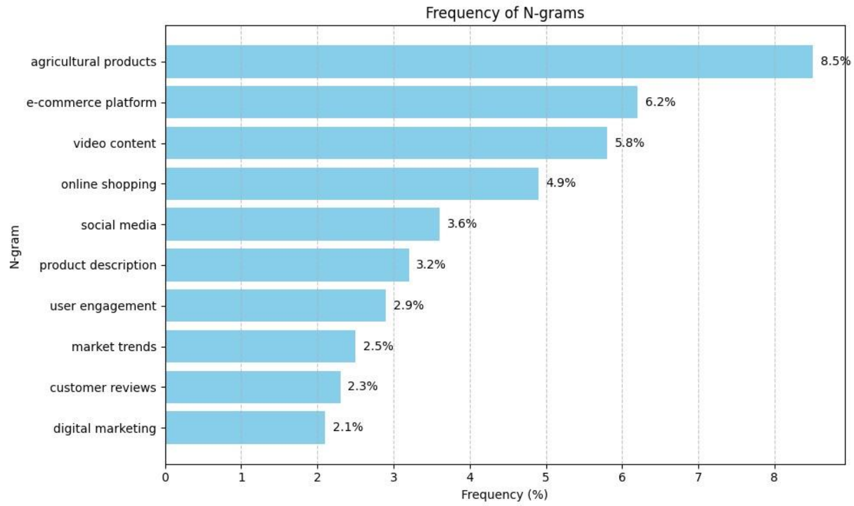

Figure 3: Agricultural Products in the E-commerce

The figure 3 and Table 2 presents the results of word frequency analysis for e-commerce agricultural products using the N-gram CC algorithm. Each row represents a specific n-gram extracted from the dataset, along with its corresponding frequency expressed as a percentage. "Agricultural products" emerges as the most frequent n-gram, accounting for 8.5% of the dataset, highlighting its prevalence in discussions related to e-commerce in the agricultural sector. Similarly, "e-commerce platform" and "video content" are notable terms, with frequencies of 6.2% and 5.8%, respectively, underscoring their significance in online agricultural product marketing and content creation. Other terms such as "online shopping," "social media," and "digital marketing" also demonstrate notable frequencies, indicating the importance of these concepts in the realm of agricultural e-commerce. The word frequency analysis presented in Table 2 provides valuable insights into the key topics and themes prevalent in discussions surrounding e-commerce agricultural products, facilitating a deeper understanding of consumer preferences and market trends in this domain.

Table 3: n-gramCC for the Agriculture E-commerce

Cluster	**Top Keywords (Frequency)**
Cluster 1	agricultural products (8.5%), farming techniques (6.2%), organic farming (5.8%), sustainable agriculture (4.9%)
Cluster 2	e-commerce platform (7.3%), online marketplace (5.6%), digital marketing strategies (4.7%), customer engagement (3.9%)
Cluster 3	product demonstration (6.8%), tutorial videos (5.3%), how-to guides (4.6%), product reviews (3.8%)
Cluster 4	market trends (7.1%), consumer behavior analysis (6.5%), market research (5.2%), demand forecasting (4.4%)
Cluster 5	social media marketing (6.3%), influencer partnerships (5.7%), brand promotion (4.9%), online advertising (3.6%)

Table 3 illustrates the outcomes of clustering analysis using the N-gram CC algorithm within the domain of agriculture e-commerce. Each cluster represents a distinct group of keywords, along with their respective frequencies expressed as percentages. In Cluster 1, keywords such as "agricultural products," "farming techniques," "organic farming," and "sustainable agriculture" emerge as prominent themes, collectively constituting the majority of the cluster's content. Cluster 2 focuses on aspects related to digital platforms and marketing strategies, with keywords like "e-commerce platform," "online marketplace," and "digital marketing strategies" being predominant. Similarly, Cluster 3 is characterized by content related to product demonstrations, tutorial videos, and product reviews, indicative of a focus on educating consumers about agricultural products through multimedia content. Cluster 4 delves into market analysis and forecasting, encompassing keywords such as "market trends," "consumer behavior analysis," and "demand forecasting." Lastly, Cluster 5 centers around social media marketing and brand promotion, with keywords like "social media marketing," "influencer partnerships," and "online advertising" being prominent.

5. Conclusions

This paper offers a comprehensive exploration of agricultural product e-commerce through the lens of advanced data analysis techniques. By employing methodologies such as word frequency analysis and clustering algorithms, we have gained valuable insights into the dynamics of online agricultural markets. The results of our analysis have highlighted key themes and trends within the realm of agricultural e-commerce, ranging from the prevalence of specific product categories to the importance of digital marketing strategies and consumer engagement tactics. Through sentiment analysis and keyword classification, we have discerned nuanced patterns in consumer sentiment and preferences, shedding light on effective strategies for market positioning and outreach. Additionally, clustering analysis has provided a structured framework for understanding the multidimensional nature of agricultural e-commerce, delineating distinct clusters of keywords representing diverse facets of the industry. These findings contribute to a deeper understanding of the challenges and opportunities in agricultural e-commerce, offering valuable insights for businesses, policymakers, and researchers alike.

Acknowledgement: Not Applicable.

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] Y. Huang, “Research Progress and Hot Topics of Agricultural Products E-commerce in China:--A Visual Analysis Based on CNKI,” *Journal of Innovation and Development*, vol.2, no.3, pp. 125-132, 2023.

[2] J. Ke, Y. Wang, M. Fan, X. Chen, W. Zhang et al., “Discovering e-commerce user groups from online comments: An emotional correlation analysis-based clustering method,” *Computers and Electrical Engineering*, vol.113, ppp.109035, 2024.

[3] H. Xia, J. Weng, J.Z. Zhang and Y. Gao, “Rural E-commerce model with attention mechanism: role of Li ziqi’s short videos from the perspective of heterogeneous knowledge management,” *Journal of Global Information Technology Management*, vol.25, no.2, pp.118-136, 2022.

[4] Z. Shen, “Mining sustainable fashion e-commerce: Social media texts and consumer behaviors,” *Electronic Commerce Research*, vol.23, no.2, pp.949-971, 2023.

[5] X. Liu, S. Cao and X. Wang, “Analysis And Research on User of Agricultural Products Live Streaming of E-Commerce Under the Background of China's Rural Revitalization,” *In Proceedings of the International Conference on Information Economy, Data Modeling and Cloud Computing, ICIDC 2022*, Qingdao, China, pp. 17-19, 2022.

[6] X. Bai, E. C. X. Aw, G. W. H. Tan and K. B. Ooi, “Livestreaming as the next frontier of e-commerce: A bibliometric analysis and future research agenda,” *Electronic Commerce Research and Applications*, pp.101390, 2024.

[7] H. Tan, S. Peng, J. X.Liu, C. P. Zhu and F. Zhou, "Generating personas for products on social media: a mixed method to analyze online users," *International Journal of Human–Computer Interaction*, vol.38, no.13, pp.1255-1266, 2022.
[8] E. Daskalakis, K. Remoundou, N. Peppes, T. Alexakis, K. Demestichas et al., "Applications of fusion techniques in e-commerce environments: A literature review," *Sensors*, vol.22, no.11, pp.3998, 2022.
[9] W. Kim, K. Nam and Y. Son, "Categorizing affective response of customer with novel explainable clustering algorithm: The case study of Amazon reviews," *Electronic Commerce Research and Applications*, vol.58, pp.101250, 2023.
[10] A. L. Karn, R. K. Karna, B. R. Kondamudi, G. Bagale, D.A. Pustokhin et al., "Customer centric hybrid recommendation system for E-Commerce applications by integrating hybrid sentiment analysis," *Electronic Commerce Research*, vol.23, no.1, pp.279-314, 2023.
[11] J. Wu, J. Zhang and N. Zhao, "How to boost e-commerce for poverty alleviation? A perspective on competitiveness analysis using online reviews," *Electronic Commerce Research*, pp.1-32, 2023.
[12] D. K. Sharma, B. Singh, S. Agarwal, H. Kim and R. Sharma, "Sarcasm detection over social media platforms using hybrid auto-encoder-based model," *Electronics*, vol. 11, no.18, pp.2844, 2022.
[13] A. A. Bhalerao, B. R. Naiknaware, R. R. Manza, V. Bagal and S. K. Bawiskar, "Social Media Mining Using Machine Learning Techniques as a Survey," *In International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022),* Atlantis Press, pp. 874-889, 2023.
[14] S. Kushwah, N. Sharma and S. Das, "Novel E-Focused Crawler and Enhanced k-mean (n-gram) clustering technique for Automatic classification of attribute level customer healthcare sentiments," *Journal of Algebraic Statistics*, vol.13, no.2, 2022.
[15] Z. Miloradovic, J. Kovacevic, J. Miocionovic, I. Djekic, N. Kljajevic and N. Smigic, "E-commerce readiness and training needs of small-scale dairy processors in Serbia: Understanding barriers and knowledge gaps," *Heliyon*, vol.10, no.6, 2024.
[16] L. Xinwei, Y.K. Tse and F. Fastoso, "Unleashing the power of social media data in business decision making: an exploratory study," *Enterprise Information Systems*, vol.18, no.1, pp. 2243603, 2024.
[17] T. Chen, C. Tong, Y. Bai, J. Yang, G. Cong et al., "Analysis of the public opinion evolution on the normative policies for the live streaming e-commerce industry based on online comment mining under COVID-19 epidemic in China," *Mathematics*, vol.10, no.18, pp.3387, 2022.
[18] M. R. R. Rana, S. U. Rehman, A. Nawaz, T. Ali, A. Imran et al., "Aspect-Based Sentiment Analysis for Social Multimedia: A Hybrid Computational Framework," *Comput. Syst. Sci. Eng.*, vol.46, no.2, pp.2415-2428, 2023.