VNUHCM Journal of Economics, Business and Law 2026, 10(2):6806-6818

Open Access Full Text Article

Research Article

Multi-Dimensional Sentiment Analytics of Online Customer Reviews in E-Commerce

Quoc-Tuan Truong[1,2], Hoang-Anh Tran[1,2], Thien Le[1,2], Hoanh-Su Le[1,2], Van-Ho Nguyen[1,2,*]

ABSTRACT

This study presents a model for collecting and analyzing customer reviews using machine learning. As e-commerce platforms grow, large volumes of customer opinions are generated, but most of this data is unstructured and complex to use directly. To address this, the proposed approach begins with gathering reviews from online platforms, followed by data preprocessing to clean, filter, and normalize the text. These steps ensure that the data is suitable for machine learning analysis. Classification models are then applied to perform sentiment analysis. The results show that the model can categorize customer opinions with an accuracy of over 94%. The findings are illustrated through comprehensive graphs and charts, offering clear insights into customer attitudes and behavioral patterns. Such visualizations are crucial in translating complex analytical results into an accessible form, enabling managers to identify key trends and underlying issues rapidly. By presenting data from multiple perspectives, the visual outputs facilitate more profound understanding, support comparative analysis, and enhance the overall quality of decision-making, thereby strengthening evidence-based management practices within dynamic business environments. The study highlights that online customer feedback reflects individual opinions and influences overall business performance. By systematically analyzing customer feedback, companies can identify emerging needs, enhance products and services, and design more effective business strategies. Such insights directly strengthen competitiveness in the e-commerce sector, where user experience is a critical differentiator. Beyond e-commerce, the proposed methodology demonstrates flexibility and applicability across other domains, including finance, healthcare, and education, where capturing and interpreting customer sentiment is equally important. This research establishes a robust and practical framework for transforming unstructured feedback into actionable knowledge by integrating machine learning techniques with multi-dimensional analytics. The framework supports strategic decision-making and provides organizations with a sustainable approach to leveraging data for innovation and long-term value creation.

Key words: machine learning, classification models, customer sentiment analysis, electronic commerce, multi-dimensional analytics

[1]*University of Economics and Law, Ho Chi Minh City, Vietnam*

[2]*Vietnam National University, Ho Chi Minh City, Vietnam*

Correspondence

Van-Ho Nguyen, University of Economics and Law, Ho Chi Minh City, Vietnam

Vietnam National University, Ho Chi Minh City, Vietnam

Email: honv@uel.edu.vn

History

- Received: 04-04-2025
- Revised: 21-08-2025
- Accepted: 14-03-2026
- Published Online: 28-06-2026

DOI : https://doi.org/10.32508/vnuhcmj-ebl.v10i2.1633

INTRODUCTION

The rise of e-commerce websites, as new shopping channels, has led to the rise of review sites for various services and products. This provides the opportunity to use aspect-based sentiment analysis and text-based opinion mining to help consumers decide what to buy and businesses to monitor their reputations and reputations better. Understand the needs of the market. Perspective analysis is a technique that aims to advance research beyond sentiment classification at the sentence or text level[1]. The goal is to identify the opinions expressed about entities and aspects of them. However, few techniques can produce such results based on customer ratings, usually against a limited set of predefined elements rather than free reviews. Another challenge in this process is the cold start problem because of the lack of sufficient evaluation data for a product.

In today's highly competitive market, word of mouth from the buyer's point of view is a vital step towards success for any company. Most businesses in all fields have launched websites to sell their products and services. Millions of reviews, opinions, and emotions are generated daily on online websites about products and services. It isn't easy to process and understand such large amounts of opinion-based data. Emotion analysis is the field of recording and extracting emotions from available opinion data and analyzing processes through natural language processing and text classification. It is a method to understand emotions in text. It is also becoming a challenge in many fields of research, including the field of data mining, because of the rapid increase in the number of web pages consisting of reviews of products and services. Practical data mining helps businesses see and understand customers, capture customer dissatisfaction and feed-

Cite this article : Truong Q, Tran H, Le T, Le H, Nguyen V. **Multi-Dimensional Sentiment Analytics of Online Customer Reviews in E-Commerce**.*VNUHCM J. Econ. Bus. Law.* 2026; 10(2):6806-6818.

back to improve the quality of products and services. This will improve user satisfaction, positively impact business results, and offer better market strategies, helping to cut costs and increase revenue.

This paper clearly articulate the gap in existing literature by highlighting the lack of a comprehensive framework that integrates multi-dimensional sentiment analytics, particularly one that combines sentiment classification with aspect-level, temporal, and spatial analysis in the context of customer feedback. The study's scope emphasizes the focus on the hospitality industry and the use of structured customer feedback data to develop a robust sentiment analytics model that supports strategic decision-making. We underscore the growing volume and importance of user-generated content in digital platforms and the need for more sophisticated analytical approaches. In particular, we will emphasize how multi-dimensional sentiment analysis offers actionable insights that traditional single-layer sentiment models cannot provide, highlighting both the study's theoretical contribution and practical implications in data-driven decision support systems.

The next section of the article is Section 2, which includes related studies in order to shape and identify research models and methods suitable for the set objectives. The proposed research methods and models are described in ection 3. After the experimental process, the results are found, and the discussion is mentioned in ction 4. The last section is the conclusion and the direction of the research development.

RELATED RESEARCH

Opinion Mining

The sentiment analysis process is important in opinion mining[2]. During the classification process, customer opinions about the product are considered, and the positive and negative categories are mainly identified as factors for analysis. Various statistical and rule-based methodologies are employed to examine the tone of the reviews[3]. Part of these categories exist different levels of opinion classification models based on input type, namely[4]:

(1)Document level: In the case of document level, the length of an assessment conducted for opinion classification is one or multiple paragraphs—for example, a review of a particular movie.

(2)Sentence level: In the case of sentence level, the evaluation duration is limited to one sentence.

(3)Aspect level: On the other hand, at the level of aspects, the review text can be one word or several words generally considered to be aspects or features.

Today, consumers are much more aware of the original quality of the product. Therefore, new buyers often go through reviews to make a purchasing decision. All in all, the web has become a popular medium for not only online users to make purchases, but also for all those who want to find relevant information about products and services they have previously committed to purchasing. To solve many decision-making problems, a new dynamic model is needed to uncover latent knowledge from e-commerce review data.

Analyze user opinions using machine learning methods

An imperative task for a company in the field of e-commerce is to maintain its reputation in the online market[5]. Emotion analysis systems are being applied in almost all fields of business and society because opinions are at the heart of nearly every human activity and the primary agent influencing our behavior[6]. Sentiment analysis has provided a wealth of information on customer opinions, leading to product development and better business management[7]. Our beliefs, perceptions of reality, and our choices, largely depend on how others see and evaluate the world. For this reason, when we need to make decisions, we often seek the opinions of others. This is true not only of individuals but also of organizations. Machine learning methods have been widely applied in analyzing online users' views and emotions[8–13]. In a recent study on hotel services, customer opinions were collected and analyzed using the LDA Topic Model, an unsupervised machine learning approach[14]. The model results found hidden topics and keywords with high probability that users are interested in and exchange. Applying the experimental results obtained from the proposed model will provide valuable support for the decision-making process in businesses operating in the hospitality sector, particularly in hotels and restaurants. Another recent study collected and analyzed customer feedback from food-related e-commerce platforms, specifically foody.vn and diadiemanuong.com, to better understand consumer opinions and behaviors[14]. Then, machine learning models are applied and evaluated to select the optimal model. The results show that the proposed model has an accuracy of up to 91.5%, according to experimental research results. Research findings can help business managers and service providers gain insight into customers' satisfaction with their products and services and understand their emotions to adjust and make the right business decisions. It also helps food e-commerce managers ensure better e-commerce service design and delivery. One of the limitations of

the machine learning-based approach is its dependence on the training dataset size, which must be large enough. However, labeled data is often uncommon, especially in some narrow disciplines. Most research groups have to spend time and money on data labeling [15]. In a 2022 study, it was found that tree-based models outperformed other classification approaches, with the Random Forest algorithm achieving an accuracy of 89%. Specifically, the model reached 73% accuracy for binary classification and the same accuracy (73%) for three-class classification.

PROPOSED RESEARCH METHODS AND MODELS

Research methods

In this study, qualitative and experimental research methods are applied. In which the qualitative method is used to survey secondary data, published results, and work on analyzing customer views when using mobile commerce applications in particular. online in general. From there, find out the basis to conduct model building and process to carry out empirical research; The experimental method is applied to survey and collect data, analyze, and process the data from the collected and experimental data sets by machine learning methods. This method is also used to evaluate the experimental and visual results of the customer opinion analysis.

Proposed research model

Fig 1 presents the proposed research model as follows: The first stage starts from accessing the website https://shopee.vn/ and collecting a raw dataset using Python's Scrapy library about product information, store, review score, and customer reviews about products being traded on the e-commerce platform Shopee through the API obtained from the website. After having the input data set, start surveying and preprocessing to check for inappropriate data and renormalize the data set to optimize the machine learning model building step. Divide the dataset into a dataset to train the machine learning model (data train) and a dataset to test the model's accuracy (data test) with the ratio 8:2 and 7:3. Then, use the training data to apply to machine learning evaluation methods by feature methods and then use the test data to re-evaluate these machine learning methods, to compare and select the most optimal model compared with the obtained evaluation data set. Finally, from the appropriate output data set, use data visualization tools such as Power BI and Python to visualize and provide meaningful comparisons and metrics to improve product quality or come up with more innovative and more effective customer access strategies.

Data collection

To test the proposed method, we collected over 200,000 comments from an extensive mobile commerce application in Vietnam, Shopee. The empirical dataset comprises customer feedback collected between 2017 and 2022, providing a factual basis and capturing temporal variations over these five years. *Tables 1* and *2* below show the data fields collected from this application.

Table 1: Data structure of collected data from e-commerce applications [Source: Author's own work]

Data fields	Data types	Describe
category	nvarchar(50)	List of categories on shopee
review_id	nvarchar(50)	Comment code
reviews	nvarchar(max)	Customer comments
id_user	nvarchar(50)	Customer's code
user_name	nvarchar(50)	Customer account name
order_id	nvarchar(50)	Order number
item_id	nvarchar(50)	Product code
shop_id	nvarchar(50)	Store code
rating_status	tinyint	Rating Status
scores	tinyint	Point evaluation
shipping	nvarchar(10)	Shipping area

Data preprocessing

Once you have the data set, preprocessing the data will be the first step. Data preprocessing is basically the process of renormalizing raw data, removing elements that are not meaningful for text classification, comments [16]. For example, we have an input comment like this: "beautifulppp!!!!! @shopee". After going through the data preprocessing step, it will be normalized to "so beautiful!". Non-meaningful components will reduce accuracy when building machine learning models. Preprocessing is an important step to increase accuracy in machine learning. Preprocessing on the Vietnamese data set usually includes steps shown in Fig. 2.

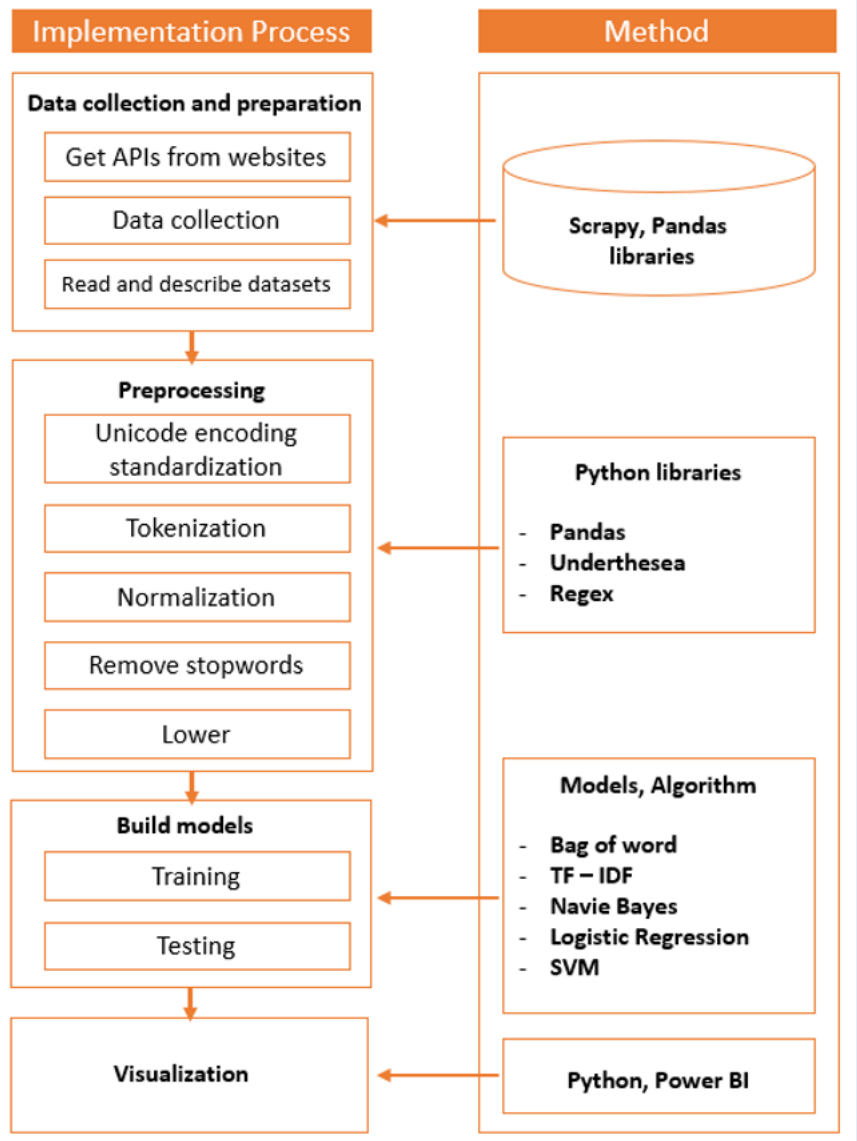

Figure 1: The process of building a customer perspective analysis model [Source: Author's own work]

Data Labeling

After we have normalized all the comment data, one more important step before we put the dataset into training is to label the data. Based on previous studies on data labeling methods, the emotional data labeling method based on the customer's rating score [17] is used to label and divide the comments into two categories as follows:

-**Rating score <= 3**: Negative comments (Negative)

-**Rating score >= 4**: Positive comments (Positive)

After assigning data labels, we get the results in Table 3.

The results show that, in more than 200,000 lines of collected review data, comments with a rating of 4 to 5 stars accounted for 97.14% with 209,698 comments negative comments with a rating score of less than 3 stars only accounted for 2.86% with 6,176 comments in total.

Feature extraction method

During data training, the computer only understands this data in vectorized form. Therefore, the data after being preprocessed and labeled will be represented in the form of a vectorized matrix. There are currently many methods to support text vectorization such as:

Figure 2: Data preprocessing process [Source: Author's own work]

Table 3: Number of comments labeled by product category [Source: Author's own work]

Product portfolio	Positive	Negative	Total
Travel sports (Thể thao du lịch)	37,114	1,180	38,294
Online Department Store (Bách hóa online)	30,645	541	31,186
Phone accessories (Điện thoại – Phụ kiện)	27,718	434	28,152
Men's shoes (Giày dép nam)	21,040	759	21,799
Women's purses (Túi ví nữ)	17,790	476	18,803
Computer – laptop (Máy tính laptop)	17,557	526	17,881
Clock (Đồng hồ)	17,149	732	15,742
Women's jewelry (Trang sức nữ)	15,023	719	13,319
Cars - Motorbikes – Micycles (Ô tô – Xe máy – Xe đạp)	13,032	287	9,118
Women's shoes (Giày dép nữ)	8,707	411	4,025
Total	209,698	6,167	215,865

Table 2: The store data structure is collected.[Source: Author's own work]

Data fields	Datatypes	Describe
avg_rating_scor	float	Average store review score
category_id	nvarchar(50)	Catalog code on shopee
followers	int	Store Followers
historical_sold	int	Total number of products sold
id_items	nvarchar(50)	Product code
shop_id	nvarchar(50)	Store code
shop_name	nvarchar(max)	Name of the store
shop_location	nvarchar(max)	Province/city
date	date	Comment time
year	int	Year of comment
month	nvarchar	Comment month
price	float	Product price

traditional methods and the most popular method is Bag of Words (BoW), or the newer method is Term Frequency - Inverse Document Frequency (TF-IDF). Many studies [18–21] have used this method as a technique for extracting and representing text in vectorized form. This method can evaluate the weight of a word in the input text. The more repeated words in the text, the higher the return value and the more critical it is. TF-IDF can also generate stopwords, such as summaries or comment categories. When calculating, all words in the comment are evaluated equally. But there will be some unimportant words like "I", "He", "the" …etc that will often appear in Vietnamese comments. We need to reduce the frequency of these words to reduce their importance in the comment. Formula (1) represents the IDF in the text data set.

$$IDF\ (t,\ D)\ = \log \frac{|D|}{|\{dD : td\}|} \quad (1)$$

In there:
- IDF (t, D): TDF value of word t in input data set
- |D|: Total number of comments in the dataset
- |{d ⊠ D: t ⊠ d }|: represents the number of comments in set D containing the word t.

Model testing and evaluation

In this process, the data set is divided into two parts, training and testing, separately using the Sklearn library with a specific scale. In this problem, we experimentally divide the data set with two ratios of 8:2 and 7:3 and compare the results. Perform the following steps:

Table 4: Ratio of training and test data sets [Source: Author's own work]

Index	Training set	Test set
1	80%	20%
2	70%	30%

- **Step 1**: Read the data from the repository and split it into two parts: the comment data and the positive or negative label column. Divide into 2 sets of train (x_train, y_train) and test (x_test, y_test) according to the ratio in Table 4.
- **Step 2**: Save two training sets and test data for training with predictive models.
- **Step 3:** Return the label (label) to a vector format to facilitate calculations using the LabelEncoder library.

- **Step 4**: Test the dataset on predictive models, including Naïve Bayes, Logistic regression, and Support Vector Machine (SVM). Then evaluate the results of these machine learning prediction models.

EXPERIMENTAL RESULTS AND DISCUSSION

In our study, we developed models not only for classifying sentiments into three categories (positive, negative) but also for analyzing sentiment across multiple dimensions such as aspect categories (e.g., product, …), temporal trends, and geographic regions. This approach allows us to gain more profound, more granular insights into how sentiment varies across different facets of customer experience.

Results of the comparison of experimental models

In this study, we selected **Naïve Bayes**, **Logistic Regression**, and **Support Vector Machine (SVM)** for sentiment classification based on several considerations. These three models are widely used and well-established baseline algorithms in sentiment analysis research [15]. Their inclusion enables effective benchmarking and comparison with results from existing studies. All three models are particularly effective for handling high-dimensional and sparse text

data, which are common characteristics in natural language processing (NLP) tasks like sentiment analysis. These models are computationally lightweight compared to more complex neural networks, which is advantageous for rapid experimentation and applications with limited resources. Logistic Regression and Naïve Bayes are also more interpretable, which aligns with our interest in explaining model decisions.

After conducting the training on the training data set in the models for training and putting the test dataset into predict the accuracy, the data set was divided into two different ratios: 8:2 and 7:3. The prediction results of the models are described in Table 5 and 6.

In this case, when comparing two tables of results. As a result, with a training/test ratio of 8:2, the model has a higher prediction probability result and a shorter prediction time compared to a training/test ratio of 7:3. Therefore, the researchers will choose the training/test ratio of 8:2 as the main ratio for future analysis.

Compare and select predictive models

The experimental results Error! Reference source not found. and Error! Reference source not found. show that the Naïve Bayes model gives the lowest prediction probability (92.86% and 89.86%), and the model training time and the prediction time are also longer. Compared with the other two models. The accuracy of the two Logistic Regression and SVM models is approximately the same, 94.12% and 90.82% for the Logistic Regression prediction model and 94.2% and 91.0% for the SVM prediction model, respectively. However, when comparing the training time of these two models, Logistic Regression returns much better results than the SVM model. After comparing the results above, we found that logistic regression is the most suitable and optimal model compared to the current data set.

Predicted results

After we finished training the model, we included any number of comments for prediction. The results are shown in **Error! Reference source not found.**, as follows:

Visualize the results

The prediction results of the completed models, the next step is to visualize the obtained results, to create valuable dashboards to support sellers and business departments. Businesses easily monitor and make policy changes or make more accurate decisions to reach customers and improve service quality easily. Fig. 4, Fig. 5, and Fig. 6 are the analyzed results.

The dashboard in Fig. 4 shows the overall stats of the stores from 2017 to the end of January 2022. The chart above shows the number of product review comments, the percentage of positive and negative comments on the total number of comments by year, the total number of positive and negative comments in each category, and the statistics of positive and negative reviews of each product that the store is trading on Shopee.

The dashboard shows the number of positive/negative reviews of each product on each store: allowing basic information of the store to be tracked (for example, number of followers on Shopee, rate of customer feedback) and the number of positive and negative reviews. At the same time, combined with pre-made filters on reports generated with Power BI, research can easily filter out different results as the research want, such as statistics on the number of positive reviews/ negative reviews of products in stores in the Ho Chi Minh City or go deeper into the stores in the districts of the Ho Chi Minh city (Fig 5).

The dashboard (Fig. 6) of evaluation according to positive/negative criteria with fixed timelines allows administrators to track essential indicators of the growth in the number of reviews over time, first, with a sticker showing the total number of comments along with the percentage of positive reviews and the percentage of negative reviews according to specific criteria. For example, in Fig. 6, the research gets the following result when the researcher selects the filter to see the indicators of Ho Chi Minh City. The returned results have 4,997 total comments, of which the number of positive comments accounted for 98.36%, and the number of negative comments accounted for 1.64%.

Businesses/managers can track the total number of comments annually, monthly, or in detail at daily and hourly levels in the statistics table. In addition, this report also helps businesses to track the number of new comments from new guests who come to evaluate and track the number of returning guests over time. From there, research came up with a customer review return index, which shows that these stores find new customers, but they still have loyal customers who return to buy again. The results show that most annual and monthly evaluation return rates are favorable, which is excellent.

Cloud chart (WordCloud) statistics the words expressing feelings and words that are mentioned many times in the evaluation dataset. The results of Table 7 show that the words most mentioned by customers in the comments are:

Table 5: Model prediction results with the ratio of 8:2 oftraining and test data sets [Soure: Author's own work]

Algorithm	Naïve Bayes		Logistic regression		SVM	
	Negative	Positive	Negative	Positive	Negative	Positive
Precision	0.33	0.91	0.74	0.94	0.79	0.94
Recall	0.04	0.95	0.12	0.91	0.14	0.96
F - score	0.23	0.94	0.21	0.96	0.24	0.98
Accuracy	92.86%		94.12%		94.2%	
Training time	3.19s		2.83s		2:13m25s	
Predicted time	43.2ms		23ms		17ms	

Table 6: Model prediction results with the ratio of 7:3 oftraining and test data sets [Source: Author's own work]

Algorithm	Naïve Bayes		Logistic regression		SVM	
	Negative	Positive	Negative	Positive	Negative	Positive
Precision	0.01	0.82	0.74	0.87	0.79	0.91
Recall	0.01	0.92	0.12	0.94	0.14	0.93
F - score	0.01	0.94	0.21	0.97	0.24	0.98
Accuracy	89.86%		90.82%		91.0%	
Training time	4s		10s		3h24m41s	
Predicted time	2.8s		2.2s		7s	

	Reviews	Label	Predict
0	sản phẩm đẹp chất lượng tốt	positive	positive
1	ổn nhưng không thấy thẻ giấy bảo hành gì hết	positive	negative
2	đế ấu lắm mọi người mình rất không hài lòng nh...	negative	negative
3	thất vọng ghê không giống ảnh	negative	negative
4	đeo lắm đã rách đứt chỉ ở mũi giày trong khi m...	negative	positive
5	đẹp mua du size minh nhung van rong khuc got c...	positive	positive
6	quai thì lõng dép thì in không đều màu tiền nà...	negative	negative
7	hàng dùng chán không tạo bọt không sạch	negative	negative
8	không khớp mặt đồng hồ lắm bị dư ít ở hài bên rìa	positive	positive
9	sản phẩm ổn đẹp cho thêm nin dự phòng tuy nhiê...	positive	positive
10	đồng hồ mới mua nhưng đã bị vô nước bị ẩm mặt ...	positive	positive
11	giao nhầm màu thôi còn mọi thứ đều ổn với lại ...	positive	positive
12	đồng hồ đẹp lắm ạ mua đúng dịp săn sale nên gi...	positive	positive
13	đồng hồ rất đẹp mỗi tội đồng hồ cơ nài deo trê...	positive	positive

Figure 3: Some predicted results [Source: Author's own work]

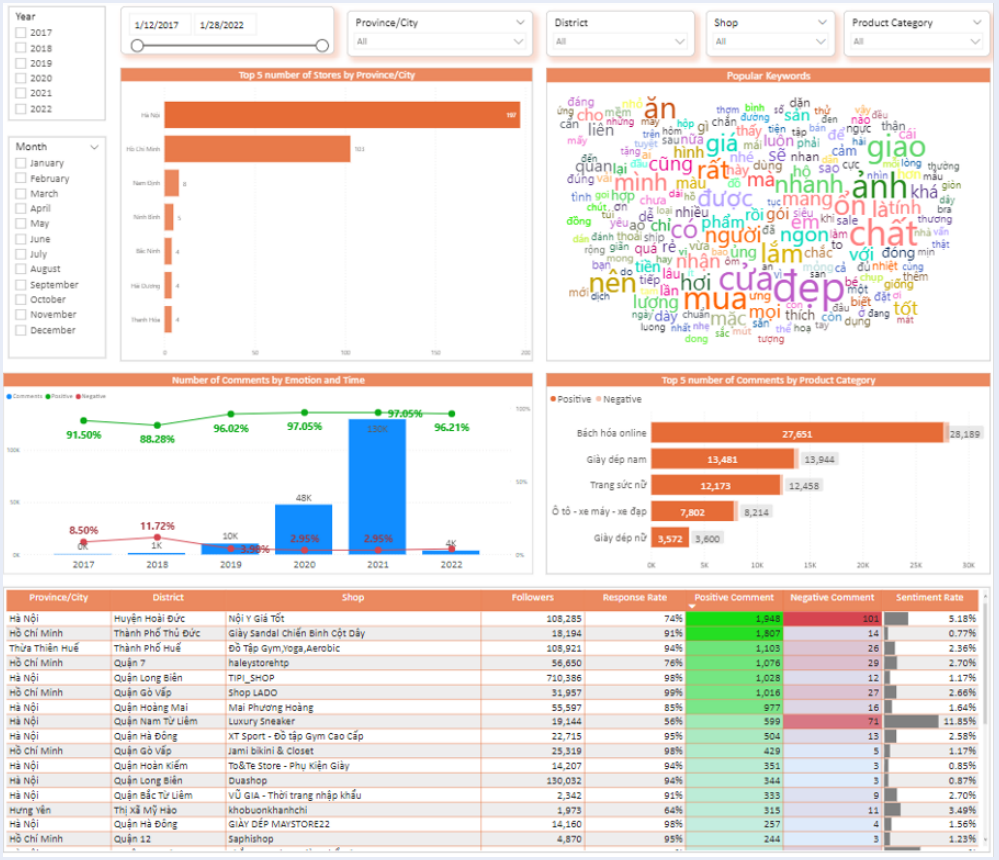

Figure 4: Dashboard of overall stats of the stores [Source: Author's own work]

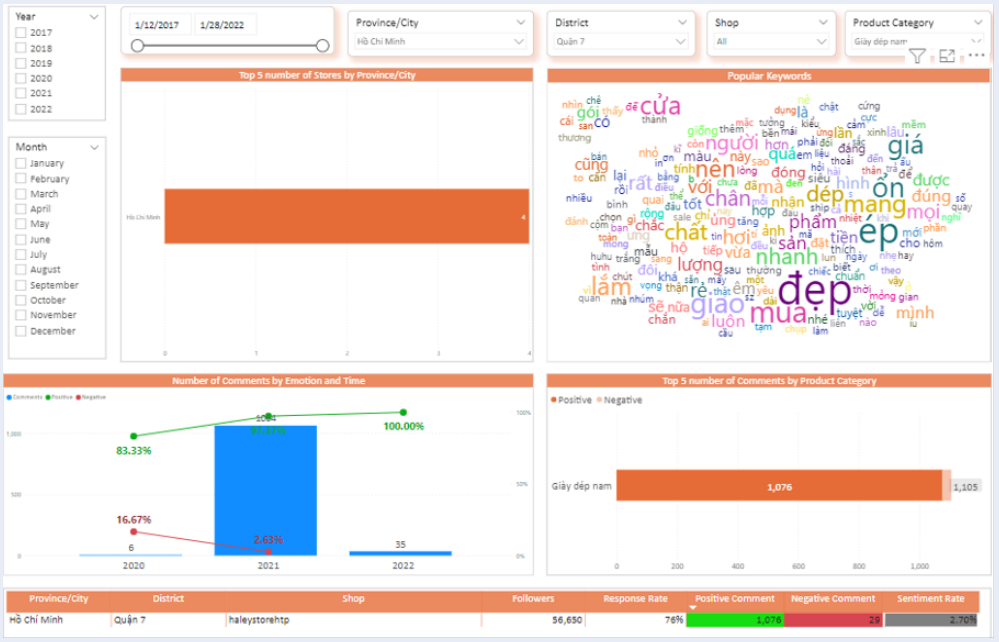

Figure 5: The dashboard shows the number of comments from Ho ChiMinh City.[Source: Author's own work]

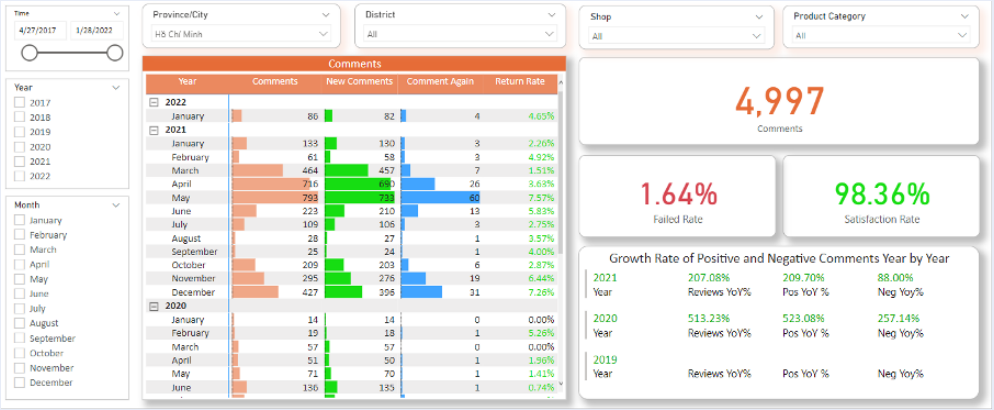

Figure 6: Dashboard shows comment trends over time. [Source: Author's own work]

Table 7: Frequency of occurrence of words.[Source: Author's own work]

Index	Word name	Number of mentions
1	beautiful (đẹp)	790
2	quality (chất lượng)	648
3	buying (mua)	576
4	delivery (giao hàng)	515
5	fine (ổn)	455

When writing a review, these words will help determine what customers want and feel about the issue. For example, with the word "beautiful (đẹp)", we can predict that customers are interested in other products in terms of fashion and images; the word "buy (mua)" shows that the customer is saying that he has bought this product and can also mean that the customer is liking this product and recommending people to buy it, or in a negative perspective, the customer is warning to People should not purchase this product. Or the keywords "quality (chất lượng)", "fast (nhanh)", "slow (chậm)" are also mentioned. This will help retailers keep track of what problems customers are interested in their products, be it about good or bad quality, fast or slow shipping, to make recommendations. Timely solutions to help increase customer experience and improve business profitability.

CONCLUSION AND FUTURE WORK

Conclusion

With this study, the online customer perspective analysis model in e-commerce is proposed and tested on the data set of Vietnamese comments left by users on e-commerce sites. Shopee. Experimental results show that the model has high accuracy. The collected empirical data set includes customer feedback from 2017 to 2022, ensuring factuality and reflecting data fluctuations over time. This helps analysts find trends and make better predictions for the future, thereby having reasonable investment strategies that minimize risk. Visual representations of results by graphs and charts help administrators conveniently and timely capture information, allowing them to see problems from different perspectives (analytical dimensions).

Future works

In the following studies, the model will be developed to identify fake comments (fake comment detection) and then put into training and analysis. Some modern methods in natural language processing, such as topic modeling and lexical analysis methods, will be applied to improve the performance and accuracy of the model. In addition, model development and integration into the real-time forecasting system, continuously collecting and predicting based on customer feedback, is also one of the appropriate development directions of the research. Deep Learning models provide a more objective comparison. Besides, the model can be improved and applied in relevant industries and interdisciplinary fields such as restaurant and hotel, customer behavior and emotion analysis, and customer experience analysis. In this study, the dataset is highly imbalanced, with negative reviews comprising only about 3%. This imbalance is a limitation, as it can bias the model toward majority classes and reduce its ability to detect negative sentiment accurately. While we applied standard preprocessing and evaluation techniques, we recognize that the skewed class distribution may have impacted the model's performance, particularly its recall for negative senti-

ments. In future work, we plan to address this issue by implementing advanced resampling techniques (e.g., SMOTE, undersampling), using class-weighted algorithms, or exploring anomaly detection approaches to improve the detection of minority classes.

ACKNOWLEDGEMENTS

This research is funded by the University of Economics and Law, Vietnam National University, Ho Chi Minh City, Vietnam.

ABBREVIATIONS

AI: Artificial Intelligence
API: Application Programming Interface
BoW: Bag of Words
IDF: Inverse Document Frequency
LDA: Latent Dirichlet Allocation
ML: Machine Learning
NLP: Natural Language Processing
SVM: Support Vector Machine
TF-IDF: Term Frequency – Inverse Document Frequency

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

AUTHOR CONTRIBUTIONS

All authors: Conceptualized and designed the study, wrote the original manuscript, and reviewed and edited the article.

Quoc-Tuan Truong and Van-Ho Nguyen: Conducted the data preparation, developed and implemented the algorithm, conducted data analysis, wrote the Methodology, and wrote the Experimental results.

Hoang-Anh Tran and Van-Ho Nguyen: Conducted the Literature Review and Introduction, Contributed to the visualization of the study, and Reviewed the article.

Thien Le and Hoanh-Su Le: Assisted with the data preparation, contributed to the experime tal design, and reviewed and edited the article.

REFERENCES

1. Prathi JK, Raparthi PK, Gopalachari MV. Real-time aspect-based sentiment analysis on consumer reviews. Singapore: Springer; 2020. Available from: https://www.doi.org/10.1007/978-981-15-1097-7_67.
2. Zvarevashe K, Olugbara OO. A framework for sentiment analysis with opinion mining of hotel reviews. In: 2018 Conference on Information Communications Technology and Society (ICTAS). Durban, South Africa: IEEE; 2018. p. 1–4. Available from: https://www.doi.org/10.1109/ICTAS.2018.8368746.
3. Yu C, Zhu X, Feng B, Cai L, An L. Sentiment analysis of Japanese tourism online reviews. Journal of Data and Information Science. 2019;4(1):89–113.
4. Syamala M, Nalini NJ. A filter based on an improved decision tree sentiment classification model for real-time Amazon product review data. Int J Intell Eng Syst. 2020;13(1):191–202.
5. Thomas KK, Anil SP, Kuriakose NGE. Sentiment analysis in product reviews using natural language processing and machine learning. Int J Inf. 2019;8(2):147–51. Available from: https://www.doi.org/10.30534/ijiscs/2019/35822019.
6. Liu B. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies. 2012;5(1):1–167.
7. Kaur A, Gupta V. A survey on sentiment analysis and opinion mining techniques. Journal of Emerging Technologies in Web Intelligence. 2013;5(4):367–71. Available from: https://www.doi.org/10.4304/jetwi.5.4.367-371.
8. Yang K, Cai Y, Huang D, Li J, Zhou Z, Lei X. An effective hybrid model for opinion mining and sentiment analysis. In: 2017 IEEE International Conference on Big Data and Smart Computing (BigComp); 2017. Available from: https://www.doi.org/10.1109/BIGCOMP.2017.7881759.
9. Li Z, Fan Y, Jiang B, Lei T, Liu W. A survey on sentiment analysis and opinion mining for social multimedia. Multimedia Tools and Applications. 2019;78(6):6939–67. Available from: https://www.doi.org/10.1007/s11042-018-6445-z.
10. Birjali M, Kasri M, Beni-Hssane A. A comprehensive survey on sentiment analysis: Approaches, challenges, and trends. Knowledge-Based Systems. 2021;226. Available from: https://www.doi.org/10.1016/j.knosys.2021.107134.
11. Nguyen VH, Ho T. Analysing online customer experience in the hotel sector using dynamic topic modelling and net promoter score. Journal of Hospitality and Tourism Technology. 2023;14(2):258–77. Available from: https://www.doi.org/10.1108/JHTT-04-2021-0116.
12. Nguyen N, Nguyen TH, Nguyen YN, Doan D, Nguyen M, Nguyen VH. Machine learning-based model for customer emotion detection in hotel booking services. Journal of Hospitality and Tourism Insights. 2024;7(3):1294–312. Available from: https://www.doi.org/10.1108/JHTI-03-2023-0166.
13. Nguyen VH, Nguyen N, Nguyen TH, Nguyen YN, Dinh MT, Doan D. Customer emotion detection and analytics in hotel and tourism services using multi-label classification models based on ensemble learning. Annals of Operations Research. 2025;p. 1–31.
14. Nguyen B, Nguyen VH, Ho T. Sentiment Analysis of Customer Feedbacks in Online Food Ordering Services. Business Systems Research: International journal of the Society for Advancing Innovation and Research in Economy. 2021;12(2):46–59.
15. Shahzad M, Alhoori H. Public Reaction to Scientific Research via Twitter Sentiment Prediction. Journal of Data and Information Science. 2022;7(1):97–124.
16. Vijayarani S, Ilamathi MJ, Nithya M. Preprocessing techniques for text mining-an overview. International Journal of Computer Science & Communication Networks. 2015;5(1):7–16.
17. Liu B. A practical guide to sentiment analysis. Cham: Springer; 2017. p. 11–39. Available from: https://www.doi.org/10.1007/978-3-319-55394-8_2.
18. Ramos J. Using tf-idf to determine word relevance in document queries. In: In Proceedings of the first instructional conference on machine learning. vol. Vol. 242; 2003. p. 29–48.
19. Qaiser S, Ali R. Text mining: use of TF-IDF to examine the relevance of words to documents. International Journal of Computer Applications. 2018;181(1):25–9. Available from: https://www.doi.org/10.5120/ijca2018917395.
20. Sezgen E, Mason KJ, Mayer R. Voice of airline passenger: A text mining approach to understand customer satisfaction. Journal of Air Transport Management. 2019;77:65–74. Available from: https://www.doi.org/10.1016/j.jairtraman.2019.04.001.
21. Le T, Ho T, Nguyen VH, Le HS. How to deeply understand the voice of the customer? A proposal for a synthesis of techniques for analyzing online reviews in the hospitality industry. Journal of Hospitality and Tourism Insights. 2024;7(3):1435–55. Available from: https://www.doi.org/10.1108/JHTI-07-2023-0460.

AUTHOR BIOGRAPHY

Quoc-Tuan Truong Quoc-Tuan Truong was born in Khanh Hoa, Vietnam, in 2000. He received his B.S. degree in Management Information Systems from the Faculty of Information Systems, University of Economics and Law, VNU-HCM, Vietnam, in 2022. He is currently conducting research and working in the field of data analytics with a focus on credit services. He can be contacted at truongquoctuan1703@gmail.com.

Hoang-Anh Tran Hoang-Anh Tran was born in Dak Lak, Vietnam, in 2003. He received his degree in Digital Business and Artificial Intelligence from the University of Economics and Law, VNU-HCM, in 2025. He has participated in several research projects on text mining and machine learning in collaboration with professors. He was also responsible for organizing the national data analytics competition Business-Intelligence. He can be contacted at anhth21416c@st.uel.edu.vn.

Thien Le Thien Le (M.Sc.) received his B.S. degree in Management Information Systems from the Faculty of Information Systems, University of Economics and Law, VNU-HCM, Vietnam, in 2019, and his M.Sc. degree in Information Systems from the University of Information Technology, VNU-HCM, Vietnam, in 2023. His current research interests include business intelligence, data analytics, and digital transformation. He can be contacted at thienlb.ktl@uel.edu.vn.

Hoanh-Su Le Hoanh-Su Le (Ph.D.) received the B.E in Electronics and Telecommunication, MSc in MIS and MBA degrees from Vietnam National University HCM City in 2009 and 2011. He received Ph.D. degree in E-business from Pukyong National University, Republic of Korea in 2016. Since 2011 he has been a faculty member and currently Dean (Tenure) of Faculty of Information Systems at University of Economics and Law, Vietnam National University Ho Chi Minh City (VNU HCM). His research interests are data analytics, artificial intelligence and fintech. The author can be contacted at: sulh@uel.edu.vn.

Van-Ho Nguyen Van-Ho Nguyen (M.A.) received his B.S. degree in Management Information Systems (MIS) from the Faculty of Information Systems, University of Economics and Law, VNU-HCM, Vietnam, in 2015, and his M.A. degree in MIS from the University of Economics Ho Chi Minh City, Vietnam, in 2020. He is currently a Lecturer at the Faculty of Information Systems, University of Economics and Law, VNU-HCM. His research has been published in top journals including Journal of Hospitality and Tourism Insights, Interdisciplinary Journal of Information, Knowledge, and Management, Journal of Hospitality and Tourism Technology, IEEE Access, Business Systems Research, and Journal of Information Processing Systems. He also serves as a reviewer for several Scopus/ISI indexed journals such as Journal of Marketing Analytics, Business Systems Research, and Journal of Information Processing Systems. His research interests include business analytics, business intelligence, data analytics, and machine learning. He can be contacted at honv@uel.edu.vn or nguyenvanho.uel@gmail.com.

VNUHCM Journal of Economics, Business and Law 2026, 10(2):6806-6818

Open Access Full Text Article

Bài nghiên cứu

Phân tích ý kiến đa chiều từ bình luận trực tuyến của khách hàng trong lĩnh lực Thương mại điện tử

Trương Quốc Tuấn[1,2], Trần Hoàng Anh[1,2], Lê Bá Thiền[1,2], Lê Hoành Sử[1,2], Nguyễn Văn Hồ[1,2,*]

TÓM TẮT

Nghiên cứu này trình bày một mô hình thu thập và phân tích đánh giá của khách hàng sử dụng học máy. Trong bối cảnh các nền tảng thương mại điện tử phát triển mạnh mẽ, khối lượng lớn ý kiến khách hàng được tạo ra, nhưng phần lớn dữ liệu này có cấu trúc phi tập trung và khó sử dụng trực tiếp. Để giải quyết vấn đề này, phương pháp đề xuất bắt đầu bằng việc thu thập đánh giá từ các nền tảng trực tuyến, sau đó tiền xử lý dữ liệu để làm sạch, lọc và chuẩn hóa văn bản. Các bước này đảm bảo dữ liệu phù hợp cho phân tích bằng học máy. Các mô hình phân loại sau đó được áp dụng để thực hiện phân tích ý kiến. Kết quả cho thấy mô hình có thể phân loại ý kiến khách hàng với độ chính xác trên 94%. Các phát hiện được minh họa thông qua các biểu đồ và đồ thị toàn diện, cung cấp những hiểu biết rõ ràng về thái độ và mẫu hình hành vi của khách hàng. Những biểu diễn này đóng vai trò quan trọng trong việc chuyển đổi kết quả phân tích phức tạp thành dạng dễ tiếp cận, cho phép các nhà quản lý nhanh chóng xác định xu hướng chính và các vấn đề tiềm ẩn. Bằng cách trình bày dữ liệu từ nhiều góc nhìn, đầu ra trực quan tạo điều kiện hiểu sâu hơn, hỗ trợ phân tích so sánh và nâng cao chất lượng ra quyết định, từ đó củng cố các thực hành quản lý dựa trên bằng chứng trong môi trường kinh doanh năng động. Bằng cách phân tích phản hồi khách hàng có hệ thống, các công ty có thể xác định nhu cầu mới nổi, cải thiện sản phẩm và dịch vụ, đồng thời thiết kế các chiến lược kinh doanh hiệu quả hơn. Những hiểu biết như vậy góp phần trực tiếp vào việc tăng cường năng lực cạnh tranh trong lĩnh vực thương mại điện tử, nơi trải nghiệm người dùng là yếu tố khác biệt quan trọng. Ngoài thương mại điện tử, phương pháp luận đề xuất thể hiện tính linh hoạt và khả năng ứng dụng trên các lĩnh vực khác, bao gồm tài chính, y tế và giáo dục, nơi việc nắm bắt và diễn giải tâm lý khách hàng cũng quan trọng không kém. Bằng cách tích hợp kỹ thuật học máy với phân tích đa chiều, nghiên cứu này thiết lập một khuôn khổ mạnh mẽ và thiết thực để chuyển đổi phản hồi không cấu trúc thành tri thức có thể hành động. Khuôn khổ này không chỉ hỗ trợ ra quyết định chiến lược mà còn cung cấp cho các tổ chức một cách tiếp cận bền vững để tận dụng dữ liệu cho đổi mới và tạo giá trị dài hạn.

Từ khoá: học máy, mô hình phân lớp, phân tích ý kiến khách hàng, thương mại điện tử, phân tích ý kiến đa chiều

[1]*Trường Đại học Kinh tế - Luật, Việt Nam*

[2]*Đại học Quốc gia TP. Hồ Chí Minh, Việt Nam*

Liên hệ

Nguyễn Văn Hồ, Trường Đại học Kinh tế - Luật, Việt Nam

Đại học Quốc gia TP. Hồ Chí Minh, Việt Nam

Email: honv@uel.edu.vn

Lịch sử

- Ngày nhận: 04-04-2025
- Ngày sửa đổi: 21-08-2025
- Ngày chấp nhận: 14-03-2026
- Ngày đăng: 28-06-2026

DOI : **https://doi.org/10.32508/vnuhcmj-ebl.v10i2.1633**

Trích dẫn bài báo này: Quốc Tuấn T, Hoàng Anh T, Bá Thiền L, Hoành Sử L, Văn Hồ N. **Phân tích ý kiến đa chiều từ bình luận trực tuyến của khách hàng trong lĩnh lực Thương mại điện tử**. *VNUHCM J. Econ. Bus. Law.* 2026; 10(2):6806-6818.