# Opinion Mining on E-commerce Live Broadcast of Agricultural Products Based on WLDA-Apriori Model

Binghao Wang*, Yifan Wang†, Zeqi Wang†
* Chengdu City Construction Investment and Management Group Co., Ltd, Chengdu, China
† Wuhan University, Wuhan, China
yifwang@whu.edu.cn

*Abstract*—**With the popularization of information technology in rural society, the e-commerce live broadcast of agricultural products has provided an impetus for agricultural marketing from rural to urban areas. But as an emerging industry, it has problems in live broadcast content, marketing strategies, market supervision, industry norms and agricultural aid effects. In order to fully understand the public opinion and promote the sustainable development of this industry, this paper proposes an opinion mining method based on WLDA-Apriori model. The method combines topic model with association rule algorithm, which can further analyze the implicit semantic relations among topic words. This paper also conducts the empirical study with social media data from Weibo, and obtains eight topics the Chinese public concerned. Based on topic words and their relations, the public opinion are identified and discussed from aspects of Industry Development, Market Effect, Live Broadcast Subject and Negative Effect. By encompassing as many netizens as possible, the research scope of this paper surpasses the constraints of conventional questionnaire surveys and case studies, enhancing the generalizability of the research findings and recommendations. But this paper does not integrate and analyze social media data from multiple platforms, nor does it analyze public opinion from the sentiment level.**

***Keywords*—public opinion, e-commerce live broadcast, agricultural products, topic model, association rule**

## I. Introduction

In the era of mobile Internet, e-commerce live broadcast has become a new online marketing method in which merchants interact with consumers instantly in order to achieve the goals of gathering popularity, increasing sales and building brands. With the continuous expansion of information technology application in rural, e-commerce live broadcast has gradually become one of the main ways to sell agricultural products. Compared with the image display provided by traditional e-commerce, the e-commerce live broadcast of agricultural products enable users to perceive information from more aspects, which has a positive impact on users' purchasing behavior. Especially since the outbreak of the COVID-19 pandemic in early 2020, the traditional circulation channels of agricultural products have been blocked, and the sales of a large number of fresh agricultural products have stagnated. With its innovative advantages in the transaction mode, e-commerce live broadcast of agricultural products has realized the efficient match between the agricultural supply and demand information, and become an important mode of helping farmers and poverty alleviation. In Q1 2020, online sales of agricultural products in China surpassed 90 billion yuan (around 13.9 billion US dollars), while the number of live broadcasts for agricultural e-commerce exceeded 4 million. [1].

However, as an emerging industry, the e-commerce live broadcast of agricultural products needs to be improved in terms of live broadcast content, marketing strategies, market supervision, industry norms and agricultural aid effects. How to give full play to its role and realize the sustainable development of this industry is a common concern of the government, enterprises and academia. Previous research has predominantly utilized questionnaires [2], mathematical models [3], and case analyses [4] to investigate the factors impacting the development of agricultural products e-commerce live broadcasting and suggest corresponding recommendations. Due to the limited scope of the questionnaire audience and cases, these studies could only meet the needs of specific populations, and the general applicability is relatively poor. At the same time, social media provides a rich source of information, including the basic opinions of the public on some issues. Making good use of this resource can help managers fully understand the public's expectations for the e-commerce live broadcast of agricultural products, and clarify the priorities of various matters to serve as an important basis for future work.

From the perspective of public opinion mining, this paper adopts topic model and association rule algorithm to carry out topic extraction and topic word association on the social media data of agricultural products e-commerce live broadcast. On this basis, this paper identifies the public opinion on the e-commerce live broadcast of agricultural products, and discusses the meanings or problems behind them. The research questions in this paper are as follows:

- RQ1: How to combine topic model and association rule algorithm to mine public opinion?
- RQ2: What are the topics concerned by the Chinese public about the e-commerce live broadcast of agri-

This paper is supported by the Chinese NSFC Major Program (Grant No. 72074172). Yifan Wang is the corresponding author.

cultural products?

- RQ3: What are the Chinese public opinion on the e-commerce live broadcast of agricultural products?

## II. Related Work

### A. E-commerce Live Broadcast of Agriculture Products

With the emergence of e-commerce live broadcast of agricultural products, researchers have studied its significance and mechanism. Liu et al. [2] took the rural e-commerce in Xinhua County as the research object and conducted a questionnaire survey. They found that under the influence of COVID-19, the e-commerce live broadcast had become the main way to promote poverty alleviation in counties. Lu [5] conducted empirical analysis on the questionnaire data issued by sellers selling navel oranges through e-commerce live broadcasts. It turned out that e-commerce live broadcast can effectively drive the development of fresh agricultural products e-commerce. Zeng et al. [6] demonstrated the internal mechanism of agricultural marketing from rural to urban areas shaped by e-commerce live broadcast. It is specifically reflected that e-commerce live broadcast transforms the agricultural marketing into short-chain value transmission, multi-value output and value co-creation.

In addition, researchers have also studied the influence factors and development countermeasures of e-commerce live broadcast effect of agricultural products. Some studies explored the influencing factors of consumers' purchase intention in e-commerce live broadcast, and made suggestions on e-commerce live broadcast of agricultural products. Shi et al. [3] presented a model for precise marketing classification of agricultural products using clustering, and used customer development space to achieve precise marketing for the e-commerce live broadcast platform. Combined with Stimulus Organism Response and Perceived Value Theory, Yang et al. [7] constructed the model of consumers' purchasing intention of fresh agricultural products under the live broadcast situation, and used SPSS and AMOS to verify the reliability and validity of the questionnaire data. Other studies innovatively reveal the influencing factors of whether the live broadcast of agricultural products can be adopted from the perspective of operators. For instance, Gao and Sheng [4] carried out an empirical study on the influencing factors and mechanism of e-commerce live broadcast adoption intention of new agricultural operation entities. The result showed that the acknowledge factors, subject factors and operation factors of the operators are the key influencing factors.

Existing studies have demonstrated the value and mechanism of agricultural products e-commerce live broadcast from the aspects of poverty alleviation, rural revitalization and e-commerce development. Besides, from the perspective of consumers and operators, they have respectively explored the factors affecting the development of agricultural products e-commerce live broadcast, and put forward suggestions for the government, industry and various participants. However, due to the limitations of data and experimental methods, the existing research fails to view from the perspective of public opinion, and lacks of general applicability, which needs to be further supplemented.

### B. Opinion Mining in Social Media

Text mining technology has been extensively applied and validated in various domains such as big data analysis in finance [8], news text classification [9], sentiment analysis on online platforms [10], and many others. Public opinion mining is another significant application of text mining technology, which involves analyzing text to identify and extract opinions expressed by authors or groups regarding a particular object or subject from a vast amount of information. In the context of social media, opinion mining refers to the process of extracting sensitive or highly relevant information from social media data, analyzing public sentiment and trends, and providing guidance for decision-making. [11]. Because public perspective has long been ignored in ex-post evaluation of large cross-regional projects, Wan et al. [12] used lexicon-based sentiment analysis and LDA topic modeling technologies to identify the patterns of public sentiments and topics from spatial and temporal perspectives. Fu et al. [13] used the GDELT and TWITTER platforms' data to mine the public opinion on online learning during the COVID-19 pandemic and found that there was a difference between the topics that the media focused on and the content that the audience talked about.

Researchers have mainly mined public opinion from the topic or sentiment level so far. Among them, opinion identification at the topic level is only based on separate topic words, and the implicit semantic associations among topic words have not been found. Besides, The method of combining topic model with association rule algorithm has been applied to other research problems to improve the interpretability of topics. Wang [14] constructed a mining method for ideological and political knowledge elements based on LDA model and Apriori algorithm. Ruan et al. [15] used LDA model to extract topics from text data, and then used association rules to enhance the semantics of words describing topics, so as to mine the semantic associations of text topics. Wu et al. [16] combined BTM (Biterm Topic Model) with association rules to conduct topic discovery on COVID-19 vaccine.

## III. Methodology

This paper proposes an opinion mining method based on WLDA-Apriori model. This method can be divided into the following steps. First, construct the WLDA topic model weighted by TF-IDF algorithm, and extract topic information from social media data. In this step, each topic and its separate topic words will be obtained. Next, based on the Apriori association rule algorithm, the association rules among separate topic words are mined to obtain the implicit semantic relations. Finally, the public opinion of each topic will be identified by analyzing the topic words and the semantic relations among them.

## A. Weighted LDA Topic Model

LDA (Latent Dirichlet Allocation) [17] is a three-layer Bayesian probability model that includes words, topics and documents. It obtains topic features through the probability distribution of topics, which can be used to identify topic information hidden in large-scale corpora. The LDA model uses Gibbs sampling method to estimate the posterior distribution of topics, but this method causes the distribution of topics and words to incline to high-frequency words rather than words that can represent documents, thus reducing the model's ability to describe text information. Besides, deleting stop words can not completely filter out these high-frequency words with poor representation.

To solve this problem, Wilson et al. [18] have proved that reasonable weighting strategies based on LDA model will enable the natural allocation of high-frequency words. The current weighting strategies mainly use the text style structure or the distribution characteristics of items. The latter introduces effective mathematical functions into the representation model by considering the distribution characteristics of items, so that the term set is not submerged by a large number of high-frequency words. TF-IDF (Term Frequency - Inverse Document Frequency) is a widely used weight calculation formula, and previous studies [19], [20] have combined LDA topic model with TF-IDF to improve the accuracy of text representation. TF-IDF considers the frequency of terms on the one hand, and the inverse document frequency on the other. The specific formula is as follows:

$$\mathrm{WT}\left(w_{ij}\right)=TF\left(w_i\right) * \mathrm{IDF}\left(w_i\right)=tf_{ij} * \log \left(\frac{n}{n_i+\mu}\right) \tag{1}$$

Where $TF\left(w_i\right)$ calculated by $tf_{ij}$ represents the number of occurrences of term $w_i$ in document $j$, $\mathrm{IDF}\left(w_i\right)$ calculated by $\log\left(\frac{n}{n_i+\mu}\right)$ represents the inverse document frequency of term $w_i$, $n$ is the total number of documents, $n_i$ is the number of documents containing term $w_i$, and $\mu$ is the adjustment factor.

In this paper, the weight value $WT\left(w_{ij}\right)$ based on TF-IDF is assigned to each term and extended to the WLDA model. A new Gibbs sampling formula is formed as follows:

$$\begin{aligned} & p\left(z_{ij}=k \mid z^{-ij}, \alpha, \beta\right) \\ & \propto \frac{\sum_{j=1}^M WT(w_{ij}) n_{ijk}^{-ij}+\beta}{\sum_{i=1}^V \sum_{j=1}^M WT(w_{ij}) n_{ijk}^{-ij}+V\beta} \times \frac{\sum_{i=1}^V WT(w_{ij}) n_{ijk}^{-ij}+\alpha}{\sum_{k=1}^K \sum_{i=1}^V WT(w_{ij}) n_{ijk}^{-ij}+K\alpha}\end{aligned} \tag{2}$$

Where $z_{ij}$ represents the topic variable of the term $w_i$ in the document $d_j$, $-ij$ represents the term $w_i$ in the excluded document $d_j$, $n_{ijk}$ represents the number of times the term $w_i$ in the document $d_j$ is assigned to the topic $k$, $\alpha$ represents the Dirichlet prior distribution of topics, $\beta$ represents the Dirichlet prior distribution of words, $K$ represents the number of topics, and $V$ represents the total number of terms in the set.

## B. Apriori Association Rule Algorithm

Association rule is a common data mining method, which is originally designed for Market Basket Analysis, and then used to discover potential rules in a large amount of data, so as to explore the correlation among data. Apriori is a classical algorithm for association rule analysis. It adopts a layer-by-layer search method, and is easy to implement. This paper uses Apriori association rule algorithm to mine association rules among topic words.

To illustrate the process of mining association rules using Apriori, the following assumptions are made: the word set is $I$, and any subset of $I$ is called the itemset, which is represented by $X$; the number of words contained in each itemset is called the length of the itemset, and an itemset whose length is $k$ is called the $k$-itemset; the whole corpus is $D$, and $|D|$ indicates the number of documents contained in corpus $D$; the number of occurrences of an itemset in corpus $D$ is called frequency, which is expressed as $count(X)$.

First, generate frequent itemsets based on support. Support indicates the probability of occurrence of itemset $X$ in all itemsets. Given a minimum support, if the support of $X$ is bigger than the minimum support, it is called a frequent itemset. The specific formula is as follows:

$$\operatorname{support}(X)=\operatorname{count}(X)/|D| \tag{3}$$

Second, generate strong association rules based on confidence. Confidence indicates the possibility of containing the feature word $Y$ in the item set containing the feature word $X$. Given a minimum confidence, when the confidence is bigger than the minimum confidence, the corresponding rule is called a strong association rule. The specific formula is as follows:

$$\text{confidence }(X \Rightarrow Y)=\operatorname{support}(XY) / \text{ support }(X) \tag{4}$$

Third, filter effective strong association rules based on lift. Lift indicates the ratio of the probability of containing the term $Y$ at the same time with the term $X$ to the probability of the occurrence of the term $Y$. Association rules that meet the minimum support and minimum confidence need to calculate corresponding lift. Generally, when the lift is bigger than 3, the association rule can be considered as a effective strong association rule. The specific formula is as follows:

$$\text{lift }(X \Rightarrow Y)=\operatorname{support}(XY) / \text{ support }(X) \text{ support }(Y) \tag{5}$$

# IV. Result & Discussion

## A. Data Preparation

With an increasing number of users turning to it as their primary channel for event exposure and emotional expression, Weibo has become a platform in China that has a significant impact on shaping public opinion. Therefore, this paper selected Weibo as the data source, and used "e-commerce live broadcast" combined with "agricultural products", "help farmers" and other related words as keywords to obtain social media data. At the same time, in order to better reflect the users' opinion, only original microblogs were collected. In August 2022, 8488 microblogs related to agriculture products e-commerce live broadcasts were collected, which included information such as the texts, release time, and publishers. These microblogs spanned from September 2017 to August 2022. The collected data underwent preprocessing, which involved segmenting the microblog texts using the Jieba toolkit in Python, with continuous improvement through the addition of a custom dictionary. A stopword dictionary was also created to remove irrelevant words. As a result of these steps, a social media dataset about agriculture products e-commerce live broadcasts was created.

### B. Topic Extraction

Based on the Gensim toolkit in Python, this paper introduced TF-IDF algorithm and constructed a weighted LDA model to achieve topic extraction of the data set. After the calculation of perplexity, it is found that when the number of topics is 8, the perplexity is lower. Therefore, this paper set the optimal number of topics $K$ to 8. Other experimental parameters were set as follows: $\alpha$ is $50/K$, $\beta$ is 0.01, and the number of iterations is 100.

After using the weighted LDA model to extract topics from the data set, eight topics and the term distribution of each topic were obtained. Under each topic, ten words with high probability were selected as the topic words, and the topic name was inferred according to the topic words. The results are shown in TABLE I.

TABLE I: Concerned Topics Related to E-commerce Live Broadcast of Agricultural Products

| ID | Topic Name | Topic Words |
|---|---|---|
| 1 | Industry Convergence | development, industry, construction, agriculture, the masses, promotion, rural, tourism, base, service |
| 2 | Rise of E-commerce Live Broadcast Industry | rural vitalization, activity, poverty alleviation, characteristic, culture, practice, sales, in field, youth, increase revenue |
| 3 | Participation of Public Welfare Forces | corn, Yugang Li, homeland, hairy crab, support for farmers, price, public welfare, deputy to the People's Congress, Back to Field, CCTV |
| 4 | Employment and Entrepreneurship | enterprise, employment, training, innovative undertaking, reward, support, policy, migrant worker, innovation, talent |
| 5 | Market and Consumption | consumption, digitization, epidemic situation, growth, prevention and control, market, economy, data, accelerate, technology |
| 6 | Participation of Well-known Enterprises or Stars | Oriental Selection, New Oriental, Minhong Yu, background, fake, Yuhui Dong, garlic, Zhan Xiao, challenge, fans |
| 7 | Violation Behavior | foodstuff, food safety, problem, renovate, production, illegal, fake, quality, propagate, joint action |
| 8 | Damage to Consumers' Rights and Interests | weather, mildew and rot, New Oriental, grape, customer, peach, fresh food, unsalable, complaint, refund |

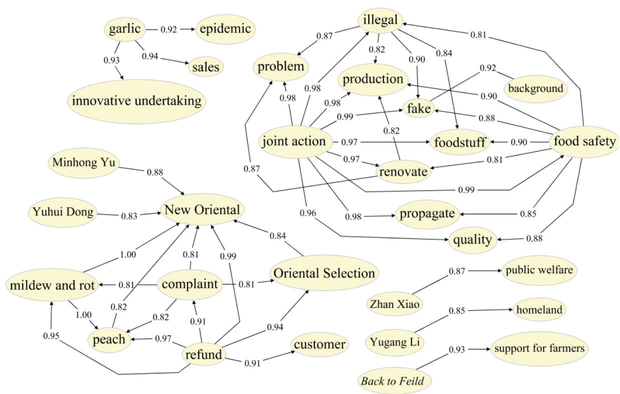


Fig. 1: Network of binomial association rules for topic words.

### C. Topic Word Association

This paper extracts association rules between subject words using the Apriori algorithm with a minimum support of 0.01, minimum confidence of 0.8, and minimum lift of 3. We obtained 1009 strong association rules (48 binomial, 961 trinomial), selected representative rules, and visualized the association rules network using CmapTools. Fig. 1 shows a network of binomial association rules for topic words. Each node represents a topic word, each directed edge represents the direction of an association rule, and the number on each edge represents the confidence of the association rule. We can see that there are lots of separate binomial association rules. For example, under the condition that the topic word "Back to Field" (a Chinese variety show) appears, there is 0.93 probability that "support for farmers" will appear. There are even several networks among the topic words. For example, the network centered at "New Oriental", "complaint" and "refund" reflects that consumer complaints and refunds caused by the rotten peaches sold by New Oriental; the network centered at "illegal", "food safety", "joint action", "fake" and "renovate" reflects that a series of joint regulatory actions have been taken on account of the food safety and false publicity.

### D. Public Opinion Identification

In this paper, the eight topics obtained from the experiment were summarized into four aspects: Industry Development, Market Effect, Live Broadcast Subject and Negative Effect. The public opinion in each topic were identified according to the topic words and the association rules among them. The results are shown in TABLE . According to the above topic words and their associations, this paper identifies the public opinion under each topic, as shown in Table II. At the same time, this paper divided the eight topics into four aspects: Industry Development, Market Effect, Live Broadcast Subject and Negative Effect. And the specific meaning of public opinion were explained and discussed from these aspects.

TABLE II: Public Opinion on the E-commerce Live Broadcast of Agricultural Products

| Aspect | Topic | Public Opinion |
|---|---|---|
| Industry Development | Rise of E-commerce Live Broadcast Industry | to increase farmers' incomes, Conducive to poverty alleviation and rural revitalization, to propagate local characteristics and culture, to attract young people to the countryside |
| | Industry Convergence | to promote the integration and upgrading of traditional industries, to do practical work for the people, to improve rural service level |
| Market Effect | Market and Consumption | to implement poverty alleviation through consumption, to revitalize the agricultural product market, to take advantage of the potential of the sinking market, to integrate technology and other resources, to achieve economic growth |
| | Employment and Entrepreneurship | to support farmers as streamers, to increase employment opportunities, to carry out entrepreneurship, to issue incentive policies, to provide training |
| Live Broadcast Subject | Participation of Public Welfare Forces | to support farmers, to speak for the agricultural products of hometown, to lower the price |
| | Participation of Well-known Enterprises or stars | to launch the business of agricultural products e-commerce live broadcast, to attract followers, to raise doubts |
| Negative Effect | Violation Behavior | false propaganda, cannot guarantee the quality, food safety problems, to conduct joint action |
| | Damage to Consumers' Rights and Interests | mildew and rot, complaints, to request a refund, loss of trust |

Industry Development mainly discusses the influence brought by *Rise of E-commerce Live Broadcast Industry* and *Industry Convergence*, such as rural revitalization, cultural export, talent introduction and upgrading of traditional industries. The marketing from rural to urban areas of agricultural products driven by the e-commerce live broadcast has increased farmers' income and promoted the upgrading of local industries, which is of great significance for the implementation of strategies such as poverty alleviation and rural revitalization. The unique live scenes of agricultural products in the field can better present agricultural products and bring more intuitive experience to users [21], which increased the product reliability and users' purchase intention. Besides, this way also plays a role in the promotion of local culture and scenery. However, the current live broadcasts are single in form and repetitive in content, which is difficult to highlight the local characteristics and cultural depth.

Market Effect includes *Market and Consumption* and *Employment and Entrepreneurship*, focusing on the phenomenon of market sinking and consumption upgrading brought by the e-commerce live broadcasts of agricultural products, as well as their role in market activation, economic growth and technological development. The e-commerce live broadcast of agricultural products enabled many links of agricultural products to be integrated into the Internet, and a large number of advantageous resources (such as talents, capital and technology) poured in [22]. This has led to the revitalization of the rural market. At the same time, e-commerce platforms have begun to use the huge potential of sinking markets, expanded the market of agricultural products. Due to the expansion of agricultural products market and low threshold of e-commerce live broadcast, farmer streamers came into being. There are also farmers, migrant workers or the young generation [23] who start their own businesses through e-commerce live broadcast of agricultural products. In addition, local governments and live broadcast platforms also encourage farmer streamers through a series of measures, and provide public welfare training opportunities, which has brought a virtuous circle to the development of this industry [24]. On the other hand, it is not uncommon to fail to bring about an increase in real income to operators, or even cause losses.

Live Broadcast Subject mainly discusses the advantages and doubts caused by the participation of multiple live broadcast subjects such as *Participation of Public Welfare Forces* and *Participation of Well-known Enterprises or stars*. Public welfare forces mainly use low prices and subsidies to promote unsalable agricultural products or advocate for agricultural products in their own hometown, of which officials' live broadcasts is a typical representative. Officials' live broadcasts means that local officials act as the streamers to endorse local products, so as to facilitate the export of local agricultural products or the promotion of cultural and tourism services. In addition, the participation of well-known enterprises and stars has brought huge flow and purchasing power to the e-commerce live broadcast of agricultural products [25]. But it is difficult for these streamers to promote specific agricultural products for a long time. The participation of these subjects has also raised doubts. For instance, the demand for agricultural products has surged in a short period, resulting in temporary production and hasty packaging, which could affect consumers' purchasing experience [26].

Negative Effect includes *Violation Behavior* and *Damage to Consumers' Rights and Interests*, and discusses the problems arising from obstructed agricultural sales processes and inadequate market regulation. On the one hand, Due to the lack of a mature market supervision mechanism, a series of violations (such as false publicity, incomplete product information and food safety problems) occur in the e-commerce live broadcast of agricultural products, which may affect consumers' health in serious cases. Besides, A series of joint regulatory actions have been carried out to monitor these issues. On the other hand, whether because of substandard quality at the source or problems during transportation, agricultural products purchased are easy to become stale. This may lead to a series of consumer rights protection behaviors such as complaints and refunds, and even reduce consumers' trust in e-commerce live broadcast of agricultural products [25].

## V. Conclusions

To gain a comprehensive understanding of public opinion regarding e-commerce live broadcasting of agricultural products, this paper has undertaken two key initiatives. Firstly, a public opinion mining approach based on the WLDA-Apriori model is presented. This approach employs a combination of topic modeling and association rule algorithms, enabling the analysis of implicit semantic relationships between subject words and increasing interpretability. Secondly, an empirical study is conducted using social media data from Weibo. The results show that the topics of Chinese public concern are *Industry Convergence*, *Rise of Live Broadcast Industry*, *Participation of Public Welfare Forces*, *Employment and Entrepreneurship*, *Market and Consumption*, *Participation of Well-known Enterprises or stars*, *Violation Behavior* and *Damage to Consumers' Rights and Interests*.

Taking public opinion into account, this paper provides an analysis of the current state and prevalent issues surrounding e-commerce live broadcast in the agricultural industry.Compared with traditional questionnaires and case studies, the results of this paper are more universal, and the suggestions based on these will be more consistent with the needs of the public. However, this article is subject to a limitation whereby the data is derived from a sole platform. Unlike social media platforms, users on e-commerce platforms such as Douyin and Taobao exhibit a greater proclivity towards live-streamed content, live-streamed formats, and product quality, which results in data that encompasses semi-structured and unstructured formats. Thus, it is imperative to incorporate these additional data sources to facilitate a more comprehensive interpretation of user demand. Furthermore, incorporating emotional tendencies in opinion mining can provide a more comprehensive representation of public sentiment. Future research should strive to construct a multidimensional opinion mining model that accounts for topics, emotions, and other relevant factors.

## References


[1] W. Shang, "More than 4 million live e-commerce broadcasts in the first quarter," *The People's Daily*, no. 19, 2020.

[2] K. Liu, B. Liu, H. Xu, Y. He, and Y. Cao, "Research on e-commerce live broadcasts helping poverty alleviation under the influence of the covid-19:———take xinhua county, hunan province as an example," in *2020 2nd International Conference on Economic Management and Model Engineering (ICEMME)*. IEEE, 2020, pp. 858–865.

[3] Y. Shi, X. Li, S. ZhouGong, X. Li, and H. Wang, "Precise marketing classification of agricultural products for e-commerce live broadcast platform using clustering," *Mobile Information Systems*, vol. 2022, 2022.

[4] K. Gao and Y. Sheng, "Research on the influencing factors and mechanism of livestreaming commerce adoption intention of new agricultural operation entities," *China Business and Market*, vol. 35, no. 10, p. 9, 2021.

[5] Z. Lu, "Webcasting and the development of online retailers of fresh agricultural products: Driving mechanism and empirical test," *China Soft Science*, 2021.

[6] Y. Zeng, C. Ma, L. Li, and H. Guo, "Live streaming e-commerce and the value recreation of agricultural marketing from rural to urban areas: Mechanism and implementation paths," *Issues in Agricultural Economy*, no. 2, p. 10, 2022.

[7] W. Yang, X. Yang, Y. Bai, H. Yi, M. He, and X. Li, "Research on purchase intention of fresh agricultural products based on sor theory under live broadcast situation," in *2021 2nd International Conference on E-Commerce and Internet Technology (ECIT)*. IEEE, 2021, pp. 416–424.

[8] S. Pejić Bach, Krstić and L. Turulja, "Text mining for big data analysis in financial sector: A literature review," *Sustainability*, no. 11, p. 5, 2019.

[9] A. Kadriu and L. Abazi, "A comparison of algorithms for text classification of albanian news articles," *Proceedings of the ENTRENOVA - ENTerprise REsearch InNOVAtion Conference*, no. 3, pp. 1–7, 2017.

[10] B. Nguyen, V.-H. Nguyen, and T. Ho, "Sentiment analysis of customer feedback in online food ordering services," *Business Systems Research Journal*, no. 12, pp. 46–59, 2021.

[11] K. Guo, L. Shi, W. Ye, and X. Li, "A survey of internet public opinion mining," in *2014 IEEE international conference on progress in informatics and computing*. IEEE, 2014, pp. 173–179.

[12] X. Wan, R. Wang, M. Wang, J. Deng, Z. Zhou, X. Yi, J. Pan, and Y. Du, "Online public opinion mining for large cross-regional projects: Case study of the south-to-north water diversion project in china," *Journal of Management in Engineering*, vol. 38, no. 1, p. 05021011, 2022.

[13] Z. Fu, M. Yan, C. Meng, W. Wang, X. Hu, Y. Li, J. Wang, Z. He, and Z. Wang, "Opinion mining about online education basing on gdelt and twitter data," in *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. IEEE, 2020, pp. 741–745.

[14] L. Wang, "Research on the mining of ideological and political knowledge elements in college courses based on the combination of lda model and apriori algorithm," *Applied Mathematics and Nonlinear Sciences*.

[15] G. Ruan and L. Xia, "Mining document topics based on association rules," *Data Analysis and Knowledge Discovery*, no. 12, pp. 50–56, 2016.

[16] P. Wu, T. Shi, and C. Ling, "Topic discovery of covid-19 vaccines based on weibo platform," *Information Science*, vol. 40, no. 7, p. 8, 2022.

[17] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[18] A. Wilson and P. A. Chew, "Term weighting schemes for latent dirichlet allocation," in *human language technologies: The 2010 annual conference of the North American Chapter of the Association for Computational Linguistics*, 2010, pp. 465–473.

[19] N. Lee, E. Kim, and O. Kwon, "A real-time combination of text-mining methods for flexible movie recommendation in human robot interaction," in *2016 International Conference on Platform Technology and Service (PlatCon)*. IEEE, 2016, pp. 1–6.

[20] L. Posch, M. Panahiazar, M. Dumontier, and O. Gevaert, "Predicting structured metadata from unstructured metadata," *Database*, vol. 2016, 2016.

[21] M. Zan and Z. Wang, "E-commerce live streaming of agricultural products: a new mode of poverty alleviation," *Issues in Agricultural Economy*, no. 11, p. 10, 2020.

[22] L. Song and X. Ni, "E-commerce consumption for poverty alleviation helps agricultural products out of difficulties," *People's Tribune*, no. 29, p. 2, 2020.

[23] Z. Wang, "How "live broadcast + e-commerce" helps rural revitalization," *People's Tribune*, no. 15, p. 2, 2020.

[24] J. Dai and W. Luo, "Research on the key path to improve user stickiness under the scenario of agricultural products live streaming," *China Business and Market*, vol. 36, no. 5, p. 12, 2022.

[25] Y. Yuan and W. Zhang, "The dilemma and its solution of rural livestreaming marketing from the perspective of social capital," *China Business and Market*, vol. 35, no. 10, p. 8, 2021.

[26] H. Guo and J. Qv, "Study on the sustainable development of e-commerce live broadcast assisting agriculture," *People's Tribune*, no. 20, p. 3, 2020.