sciendo

Sentiment Analysis of Customer Feedback in Online Food Ordering Services

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

Abstract

Background: E-commerce websites have been established expressly as useful online communication platforms, which is rather significant. Through them, users can easily perform online transactions such as shopping or ordering food and sharing their experiences or feedback. **Objectives:** Customers' views and sentiments are also analyzed by businesses to assess consumer behavior or a point of view on certain products or services. **Methods/Approach:** This research proposes a method to extract customers' opinions and analyse sentiment based on a collected dataset, including 236,867 online Vietnamese reviews published from 2011 to 2020 on foody.vn and diadiemanuong.com. Then, machine learning models were applied and assessed to choose the optimal model. **Results:** The proposed approach has an accuracy of up to 91.5 percent, according to experimental study findings. **Conclusions:** The research results can help enterprise managers and service providers get insight into customers' satisfaction with their products or services and understand their feelings so that they can make adjustments and correct business decisions. It also helps food e-commerce managers ensure a better e-commerce service design and delivery.

Keywords: online feedback; food ordering services; Vietnamese sentiment analysis; text analytics

Introduction

Today, advanced information technology has changed the way of communication; it helps users easily access information and exchange their opinions about products and services on a large scale in real-time. The advent of social media and review websites allows users to express their opinions (Akila et al., 2020). The explosion of big data has made online community comments or reviews need to be collected and mined automatically, allowing enterprises to track customers' shopping behavior, interests, and satisfaction with products and services (Yadav, 2015; Akter et al., 2016). Hidden in those comments are the happy, sad, love, and hate feelings. Such "emotional" things it is a big challenge for computers without human reading and self-understanding. From an e-commerce standpoint, detecting the correct customer emotions will help us display better advertising content. For example, spotting a person in a tired mood can suggest some energy drinks, an entertainment venue, or simply play a piece of gentle music. The research direction is not a new one. However, each method has its advantages and disadvantages, and no method is accurate. Because of the intricacy of the Vietnamese language structure, using a lexicon-based technique for opinion mining poses a significant barrier for academics. To deal with the Vietnamese language, there aren't many sets of emotional vocabulary or handling methods. Small businesses are beginning to see the value of social media in achieving their objectives (Balan et al., 2017). Recently, Nguyen et al. (2020) proposed exploring user experience in the hotel sector by using the Topic Model, which is also an effective method in analyzing and extracting information from the corpus of customers' opinions. Therefore, the application of machine learning methods and evaluation of the Accuracy is necessary to choose the most suitable method through collected datasets.

The goal of this study is to analyze opinion mining studies and suggest the use of a machine learning approach to exploit consumer comments in Vietnamese. This research applies the knowledge mining method from data collected by automatic programs, including 236,867 reviews from customers on online ordering services and eating places review channels, namely foody.vn and diadiemanuong.com, which are famous for e-commerce websites in Vietnam. Then, data preprocessing was conducted, and machine learning methods were applied to find the best model and predict sentiment scores for the rest of the corpus.

The structure of this paper is divided into five sections. Section 1 describes the necessity of the research. Theoretical bases related to the research are presented in Section 2. In Section 3, the author describes the research method and experimental designs. The research results are detailed in Section 4. Finally, conclusions and future research are presented in Section 5.

Related works

This section focuses on exploring related research in customer opinion mining sentiment analysis, especially in the online service field. The machine learning and lexicon-oriented approaches in some research are also explored and analyzed to form the basis of this research.

Customer Opinion Mining in online services

The development of technology and using social media on a large scale has created opportunities to get useful insights from data without proper schema. Opinion mining in big data is used to categorize customers' opinions with different emotions and gauge customer mood. Opinion mining has gained significant results over time based on many comments available online. Customers have shared their opinions on products and

services in restaurants, schools, hospitals, vacation destinations, etc. The value of a user's comment, review, or rating about some product or service is their thoughts, judgment, psychological or feelings about its quality, appearance, or price. Depending on individual perceptions, opinions can be positive, negative, or neutral. Users may now express their opinions and make them visible to anybody on the internet thanks to social media. Based on that, enterprises can improve their products, services, and marketing strategies, to detect the latest trends opportunities or measure the effectiveness of their marketing activities (Pejić Bach et al., 2019). Currently, the community of scientists has much research on opinion mining methods and the applications of opinion mining at many different levels. In the study of Akila et al. (2020) and Nagpal et al. (2020), the authors have proposed tools and methods to collect and analyze customer comments using machine learning and topic models. In another study by Patel et al. (2020), the author analyzed users' emotions based on the customer rating score of the products and services they used in the food services. From the results of domestic and foreign researches, the author found that there are two popular approaches in opinion mining: (1) Based on machine learning (Kadriu et al., 2019; Khairnar et al., 2013; Le et al., 2017) and (2) based on lexicon (Liu, 2012, 2017; Vu et al., 2017; Li et al., 2019). In addition, to increase the efficiency of the opinion mining method, the research has used a hybrid method combining machine learning and lexicon (Mudambi et al., 2010; Maks et al., 2012; Sun et al., 2017; Yang et al., 2017).

One of the limitations of the machine learning-based method is its dependence on the training dataset size, which is labeled and must be large enough. However, labeled data is often uncommon, especially in some narrowly specialized majors. Most research teams must spend time and cost on labeling the data.

Machine learning-based customer sentiment analysis

Emotions and sentiment are a problem that many scientists are interested in and researched (Akter et al., 2016; Lugović et al., 2016). So, there are different views about the number of emotions. Based on the nature of emotions, emotions can be divided into 2 categories: positive emotions and negative emotions. Based on expression and content, we can divide emotions into six basic categories: happy, sad, angry, surprised, hate, scared. Under the impact of different stimuli in different conditions and circumstances, human emotions sometimes intertwine, mix with others, and coexist simultaneously. And this created a series of other emotions.

For the most part, sentiment analysis was characterized as "the study computation of views, feelings, and emotions represented in the text" (Nagpal et al., 2020). In other words, opinion mining, as a way of obtaining the viewpoint of the person who generated a certain document, has lately been the most popular study topic in general social networks (Pang et al., 2008; Ohana et al., 2009). The importance of sentiment analysis has grown with the rise of social network media such as reviews, discussions forum, and social media. Especially in the era of digital development with the explosion of the internet, a lot of this research has focused on social networking domains (Facebook, Twitter...), as in Dunđer et al., (2016), Krstić et al., (2019). Due to some characteristics of the language on social networks, such as a limited number of characters or emotions depending heavily on what users are reading and listening to, the emotional classification of users in social networks is a challenging issue. Machine learning has been applied and has achieved some success in sentiment analysis (Khairnar et al., 2013; Kadriu et al., 2019).

Lexicon-based customer sentiment analysis

Opinions and comments of customers are natural written form (Liu, 2012). In some research by Maks et al. (2012), Akter et al. (2016) gave some methods and techniques of natural language processing in analyzing opinions and sentiment of customers through online commentary. Previous research mainly focuses on vocabulary – lexicon-based and machine learning-based methods. For the lexicon-based approach, the outcome depends heavily on the quality of the emotional words. In a subtle way, the outcomes of machine learning-based approaches, such as SVM and Nave Bayes, are significantly reliant on feature selection methods, such as n-gram or lexicon-based. The research of Vu et al. (2011) has given ways or reviews that explore words in Vietnamese comments in general, but it is almost absent in favor of the user emotions.

The lexicon-based method of analysis depends on the emotional vocabulary sources. An emotional vocabulary source, which is often understood as a dictionary, is a collection of words expressing emotions, with each word assessed as polarizing by a real number. These dictionaries can be built by hand or semi-hand. The advantage of this approach is that there is no training required since there is no need for labeled data. This method is commonly used for sentiment analysis on common text types: blog posts, comments on film, product, or forums.

The research of Ohana et al. (2009) used the SentiWordNet dictionary to evaluate the polarization of film comments. SentiWordNet is an automatically generated dictionary based on a WordNet database, and the best results get an accuracy of 69.35%. The authors conclude that using a SentiWordNet dictionary is as effective as using a hand-built dictionary. Other research has built their dictionaries based on different sources. Research by Taboada et al. (2011) and Liu (2012) affirms that dictionary building helps to establish a solid foundation for this approach.

Support Vector Machine – A classificational algorithm

SVM is a machine learning taxonomy using the kernel function to map a space of data points that cannot be linearly separated into a new space with error classification. For instruction on SVM and their recipe details, we refer readers to Burges (1998). A detailed treatment of the application of these models for text classification is possible found in Joachims (2002).

SVM is essentially an optimal problem; the goal of this algorithm is to find a space F and the super-plane decision f over F such that the classification error is lowest. Let the sample set $\{(x_1, y_1), (x_2, y_2), ... (x_f, y_f)\}$ with $x_i \in R^n$ belong to two classes of labels: $y_i \in \{-1,1\}$ is the corresponding class label of x_i (-1 represents class I, 1 represents class II). We have, the super-plane equation contains the vector x_i in space: $x_i.w + b = 0$

$$\text{Set } f(x_i) = \text{sign}(x_i.w + b) = \begin{cases} +1, x_i.w + b > 0 \\ -1, x_i.w + b < 0 \end{cases} \tag{1}$$

Thus, in the equation (1), $f(x_i)$ represents x_i's classification into the two stated classes. We say $y_i =+1$ if $x_i \in$ class I and $y_i = -1$ if $x_i \in$ class II. Then, to have a super-plane f we will have to solve the following problem: Find min $\|w\|$ with W satisfying the following conditions: $y_i(\sin(x_i.W + b)) \geq 1$ where $i \in [1, n]$.
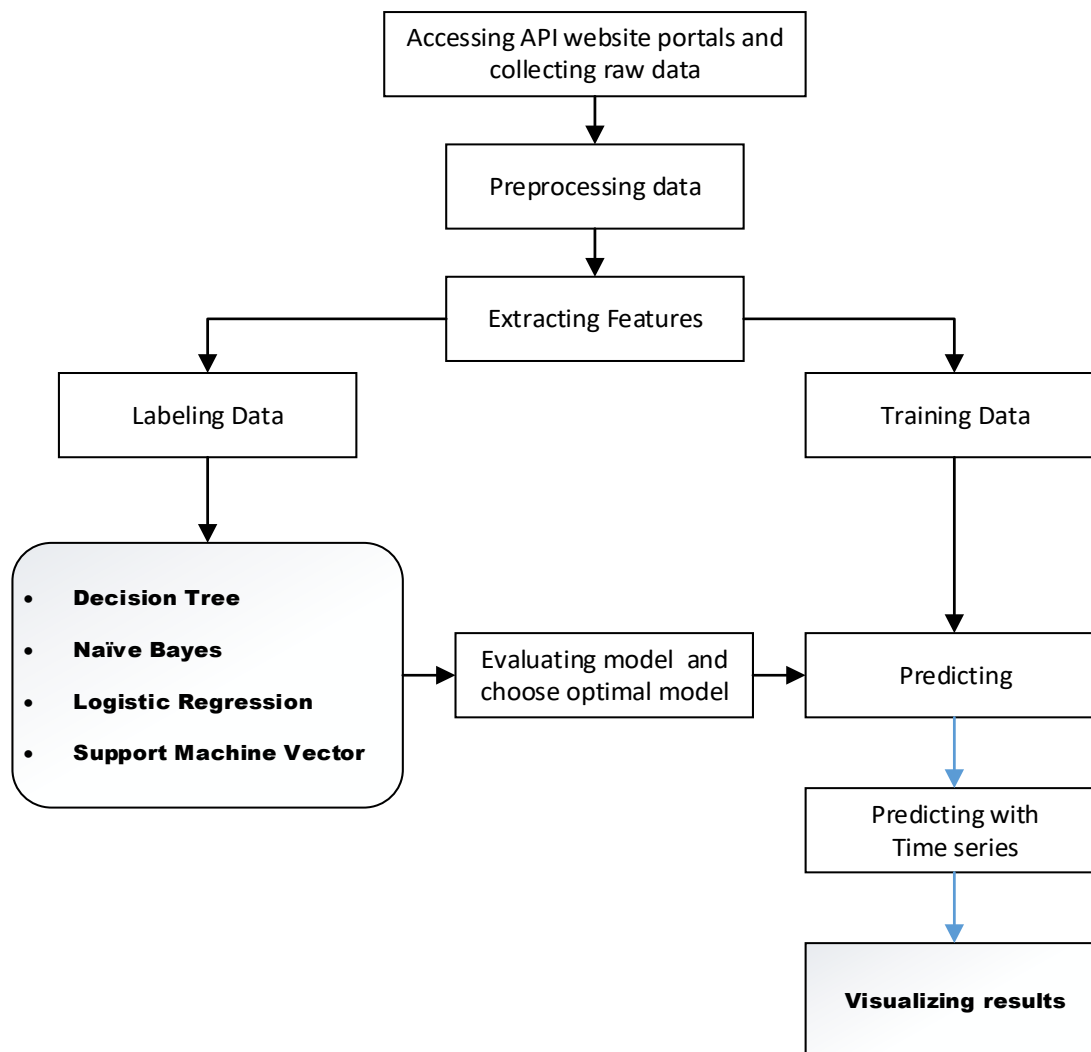
Methodology

This section describes the General Model that the research proposes. Followed by steps to preprocess the data, train, evaluate the model, and conduct data analysis with the time factor.

Overview model and methods

The research data was collected for research purposes, containing raw data from the Foody.vn and diadiemanuong.com websites. Before the machine learning procedure, the raw data is preprocessed, sampled, and labeled. Training, validation, and test data are the three types of sampling data. The training dataset is used during the learning process and is used to fit the parameters; the validation dataset is a dataset of examples used to tune the hyperparameters of a classifier. Test datasets are used only once as the final step to reporting estimated error rates for future predictions. Figure 1 is an overview of the research model which we have done.

Figure 1
Proposed Overview Model and Methods


```
┌─────────────────────────────┐
│  Accessing API website       │
│  portals and collecting      │
│  raw data                    │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│     Preprocessing data       │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│     Extracting Features      │
└─────────────────────────────┘
        │             │
        ▼             ▼
┌──────────────┐  ┌──────────────┐
│ Labeling Data│  │ Training Data│
└──────────────┘  └──────────────┘
        │                 │
        ▼                 │
┌──────────────────┐      │
│ • Decision Tree  │      │
│ • Naïve Bayes    │      │
│ • Logistic       │      │
│   Regression     │      │
│ • Support        │      │
│   Machine Vector │      │
└──────────────────┘      │
        │                 │
        ▼                 ▼
┌──────────────┐  ┌──────────────┐
│ Evaluating   │→ │  Predicting  │
│ model and    │  └──────────────┘
│ choose       │         │
│ optimal model│         ▼
└──────────────┘  ┌──────────────┐
                  │ Predicting   │
                  │ with Time    │
                  │ series       │
                  └──────────────┘
                         │
                         ▼
                  ┌──────────────┐
                  │ Visualizing  │
                  │ results      │
                  └──────────────┘
```

Source: Author's proposal

Data crawling

The Beautiful Soup and Selenium libraries in Python language collect data on the websites. The data collection is based on the Hypertext Markup Language (HTML) structures of foody.vn and diadiemanuong.com. If we want to collect some information data, we proceed to retrieve the data corresponding to the HTML tag containing that information. The result of this step will collect all website data in HTML or TXT formats. This data will be processed in the following steps.

Result of data collecting
The collected dataset had 236,867 records, shown in Table 1, including store name, address, commented customer name, commented time, comment content, rating of a customer for that store. The number of reviews gathered from foody.vn is 214,835 comments; for the diadiemanuong.com is 22,032 comments. This dataset will go into the preprocessing and cleaning step to provide input to the later steps of the models.

Table 1
Results of data crawling

Sources	Number of reviews
foody.vn	214,835
diadiemanuong.com	22,032
Total	236,867

Source: Authors' work

Data preprocessing
Collected data is raw unprocessed so that the data may be empty, misspelled, too short, too long, or contain icons. This will affect the analysis results, so we need to clean up the data. The steps are as below:
- **Remove icon and special characters:** special characters do not have any definite meaning, on the other hand, cause interference in the analysis. Convert all to lower case: each character represents a binary sequence in computer memory. Because the upper-case characters will have a Unicode code that is different from the lower case, which has the same semantically, the computer will not be able to distinguish the input data so that the prediction may be affected. Therefore, converting the entire text to lowercase is reasonable for the analysis and prediction system.
- **Transform words to normal form**: conversion to clear words is required for the preprocessing of the data. Comments on Foody (commented by users in Vietnamese) may have acronyms or misspellings. For example, words in Vietnamese: "ko ngon" (not delicious), "vs" (with), "15k" (15,000 VND) ... or data is not normalized, not standardized. This will interfere with the results of the analysis. During machine learning training, the input is "không ngon", but when predicting the output, the phrase "ko ngon" does not appear during the training, so it will be difficult to identify emotional and predictable results.
- **Remove blank/NULL data**: the collected dataset will have a lot of blank data, which does not make sense in the analysis process, causing a waste of storage memory.

Data labeling
Normally, the data labeling in research applying machine learning will be built by hand. However, after randomly reviewing the content of the collected comment dataset and based on the results of the rating (the rating field in the dataset), founding that comments with a rating less than 5.0 have a negative meaning, and vice versa, comments with a rating equals or greater than 5.0 have a positive meaning. To perform the data labeling process before being trained, the research applied the classifying emotions method according to the customer rating (Liu, 2017; Patel et al., 2020) to divide the collected dataset into 2 datasets, labeled according to the following rules:
- Rate < 5: reviews below 5 stars will be labeled negative.
- Rate >= 5: Review comments rated above 5 stars will be labeled as positive.

The labeling results showed that most of the data were positive comments which accounted for 81.9% of the total comments; the negative comments accounted for 18.1% of the total comments, as table 2 below:

Table 2
Labeled data

Type	Number of reviews	% of total
Negative	42,799	18.1%
Positive	194,068	81.9%
Total	236,867	100.0%

Source: Authors' work

Training and Evaluating model

Normally, the efficiency of opinion classification models is evaluated based on four indicators: Accuracy, Precision, Recall, and F1_Score (known as a harmonic average of Precision and Recall in Table 3). They are formulas (2), (3), (4), and (5), respectively. In addition, this research also considers the training time and the predicting time of each model.

Table 3
Confusion matrix

	Predict: Positive	Predict: Negative
Actual: Positive	True Positive (TP)	False Negative (FN)
Actual: Negative	False Positive (FP)	True Negative (TN)

Source: Authors' work

There is,

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1\_Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{5}$$

Results and Discussion

The results of data preprocessing, training, and model evaluation are presented in this section. Along with that, the results are visualized, and discussions related to the research topic are presented.

Result of training and Evaluating model

This is the most important stage of opinion mining research to determine whether a customer comment is "positive" or "negative". This research applies some classification methods of the Supervised Machine Learning group that are considered the best. Based on the results of the previous research related to the topic, find the most suitable model for the dataset, which is the classified comments. Then, forecasting the unsorted comment data or new comment data arises without retraining.

Table 4 shows the experimental results of the methods. The Accuracy of Decision Tree is 89%, Naïve Bayes 82.5%, Logistic Regression 90%, and Support Machine Vector 91%. In addition, it also shows the training and prediction time of each method. The Decision Tree method has a training time of 1h 4m 32s and a prediction time of 14,300 ms, while the Support Machine Vector (SVM) has a training time of 6,320 ms and a prediction time of 31.25 ms.

Table 4
Results of training and evaluating model

Models	Decision Tree		Naïve Bayes		Logistic Regression		SVM	
	Positive	**Negative**	**Positive**	**Negative**	**Positive**	**Negative**	**Positive**	**Negative**
Precision	0.88	0.97	0.82	0.99	0.90	0.88	0.92	0.86
Recall	1.00	0.37	1.00	0.04	0.99	0.50	0.98	0.64
F1_score	0.93	0.54	0.90	0.08	0.94	0.64	0.95	0.73
Accuracy	89.00%		82.50%		90.00%		**91.50%**	
Training time	1h 4min 32s		1.260 ms		53.700 ms		**6.320 ms**	
Predicting time	14.3 s		66.1 ms		31.5 ms		**31.25 ms**	

Source: Authors' work

A clustered bar chart shows the experimental results of the model in Figure 2 below. In this chart, we can see the column that shows the SVM algorithm's Accuracy is highest with 91,5%.

Figure 2
Results of training and evaluating model (Precision, Recall, F1_Score, and Accuracy)
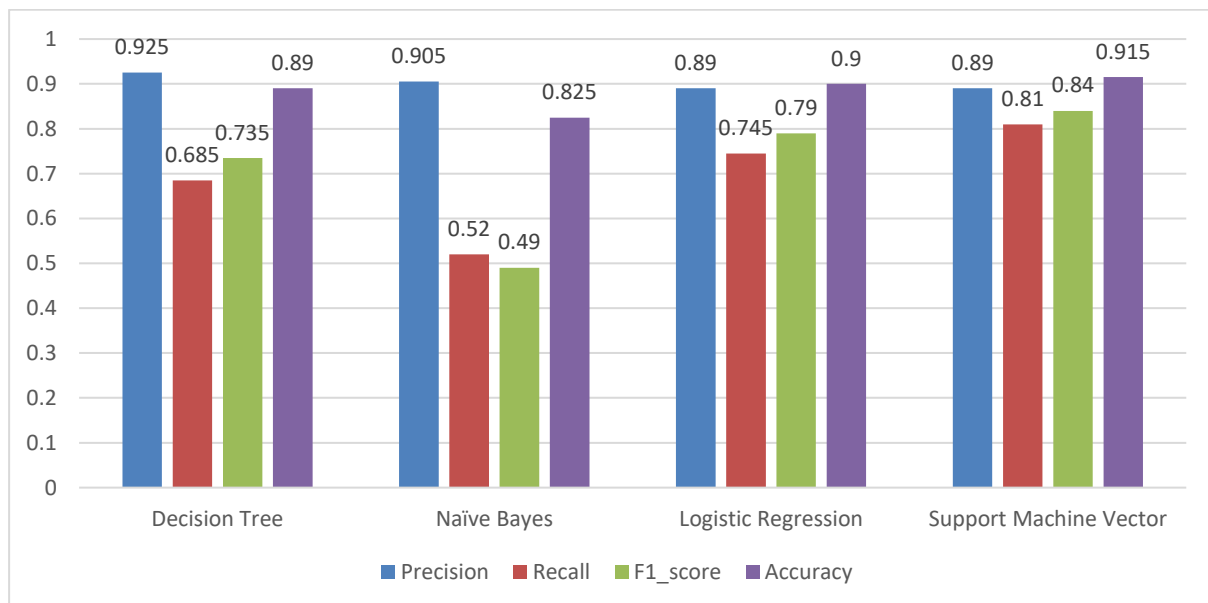
Source: Authors' work

Result of visualization

The visualization results in Figure 3 include the following four charts: Rating by Store, Top stores with a high review, Criteria Scores by Year, and Sentiment by District. Reports are filtered, and information is displayed only in 2020.

The **Rating by Store** chart shows the average customer rating information for each store. In addition, it also shows the average rating of all stores, which is 5.904, through which we can correlate the rating of the store with the average value. For example, "3 Râu" - the fried chicken store has an average rating of 10.00, and R&B milk tea has 9.7.

The **Top Stores** with high reviews chart show the total number of customer comments for each store. The chart shows "Mực nướng Đảo Ngọc", "Baozi - Ẩm thực Đài" are stores that are more interested in and commented on by customers than the rest of the shops.

The **Criteria Scores by Year** chart shows the total customer rating according to the criteria (location, price, quality, service, space). In 2020, the total rating by location is 2366, by price is 2319, by services is 2520, by quality is 2411, and by space is 2452.

The **Sentiment by District** chart shows information about total negative and positive comments distributed by districts in Ho Chi Minh. For example, District 1 has a positive comment rate of 63%, and negative comment rate is 37%, or Binh Thanh district has a positive comment rate of 64%, negative comment rate is 36%.

Figure 3
Dashboard Sentiment Analytics
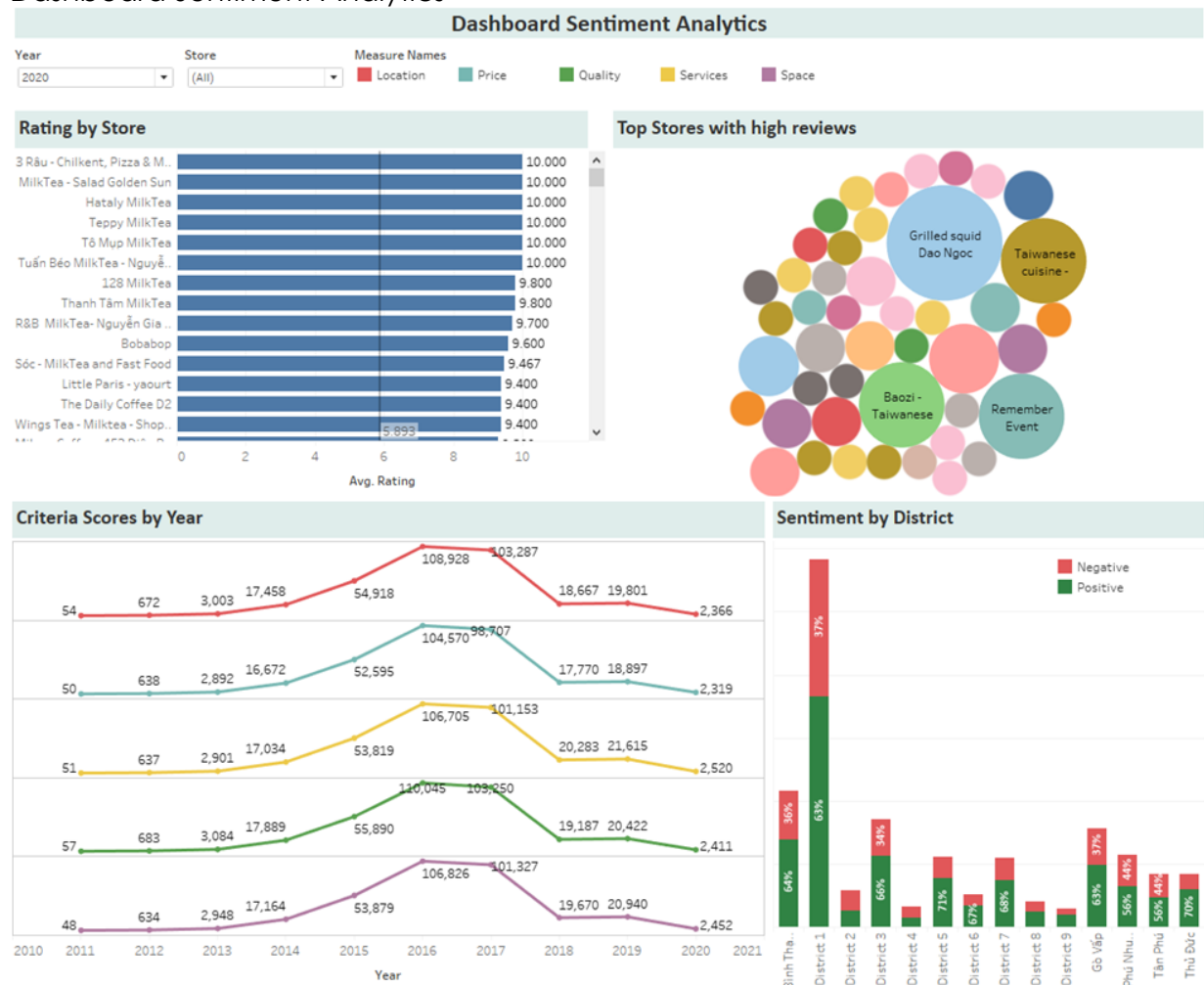
Source: Authors' work

The Word Cloud chart represents negative and positive keywords, making it easy for viewers to catch up with and compare them. In Figure 4, it's easy to see which words are mentioned the most in customers' comments, and the bigger words, the more mentioned. In the WordCloud_Positive chart, the word "món ngon" (delicious plates) appears most in the customers' reviews. Similarly, in the WordCloud_Negative chart, the word "thất vọng" (disappointed) was mentioned most.

Figure 4
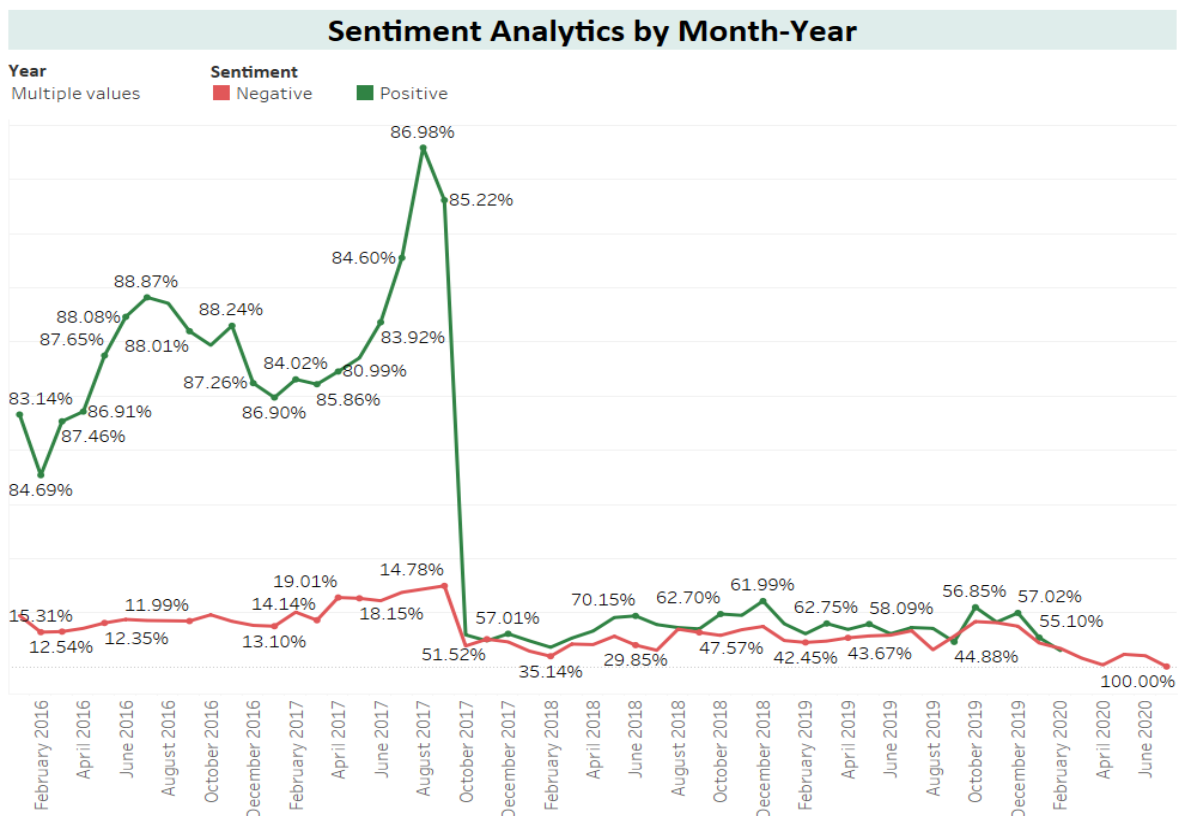
Vietnamese Word Cloud by Positive and Negative
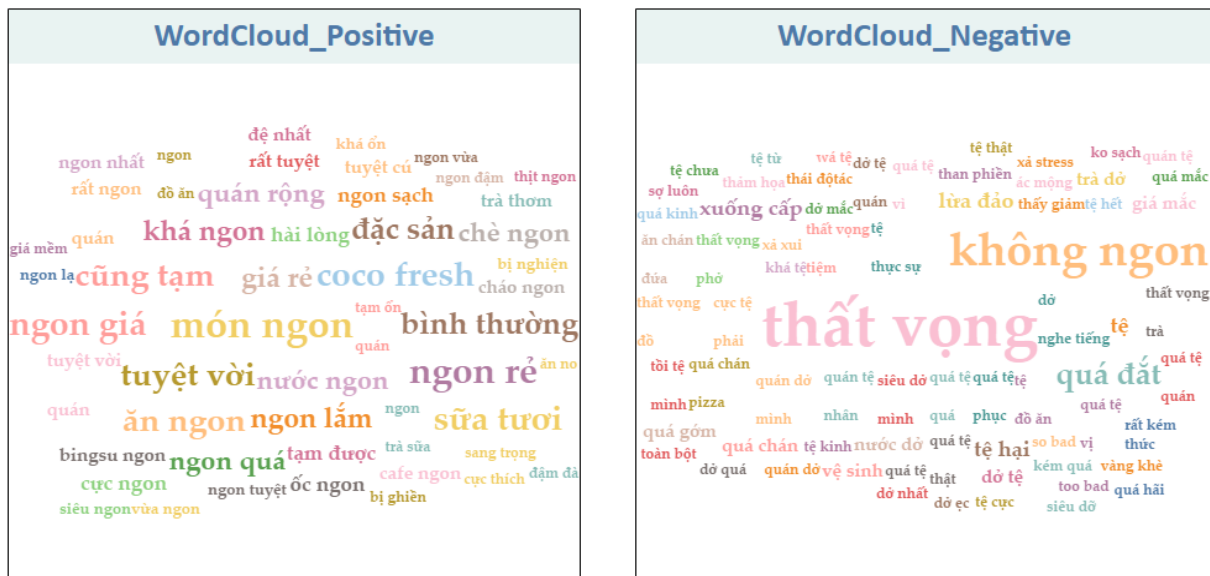
Source: Authors' work

Result of training and evaluating model over time

Figure 5
Sentiment Analysis by Month-year

Source: Authors' work

The research has conducted experiments on the dataset for the SVM method combining the time factor. The results are shown in Figure 5; the Sentiment by Month-

Year chart shows the percentage of positive and negative comments over time. For example, in February 2016, the rate of positive comments was 83.14%, and the negative comments rate was 16.31%; in September 2016, the rate of positive comments was 88.01%, and the negative comments rate was 11.99%. This dashboard lets managers capture customers' emotions very promptly and quickly. This makes a lot of sense in business and management.

Figure 6 below is the accuracy result from 2015-2020 of the SVM method. The chart is the experimental results of the SVM method for the dataset grouped by year. Including 6 datasets (2015, 2016, 2017, 2018, 2019, and 2020). The SVM accuracy for the 2015 dataset was 89%, 2016 was 92%, and 2020 it was 92%.

Figure 6
SVM's Accuracy by year (2015-2020)
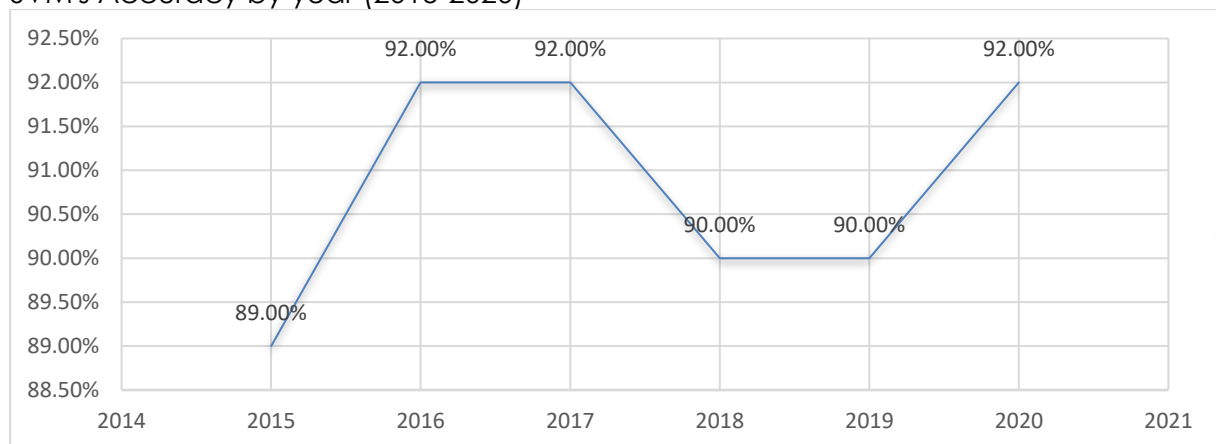
Source: Authors' work

Conclusion

In this paper, the research experimented, compared, and selected suitable machine learning methods to analyze and classify sentiment based on customers' opinions. The applications of the opinion categorization depend on the field, the analysis model, and the source of the collected data. In this research, we have proposed an application solution in natural language analysis, namely, customer sentiment analysis based on comments posted on foody.vn and diadiemanuong.com websites. The solution is tested on many different machine learning methods to compare the pros and cons of the model and select the best model through F1-Score measurement. The research results implemented on the corpus from 2011 to 2020 show that the SVM algorithm has the highest Accuracy with 91,5%. Especially creating visual reports, the analysis combined with the time factor to serve the decision-making needs of businesses. Solving the data explosion problem is to provide customer experience information in locations. The research provides a fundamental architecture in exploiting customer opinions from text data in Vietnamese on social networks, creating the basis for further research in exploiting Big Data in each industry field, creating value for business and consumers. In addition, the research results also significantly contribute to the practical application of social network data mining in the process of understanding users' needs, thereby making appropriate business decisions and management of an enterprise. At the same time, the results also open the application direction for regulators in gathering people's comments on drafts and management policies before being promulgated through social networks. The food and beverage sector will have strategies to develop better services and products to attract better and retain customers. In addition, the

research will be the premise for data analysis applications, using this solution to integrate into applications with the purpose of surveying customer experience feelings for all products and services, especially applying in Vietnamese language processing.

We will expand by installing the system to automatically update data in further research. Data will be automatically extracted from the website and remove duplicate entries before saving to the database. Collect more data from multiple sources and develop research towards big data analysis. The application of analyzing customer opinion reports on the website, especially on mobile devices, helps enterprises more convenient in viewing reports and making better decisions.

References

1. Akila, R., Revathi, S., Shreedevi, G. (2020), "Opinion Mining on Food Services using Topic Modeling and Machine Learning Algorithms", in 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 1071-1076.
2. Akter, S., Aziz, M. T. (2016), "Sentiment analysis on facebook group using lexicon based approach", in 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pp. 1-4.
3. Balan, S., Rege, J. (2017), "Mining for social media: Usage patterns of small businesses", Business Systems Research: The Journal of Society for Advancing Innovation and Research in Economy, Vol. 8 No. 1, pp. 43-50.
4. Burges, C. J. (1998), "A tutorial on support vector machines for pattern recognition", Data mining and knowledge discovery, Vol. 2 No. 2, pp. 121-167.
5. Dunđer, I., Horvat, M., Lugović, S. (2016), "Word occurrences and emotions in social media: Case study on a Twitter corpus", in Biljanović, P. (Ed.), Proceedings of the 39th International Convention on Information and Communication Technology, Electronics and Microelectronics MIPRO 2016, Croatian Society for Information and Communication Technology, Electronics and Microelectronics - MIPRO, Rijeka, pp. 1557-1560.
6. Joachims, T. (2002), Learning to classify text using support vector machines, Springer Science & Business Media.
7. Kadriu, A., Abazi, L., Abazi, H. (2019), "Albanian Text Classification: Bag of Words Model and Word Analogies", Business Systems Research: The Journal of the Society for Advancing Innovation and Research in Economy, Vol. 10 No. 1, pp. 74-87.
8. Khairnar, J., Kinikar, M. (2013), "Machine learning algorithms for opinion mining and sentiment classification", International Journal of Scientific and Research Publications, Vol. 3 No. 6, pp. 1-6.
9. Krstić, Ž., Seljan, S., Zoroja, J. (2019), "Visualization of Big Data Text Analytics in Financial Industry: A Case Study of Topic Extraction for Italian Banks", Entrenova, Vol. 5 No. 1, pp. 67-75.
10. Le, H. S., Trieu, C., Ho, T., Lee, J. H., Lee, H. K. (2017), "Applying Artificial Neural Network for Sentiment Analytics of Social Media Text Data in fastfood industry", Internet e-commerce research, Vol. 17 No. 5, pp. 113-123.
11. Li, Z., Fan, Y., Jiang, B., Lei, T., Liu, W. (2019), "A survey on sentiment analysis and opinion mining for social multimedia", Multimedia Tools and Applications, Vol. 78 No. 6, pp. 6939-6967.
12. Liu, B. (2012), "Sentiment analysis and opinion mining", Synthesis lectures on human language technologies, Vol. 5 No. 1, pp. 1-167.
13. Liu, B. (2017), "Many facets of sentiment analysis", in A practical guide to sentiment analysis, pp. 11-39.
14. Lugović, S., Dunđer, I., Horvat, M. (2016), "Techniques and applications of emotion recognition in speech", in Proceedings of MIPRO, pp. 1278-1283.
15. Maks, I., Vossen, P. (2012), "A lexicon model for deep sentiment analysis and opinion mining applications", Decision Support Systems, Vol. 53 No. 4, pp. 680-688.
16. Mudambi, S. M., Schuff, D. (2010), "What makes a helpful review? A study of customer reviews on Amazon.com", MIS Quarterly, Vol. 34 No. 1, pp. 185-200.

17. Nagpal, M., Kansal, K., Chopra, A., Gautam, N., Jain, V. K. (2020), "Effective Approach for Sentiment Analysis of Food Delivery Apps", in Soft Computing: Theories and Applications, pp. 527-536.
18. Nguyen, H., Ho, T. (2020), "Topic modeling for analyzing online reviews in hotel sector", Science & Technology Development Journal - Economics - Law and Management, Vol. 4 No. 4, pp. 1081-1092.
19. Ohana, B., Tierney, B. (2009), "Sentiment classification of reviews using SentiWordNet", in the 9th IT&T conference, pp. 18-30.
20. Pang, B., Lee, L. (2008), "Opinion mining and sentiment analysis", Foundations Trends Information Retrieval, Vol. 2 No. 1-2, pp. 1-135.
21. Patel, R., Sornalakshmi, K. (2020), "Sentiment Analysis of Food Reviews Using User Rating Score", in Artificial Intelligence Techniques for Advanced Computing Applications, pp. 415-431.
22. Pejić Bach, M., Krstić, Ž., Seljan, S. (2019), "Big data text mining in the financial sector", in Metawa, N., Elhoseny, M., Hassanien, A. E., Hassan, M. K. (Eds.), Expert Systems in Finance: Smart Financial Applications in Big Data Environments, Routledge, pp. 80-96.
23. Sun, S., Luo, C., Chen, J. (2017), "A review of natural language processing techniques for opinion mining systems", Information fusion, Vol. 36, pp. 10-25.
24. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M. (2011), "Lexicon-based methods for sentiment analysis", Computational linguistics, Vol. 37 No. 2, pp. 267-307.
25. Vu, L., Le, T. (2017), "A lexicon-based method for Sentiment Analysis using social network data", in Proceedings of the International Conference on Information and Knowledge Engineering (IKE), pp. 10-16.
26. Vu, T. T., Pham, H. T., Luu, C. T., Ha, Q. T. (2011), "A feature-based opinion mining model on product reviews in Vietnamese", in Semantic Methods for Knowledge Management and Communication, pp. 23-33.
27. Yadav, S. K. (2015), "Sentiment analysis and classification: a survey", International Journal of Advance Research in Computer Science and Management Studies, Vol. 3 No. 3, pp. 113-121.
28. Yang, K., Cai, Y., Huang, D., Li, J., Zhou, Z., Lei, X. (2017), "An effective hybrid model for opinion mining and sentiment analysis", in 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 465-466.

About the authors

Bang Nguyen received a B.S degree in Management Information System from the Faculty of Information Systems, University of Economics and Law (VNU–HCM), Vietnam, in 2020. He is Business Intelligence Specialist at an outsourcing company in Vietnam. His research interests are Business Intelligence, Social media analytics, and Visualization Platforms. He can be contacted at **bangndlk16406@st.uel.edu.vn**

Van-Ho Nguyen received a B.S degree in Management Information System (MIS) from the Faculty of Information Systems, University of Economics and Law (VNU–HCM), Vietnam in 2015, and a Master degree in MIS from the University of Economics Ho Chi Minh City, Vietnam in 2020, respectively. He is currently a lecturer in the Faculty of Information Systems, University of Economics and Law, VNU-HCM, Vietnam. His current research interests include Business Analytics, Business Intelligence, Data Analytics, and Machine Learning. The author can be contacted at **honv@uel.edu.vn**

Thanh Ho received an M.S degree in Computer Science from the University of Information Technology, VNU-HCM, Vietnam, in 2009 and a Ph.D. degree in Computer Science from University of Information Technology, VNU-HCM, Vietnam in 2018. He is currently a lecturer in the Faculty of Information Systems, University of Economics and Law, VNU-HCM, Vietnam. His research interests are Data mining, Data Analytics, Business Intelligence, Social Network Analysis, and Big Data. The author can be contacted at **thanhht@uel.edu.vn**