Sentiment Analysis and Machine Learning-based Classification of Online Consumer Reviews

Xinye Zhou

Business School, Geely University, Chengdu, China
Email: 814408118@qq.com
Manuscript received August 28, 2026; accepted October 21, 2026; published November 10, 2026.

***Abstract*—With the development of electronic commerce and online service systems, consumer evaluations have become an important way to evaluate the quality of products and service experiences. Because of the characteristics of online reviews such as their large number, complex structures and various emotional expressions, this paper studies the sentiment classification of consumer reviews by applying machine learning techniques. First, the 5000 online consumer reviews are processed by eliminating the noise, segmenting the words and deleting the stop-words, and then the TF-IDF method is used to convert the texts into numerical features. Then, Support Vector Machines (SVM) are used as the main classification algorithm to classify the reviews into positive and negative sentiments. Additionally, Logistic Regression and Naive Bayes are introduced as traditional machine learning methods for a comparison. The experimental results demonstrate that the SVM obtains Accuracy, Precision, Recall and F1-Score of 91.27%, 91.54%, 90.76% and 91.15% respectively, which indicates its excellent comprehensive classification ability compared with the other two comparative algorithms. The research suggests that the combination of TF-IDF and SVM can effectively identify the basic sentiments of online consumer reviews, providing useful advice for the analysis of customer feedback and service improvement.**

***Keywords*—online consumer reviews, sentiment analysis, TF-IDF, support vector machines, machine learning**

I. Introduction

With the rapid progress of electronic commerce and online service systems, consumer evaluations have become an important source of information about the quality of products, the degree of customer satisfaction and the service standards. Generally, these evaluations are given in unstructured texts which include useful opinions and demands of the customers. Research indicates that the online review data have different contents and that machine learning techniques can efficiently recognize and analyze consumer attitudes (Jain, Pamula, & Srivastava, 2021). Hence, it is necessary to carry out sentiment analysis on consumer evaluations using natural language processing and machine learning methods. Sentiment analysis can determine the emotional tendencies of consumers from the review texts and convert the complex information into clear classification outcomes. This technique is widely applied in many text cases such as product reviews in electronic commerce and service evaluations on the Internet, where the effectiveness of the classification model is assessed by accuracy, precision, recall and F1-Score (Zhao, Liu, Yao, & Yang, 2021). Considering the features of consumer evaluations, including high dimensionality, sparsity and various types of emotional expressions, this paper extracts features from the review texts by using the TF-IDF method and chooses Support Vector Machines (SVM) as the main classifier to automatically classify positive and negative consumer comments. Moreover, Logistic Regression and Naive Bayes are also introduced as alternative classifiers to verify the classification results of SVM. The performance of the models is evaluated according to accuracy, precision, recall and F1-Score. At last, the experimental results and their practical significance are discussed to offer a reference for the automatic analysis of online consumer feedback.

II. Consumer Review Data Processing and Sentiment Classification Methods

A. Features of Online Consumer Review Data and Text Preprocessing

Online consumer reviews are usually in the form of unstructured texts which contain information about the product quality, function, price, delivery service, after-sale service and personal experience. Compared to structured data, the characteristics of these review texts include a high word number, great variation in expressions and complex emotional tendencies, therefore necessitating preprocessing before they can be applied in machine learning classification. In this paper, the original review data are firstly cleaned by deleting duplicate reviews, abnormal characters and meaningless information to minimize the impact of irrelevant content on the classification results. Then, the reviews are divided into words and those words which are irrelevant to sentiment analysis are removed, while words expressing emotional meanings are kept. After these steps, each review is converted into a relatively standardized text data. The original consumer reviews are processed sequentially through data cleaning, word segmentation and removal of stop words. TF-IDF is used to extract text features, which are then inputted into a SVM model for sentiment classification. The performance of the model is evaluated using classification assessment criteria. By means of this process, the original consumer reviews are gradually transformed into numerical features recognizable by machine learning models, providing a uniform data basis for further classification experiments. Fig. 1 illustrates the basic technical flow of the paper, from the initial reviews to the final sentiment classification outcomes. This flow combines the text processing with the classification model, enabling comparative experiments on the classification abilities of different machine learning methods based on a common data processing platform.

doi:10.18178/joebm.2026.14.4.968

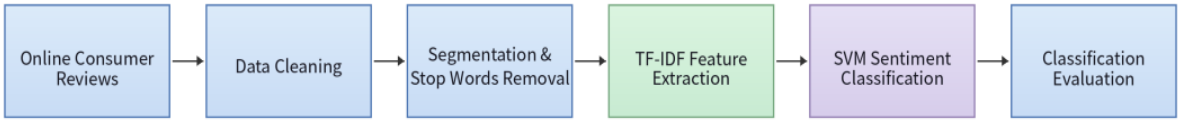


Fig. 1. Basic workflow for online consumer review sentiment classification.

B. Text Feature Representation and Sentiment Classification

After completing text preprocessing, the text must be further converted into a numerical form recognizable by machine learning models. This paper employs the TF-IDF method for text feature representation. TF-IDF assigns different weights to different words by comprehensively considering both the frequency of a word's occurrence within a single review and its prevalence across the entire review corpus (Alantari, Currim, Deng, & Singh, 2022).

Specifically, term frequency reflects the importance of a term within the current review, while inverse document frequency is used to reduce the weight of common terms that appear frequently across a large number of reviews. When a term appears frequently in the current review but rarely in other reviews, it typically possesses strong text discriminative power and thus receives a higher TF-IDF weight. In this way, representative term features can be extracted from a large corpus of review texts.

Suppose the term frequency of the word t in the comment d is $TF(t,d)$, and its inverse document frequency is $IDF(t)$. Then, TF-IDF can be expressed as:

$$TF-IDF(t,d) = TF(t,d) \times IDF(t) \quad (1)$$

Ultimately, each review can be converted into a feature vector composed of the weights of multiple words, which serves as input for machine learning classification models.

In terms of sentiment classification, this paper primarily focuses on the emotional tendencies expressed in consumer reviews, categorizing reviews into corresponding sentiment classes based on existing labels. If the dataset uses binary labels, the classification primarily distinguishes between positive and negative reviews; if the raw data includes a neutral category, multi-class classification is performed according to the actual label configuration. Classification labels should be based on the original dataset to ensure that the experimental results accurately reflect the sentiment distribution of consumer reviews.

It should be noted that while TF-IDF effectively represents the statistical importance of words, it takes limited account of contextual relationships and semantic connections between words (Koufakou, 2024). For example, the same word may express different emotions in different contexts, and traditional term frequency features struggle to adequately capture such differences. Therefore, while this paper employs TF-IDF for basic feature representation, it also analyzes the performance of traditional text classification methods in consumer review sentiment recognition by comparing the results of different machine learning classification models.

C. Support Vector Machine Sentiment Classification Model and Evaluation Metrics

This paper employs the Support Vector Machine (SVM) as the core algorithm for sentiment classification in consumer reviews. SVM determines the sentiment category of a review by identifying a classification boundary in the feature space that effectively distinguishes samples from different categories (Agustina, Novita, Mustakim, & Rozanda, 2024). After TF-IDF processing, consumer reviews typically yield high-dimensional and sparse text features. SVM is well-suited for this research task as it can effectively utilize these features for classification. In the classification process, the review feature vectors generated by TF-IDF serve as the model input, and the sentiment category of the review serves as the output. The model first learns the text features corresponding to different sentiment categories using the training set, and then evaluates its classification performance using the test set. This paper primarily focuses on the classification performance of SVM in consumer review sentiment recognition, while logistic regression and Naive Bayes are used only as auxiliary comparison models to verify the stability of the SVM experimental results. Model performance is evaluated using Accuracy, Precision, Recall, and F1-Score. Accuracy reflects the model's overall classification accuracy; Precision represents the proportion of samples predicted by the model to belong to a certain category that actually belong to that category; Recall reflects the extent to which samples that actually belong to a certain category are correctly identified; the F1-Score comprehensively considers both Precision and Recall, providing a more holistic reflection of the model's classification capability, as shown in Table 1. Through the aforementioned data processing, feature representation, and classification methods, this paper has established an experimental methodology centered on TF-IDF feature extraction and SVM sentiment classification.

Table 1. Major evaluation metrics for machine learning sentiment classification

Evaluation Metric	Main Meaning	Analytical Role
Accuracy	The proportion of correctly classified samples out of the total number of samples	Measures the overall accuracy of classification
Precision	The proportion of samples correctly classified as the target class	Measures the accuracy of classification results
Recall	The proportion of samples that actually belong to the target class that are correctly identified	Measures the ability to identify categories
F1-Score	A composite metric combining Precision and Recall	Comprehensive evaluation of a model's classification performance

III. SVM Sentiment Classification Experiment on Online Consumer Reviews and Analysis of Results

A. Sentiment Distribution of Review Data and Experimental Setup

This paper constructs an experimental dataset comprising 5,000 online consumer reviews, which are categorized into positive and negative reviews based on their overall sentiment. Among these, there are 2647 positive reviews, accounting for 52.94%, and 2353 negative reviews, accounting for 47.06%. The number of reviews in each category is relatively close, with no significant class imbalance, thereby minimizing the impact of sample distribution differences on classification results.

During the data preprocessing stage, the raw reviews were first cleaned to remove duplicates, information lacking substantive meaning, and anomalous characters, followed by text segmentation and stop-word removal. While preserving as many sentiment-expressing terms as possible, the review texts were converted into numerical features using the TF-IDF method. The processed text can be input into various machine learning classification models in the form of feature vectors, providing a unified data foundation for subsequent experiments.

This paper divides the dataset into a training set and a test set in an 80% to 20% ratio. The training set contains 4000 reviews, comprising 2118 positive reviews and 1882 negative reviews; the test set contains 1000 reviews, comprising 529 positive reviews and 471 negative reviews. The training set is used to learn the relationship between comment text features and sentiment categories, while the test set is used to evaluate the model's ability to classify samples not included in the training set.

Judging by the content of the reviews, consumers' emotional expressions exhibit significant diversity. Positive reviews primarily concern product quality, functional performance, user experience, and service attitude, while negative reviews focus more on product defects, logistics experience, price perceptions, and after-sales service issues. Some reviews also include multiple evaluation targets, resulting in a single review containing sentiment information of different orientations. Therefore, consumer reviews are not simply a collection of sentiment words, but rather unstructured text with specific contextual and semantic relationships, which further increases the difficulty of automatic sentiment classification.

B. Comparison of Consumer Review Sentiment Classification Results Between SVM and Traditional Algorithms

After completing the preprocessing of consumer review text and TF-IDF feature extraction, this paper employs SVM as the primary sentiment classification model and selects two relatively traditional text classification algorithms—Logistic Regression (LR) and Naive Bayes (NB)—for comparison. All three models use the same training set, test set, and TF-IDF text features to ensure comparability among the different algorithms. The experimental results are shown in Table 2.

Table 2 Sentiment classification results of different machine learning models

Classification Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	91.27	91.54	90.76	91.15
Logistic Regression	88.64	88.91	87.83	88.36
Naive Bayes	85.73	86.12	84.57	85.33

Based on the experimental results, SVM achieved high classification performance across all four evaluation metrics. Specifically, Accuracy was 91.27%, Precision was 91.54%, Recall was 90.76%, and F1-Score was 91.15%. In comparison, logistic regression achieved an accuracy of 88.64% and an F1-Score of 88.36%, while Naive Bayes achieved 85.73% and 85.33%, respectively. The SVM's accuracy was 2.63 percentage points higher than that of logistic regression and 5.54 percentage points higher than that of Naive Bayes, indicating that the SVM is effective at utilizing text features derived from TF-IDF to distinguish sentiment categories.

In terms of Precision, Recall, and F1-Score, SVM also maintained relatively stable classification performance, indicating that it can not only accurately determine comment categories but also sufficiently identify comments with different sentiment tendencies. Logistic regression and Naive Bayes still possess some classification capability, but their overall results are lower than those of SVM. The above comparison is primarily used to verify the relative stability of SVM's classification results and does not alter SVM's position as the main research algorithm in this paper.

C. Discussion

From the experimental results, the Support Vector Machine (SVM) shows a good classification performance in the sentiment analysis of online customer reviews, obtaining an accuracy of 91.27%, with precision, recall and F1-value being 91.54%, 90.76% and 91.15% respectively. These three indices are quite similar, indicating that the model has a stable classification ability for both positive and negative reviews. Compared with the results of logistic regression and Naive Bayes, the classification effects are relatively lower, which implies that under the TF-IDF text feature conditions of this experiment, SVM has good compatibility with high-dimensional and sparse review texts. In the experiment, all the models can perform a basic sentiment classification, but SVM has higher comprehensive classification efficiency; therefore, it is chosen as the main study algorithm in our paper. Moreover, some reviews cannot be correctly classified due to the existence of negation, change in conversation, implicit sentiment and the coexistence of positive and negative information in the customer reviews. In summary, the experimental results indicate that the sentiment classification method based on TF-IDF and SVM is feasible, but there is still some scope for improvement in recognizing complicated semantic information.

IV. Conclusion

This paper discusses the issue of sentiment classification in online customer reviews by applying TF-IDF to represent the text features and using SVM as the main classification technique to distinguish between positive and negative sentiments in 5000 customer reviews. The experimental outcomes indicate that SVM achieves an accuracy of 91.27%, with precision, recall and F1-Score being 91.54%, 90.76% and 91.15% respectively, which shows a relatively stable classification capability. A comparison with logistic regression and Naive Bayes reveals that SVM performs better than these two methods under the condition of TF-IDF feature in this experimental data set, because it can effectively use the sentiment information in high-dimensional text features. This study also suggests that traditional machine learning methods have certain practical applications in the sentiment analysis of customer reviews, but they still encounter difficulties in recognizing text elements like negation, contrast and complex semantics. In the future, research may collect more review data sources and integrate them with text feature techniques having higher semantic expression power to enhance the sentiment recognition ability for complicated customer reviews. Moreover, the sentiment categories can be divided according to different product and service sectors, so as to make the analysis results reflect more accurately the particular requirements of customers.

Conflict of Interest

The authors declare no conflict of interest.

References

Agustina, C. A. N., Novita, R., Mustakim, Rozanda, N., 2024. The implementation of TF-IDF and Word2Vec in sentiment analysis of booster vaccines using the support vector machine algorithm. ***Procedia Computer Science,*** 234: 156–163.

Alantari, H. J., Currim, I. S., Deng, Y., Singh, S., 2022. An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews. ***International Journal of Research in Marketing,*** 39(1): 1–19.

Jain, P. K., Pamula, R., Srivastava, G., 2021. A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. ***Computer Science Review***, 41: 100413.

Koufakou, A., 2024. Deep learning for opinion mining and topic classification of course reviews. ***Education and Information Technologies,*** 29: 2973–2997.

Zhao, H., Liu, Z., Yao, X., Yang, Q., 2021. A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. ***Information Processing & Management,*** 58(5): 102656.

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).