



UNIVERSITY OF WALES TRINITY SAINT DAVID(UWTSD)

Assignment

Md.Nabil Rahman

ID: 2504501

Course Code: ACCA7025

Module: Application of Machine Learning

Lecturer: Mark Huntly

MSc Software Engineering and Artificial Intelligence

April 17th 2026

Abstract

Sentiment analysis is important for marketing because it helps you understand how people feel about things and make decisions based on data. This study examines the utilization of traditional machine learning techniques for sentiment analysis of customer reviews, employing the Shopper reviews dataset.

customer reviews. We used the Term Frequency–Inverse Document Frequency (TF-IDF) method to change the labeled text reviews in the dataset, which show different sentiment classes, into numbers. A Multinomial Naïve Bayes classifier was used for sentiment classification because it works well and doesn't need much processing power. The experimental results showed that the model did okay, with an accuracy of about 50%. The small sample size, class imbalance, and the difficulty of multi-class classification are the main things that affect this result. Even with these problems, the work shows that traditional machine learning methods can be a simple and efficient way to do sentiment analysis, especially in situations where there aren't many resources available. This study underscores the efficacy and scalability of lightweight machine learning techniques, in contrast to prior research that emphasizes deep learning and extensive language models. The findings underscore the significance of feature engineering, preprocessing, and high-quality data in improving model performance. In conclusion, this study enhances our understanding of the effective application of traditional machine learning techniques for sentiment analysis, achieving a balance between computational efficiency and accuracy.

Contents

Abstract	i
List of Figures	iii
1 Introduction	1
2 Related Work	1
2.1 Traditional Machine Learning Approaches	2
2.2 Comparative Studies of Machine Learning Methods	2
2.3 Feature Representation and Text Processing	2
2.4 Modern Deep Learning and LLM-Based Approaches	2
2.5 Critical Analysis and Justification for Traditional Methods	3
3 Workflow Methodology	3
3.1 Data Description	3
3.2 Data Preprocessing	4
3.3 Feature Extraction using TF-IDF	4
3.4 Model Selection	5
3.5 Model Training	5
3.6 Model Evaluation	5
3.7 System Deployment Considerations	6
4 Discussion and Recommendations	6
4.1 Analysis of Model Performance	6
4.2 Impact of Dataset Size	7
4.3 Effect of Multi-Class Classification	7
4.4 Comparison with Literature	7
4.5 Justification of Selected Approach	7
4.6 Limitations	8
4.7 Future Improvements	8
5 Conclusion	9
6 REFERENCES	10

List of Figures

3.1 shopper reviews dataset 4

3.2 Data Preprocessing 4

3.3 Model Selection: Naïve Bayes 5

3.4 Model Evaluations 6

1 Introduction

Sentiment analysis is now an important tool for businesses that want to learn more about how customers feel and make better decisions. By looking at customer reviews, businesses can find out how happy their customers are, what problems they have, and how to make their service better. As online platforms grow quickly, they create a lot of user-generated content every day. This makes automated sentiment analysis more and more important. Recent advances in AI have led to the development of deep learning and large language model (LLM)-based methods for sentiment analysis. These methods have been very accurate in tough situations [1], [3]. However, these methods might not work for small or new businesses because they need a lot of computing power, big datasets, and special equipment. Traditional machine learning techniques are still important because they work well, are easy to use, and don't need as much computing power, as shown by earlier studies [4], [5]. This study employs traditional machine learning algorithms for sentiment analysis utilizing the Shopper reviews dataset. The main goal is to make a simple, useful, and cheap model that doesn't need deep learning and can work in situations where resources are limited. We use TF-IDF feature extraction to turn the labeled customer reviews in the dataset into numbers.

A Multinomial Naïve Bayes classifier is used because it is cheap to run and works well with text data [6]. Standard classification metrics, such as accuracy, precision, recall, and F1-score, are employed to evaluate the model's performance.

The study's results show that the classification performance is not very good, which shows how hard it is to work with small datasets and categorize sentiment into more than one class. The findings indicate that traditional machine learning methodologies can provide a viable foundational solution for sentiment analysis applications. This study enhances our understanding of the effective application of lightweight machine learning workflows in practical marketing contexts while maintaining computational efficiency.

2 Related Work

Sentiment analysis has become increasingly significant within the domain of Natural Language Processing, particularly in the evaluation of customer feedback for marketing purposes. Recent studies have investigated both modern deep learning methodologies and traditional machine learning, highlighting the trade-offs between computational efficiency and performance.

2.1 Traditional Machine Learning Approaches

Because of their effectiveness and ease of use, traditional machine learning techniques are still often utilized for sentiment analysis. Using airline review data, Fadzal et al. [6] showed that the Multinomial Naïve Bayes algorithm offers dependable classification results at a cheap computational cost. Similar to this, Nalini et al. [10] demonstrated that conventional classifiers can successfully extract insights from product reviews by using machine learning techniques for consumer sentiment analysis.

These results demonstrate that classical models are especially well suited for practical applications with constrained computational resources.

2.2 Comparative Studies of Machine Learning Methods

The efficacy of conventional methods has been further supported by comparative studies. Using both balanced and imbalanced datasets, Corak et al. [4] compared deep learning and conventional machine learning models on Amazon evaluations. According to their findings, traditional machine learning techniques continue to be competitive and perform consistently under various dataset situations, even though deep learning models can attain higher accuracy.

In a similar vein, Wuisang and Pangestu [5] discovered that traditional methods provide a more effective alternative, especially when computational resources are limited, when comparing machine learning techniques with large language models in hotel review analysis.

2.3 Feature Representation and Text Processing

Enhancing classification performance is mostly dependent on feature extraction. Using keyword vector representations, Chu et al. [7] investigated sentiment analysis, highlighting the significance of efficient text representation methods. Despite the existence of more sophisticated representations, conventional techniques like TF-IDF are still frequently employed because of their ease of use and efficiency.

A thorough examination of sentiment analysis methods was given by Sushma et al. [8], who emphasized that feature engineering and preprocessing have a big impact on model performance independent of the approach employed.

2.4 Modern Deep Learning and LLM-Based Approaches

Deep learning and large language models (LLMs) for sentiment analysis have received more attention in recent years. Özdemir et al. [1] achieved good classification accuracy by analyzing hotel

reviews using LLM-based approaches. In a similar vein, Thiagarajan [3] showed how useful LLMs are for analyzing e-commerce product reviews.

Furthermore, Liu et al. [9] shown that deep learning architectures may capture intricate language patterns and enhance prediction performance by proposing a hybrid CNN-LSTM model for sentiment categorization. In order to derive strategic insights from consumer feedback, Mishra et al. [2] investigated machine learning-driven sentiment analysis methods in more detail.

2.5 Critical Analysis and Justification for Traditional Methods

Even if deep learning and LLM-based techniques perform well, they need a lot of processing power, a lot of memory, and specialized hardware like GPUs. They also entail less interpretability and intricate training processes.

Such demands are unfeasible for startups and small marketing firms. Conversely, conventional machine learning methods offer:

- Reduced computational expense
- Reduced training time
- A more straightforward implementation
- Increased interpretability

As a result, even though contemporary methods reflect the state-of-the-art, they are inappropriate for our investigation. Because of their effectiveness and usefulness in settings with limited resources, traditional machine learning techniques are given priority.

3 Workflow Methodology

The Shopper reviews dataset, which comprises customer review text labeled according to sentiment, is used in this study’s structured machine learning methodology to create an effective sentiment analysis model. The workflow is appropriate for contexts with limited resources because it is made to guarantee high classification performance while minimizing computing cost.

3.1 Data Description

Textual consumer reviews with sentiment labels (positive and negative) are included in the Shopper reviews dataset. For the purpose of building impartial classification models, the dataset offers a balanced distribution of classes.

These datasets offer a realistic depiction of consumer input in marketing applications and are frequently utilized in sentiment analysis research [4].

```
# Load Dataset
df = pd.read_csv("shopper_reviews.csv")

print(df.head())
print(df['sentiment'].value_counts())

review sentiment
0 Love this product! Amazing quality and super f... positive
1 Terrible experience. Broke after 2 days and cu... negative
2 It's okay, works as expected but nothing speci... neutral
3 Outstanding value! Highly recommend to all my ... positive
4 Complete waste of money. Doesn't work at all. negative
sentiment
positive 10
negative 10
neutral 10
Name: count, dtype: int64
```

Figure 3.1: shopper reviews dataset

3.2 Data Preprocessing

Because textual data is naturally noisy and unstructured, preparation is necessary to enhance model performance. The preprocessing procedures listed below are used:

- Text conversion to lowercase
- Elimination of punctuation and HTML tags
- Sentences are tokenized into words.
- Getting rid of stopwords
- Lemmatization or stemming

```
# Data Preprocessing
nltk.download('stopwords')
stop_words = set(stopwords.words('english'))

def clean_text(text):
    text = text.lower()
    text = re.sub(r'<.*>', '', text) # remove HTML tags
    text = re.sub(r'["\s\W]', ' ', text) # remove punctuation
    words = text.split()
    words = [word for word in words if word not in stop_words]
    return " ".join(words)

df['clean_review'] = df['review'].apply(clean_text)

print(df[['review', 'clean_review']].head())

review \
0 Love this product! Amazing quality and super f...
1 Terrible experience. Broke after 2 days and cu...
2 It's okay, works as expected but nothing speci...
3 Outstanding value! Highly recommend to all my ...
4 Complete waste of money. Doesn't work at all.

clean_review
0 love product amazing quality super fast delive...
1 terrible experience broke days customer servic...
2 okay works expected nothing special
3 outstanding value highly recommend friends
```

Figure 3.2: Data Preprocessing

These procedures preserve important information while lowering dimensionality and noise. Preprocessing has a major impact on sentiment classification model performance, as shown by Sushma et al. [8].

3.3 Feature Extraction using TF-IDF

The Term Frequency–Inverse Document Frequency (TF-IDF) technique is used to convert the reviews into numerical representations so that machine learning algorithms can process textual data.

TF-IDF improves classification performance by giving greater weight to terms that are less common and more informative across documents. Due to its efficacy and efficiency, this method has been widely used in sentiment analysis jobs [7].

To lower computational cost and enhance model generalization, the feature space is restricted to the most pertinent terms.

3.4 Model Selection

Due to their high computing cost and resource needs, deep learning and neural network-based techniques were rejected in compliance with project requirements.

The following benefits led to the selection of the Multinomial Naïve Bayes algorithm:

```
# Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# Feature Extraction (TF-IDF)
tfidf = TfidfVectorizer(max_features=5000)
X_train_tfidf = tfidf.fit_transform(X_train)
X_test_tfidf = tfidf.transform(X_test)

# Train Model (Naive Bayes)
model = MultinomialNB()
model.fit(X_train_tfidf, y_train)

• MultinomialNB
MultinomialNB()
```

Figure 3.3: Model Selection: Naïve Bayes

- Effective management of text data with several dimensions
- Quick prediction and training
- Minimal computational overhead
- demonstrated efficacy in sentiment analysis assignments

Multinomial Naïve Bayes maintains efficiency while performing well for sentiment categorization, as shown by Fadzal et al. [6].

3.5 Model Training

An 80:20 split was used to separate the dataset into training and testing sets. The model was trained using the training set, and its performance on untested data was assessed using the testing set.

Sentiment labels for the test set were predicted using the Multinomial Naïve Bayes model, which was trained using TF-IDF features taken from the training data.

3.6 Model Evaluation

Standard classification measures are used to assess the model's performance:

```

Accuracy: 0.5

Classification Report:

      precision    recall  f1-score   support

negative    0.00    0.00    0.00         0
neutral    1.00    0.33    0.50         3
positive    0.67    0.67    0.67         3

 accuracy    0.50    0.33    0.50         6
 macro avg   0.83    0.50    0.58         6
 weighted avg 0.83    0.50    0.58         6

Confusion Matrix:

[[0 0 0]
 [1 1 1]
 [1 0 2]]

```

Figure 3.4: Model Evaluations

- Accuracy: Evaluates general correctness
- Precision: Evaluates how accurate positive forecasts are
- Recall: Evaluates the capacity to identify positive examples
- F1-score: Precision and recall harmonic mean

These metrics offer a thorough assessment of the classifier’s performance. Using numerous evaluation measures guarantees a more trustworthy assessment of model performance, as demonstrated by earlier research [4].

3.7 System Deployment Considerations

A marketing analytics solution can use the finished trained model to automatically categorize customer reviews. The model may operate on common hardware without the need for specialized infrastructure because of its cheap computational cost.

For small businesses looking for affordable sentiment analysis solutions, this makes it ideal.

4 Discussion and Recommendations

An overall accuracy of 50% was obtained from the sentiment analysis model’s experimental evaluation utilizing the Shopper reviews dataset. Even though its performance is mediocre, it offers crucial information about how well conventional machine learning techniques work in limited circumstances.

4.1 Analysis of Model Performance

The model performed differently across sentiment classes, according to the classification results. The categorization report and confusion matrix demonstrate that:

- About half of the test samples were properly classified by the model.
- The positive class performed comparatively better, with memory and precision in balance.

- Because of its small representation, the neutral class performed inconsistently.
- Because the negative class was absent from the test set, the assessment metrics' dependability was impacted.

The evaluation as a whole is greatly impacted by this uneven distribution of classes within the test set, which makes the claimed accuracy less reliable and less indicative of actual model performance.

4.2 Impact of Dataset Size

The short dataset size is a major factor affecting the model's performance. Only 30 total reviews and 6 test instances make up the Shopper reviews dataset used in this investigation.

The model's capacity to discover significant patterns and successfully generalize to new data is limited by such a short dataset. A considerable volume of data is necessary to achieve dependable performance in sentiment classification tasks, as previously noted [8].

4.3 Effect of Multi-Class Classification

The categorization process becomes more complex when three sentiment classes—positive, neutral, and negative—are included. The model must be able to differentiate between multiple categories in order to do multi-class classification, which is especially difficult when the dataset is limited.

4.4 Comparison with Literature

Previous research has shown that when trained on big datasets, conventional machine learning models can achieve excellent accuracy in sentiment analysis. For instance, strong performance employing machine learning techniques was reported by Fadzal et al. [6] and Nalini et al. [10].

Further evidence that conventional methods can compete with more complicated models comes from comparative studies like Corak et al. [4]. These studies, however, usually use much bigger datasets, which emphasizes how crucial data accessibility is to attaining high accuracy.

4.5 Justification of Selected Approach

The chosen method utilizing TF-IDF and Multinomial Naïve Bayes is still suitable for the presented situation despite the moderate accuracy. The model provides:

- Minimal computational expense
- Quick prediction and training
- Ease of use and simplicity

These qualities are especially crucial for small marketing firms with constrained processing power.

4.6 Limitations

There are various limitations to the study.

- Generalization is limited by small dataset sizes.
- Evaluation reliability is impacted by class imbalance.
- restricted capacity to provide contextual meaning
- Complexity is increased by multi-class classification.

4.7 Future Improvements

Future research ought to concentrate on:

- Increasing the dataset
- Enhancing preprocessing methods
- Investigating more conventional models, like logistic regression
- Keeping computational efficiency and correctness in check

5 Conclusion

The report has critically assessed With an emphasis on creating an effective solution appropriate for marketing contexts with limited resources, this paper has provided a thorough investigation of sentiment categorization algorithms for customer evaluations. Using the Shopper reviews dataset, this study examined the use of conventional machine learning approaches for sentiment analysis of consumer reviews. The main goal was to provide a low-cost, effective solution that didn't require computationally demanding deep learning models and could be used in contexts with limited resources.

The suggested method offered a straightforward and scalable framework for analyzing consumer sentiment. It was based on TF-IDF feature extraction and Multinomial Naïve Bayes classification. The model performed moderately, with an accuracy of about 50%. The dataset's shortcomings, especially its small size, class imbalance, and the complexity brought forth by multi-class classification, are reflected in this outcome. The study shows that conventional machine learning techniques are still a good choice for sentiment analysis tasks despite the middling accuracy, particularly in situations where CPU resources are few. The suggested approach is appropriate for small marketing firms looking for workable and affordable solutions because of its simplicity, quickness, and low resource requirements.

The results further emphasize how crucial feature engineering, dataset size, and data quality are to model performance. Accuracy could be greatly increased by making changes like expanding the dataset, making the classification issue simpler, and improving feature extraction methods.

In conclusion, whereas sophisticated models like deep learning and huge language models might attain greater accuracy [1], [3], their computing requirements restrict their applicability in specific situations. This study demonstrates that conventional machine learning techniques offer a fair trade-off between economy and performance, which makes them ideal for practical applications where scalability and simplicity are crucial factors.

6 REFERENCES

- [1] A. O. Özdemir, E. B. Giritli, and Y. S. Can, “Sentiment analysis for hotel reviews in Turkish by using LLMs,” in Proc. 2024 9th International Conference on Computer Science and Engineering (UBMK), Antalya, Turkiye, 2024, pp. 210–214, doi: 10.1109/UBMK63289.2024.10773456.
- [2] R. Mishra, N. Tyagi, M. Gulhane, N. Rakesh, and B. Balusamy, “Machine learning-driven aspect-based analysis of sentiment to make strategic inferences from customer feedback,” in Proc. 2025 IEEE 4th World Conference on Applied Intelligence and Computing (AIC), Gwalior, India, 2025, pp. 712–717, doi: 10.1109/AIC66080.2025.11212190.
- [3] K. Thiyagarajan, “Leveraging large language models for sentiment analysis in e-commerce product reviews,” in Proc. 2025 6th International Conference on Artificial Intelligence, Robotics and Control (AIRC), Savannah, GA, USA, 2025, pp. 234–239, doi: 10.1109/AIRC64931.2025.11077508.
- [4] M. I. Corak, C. Wang, and X. Zhang, “A comparative study of sentiment analysis on Amazon customer reviews: Deep learning vs. machine learning with balanced and imbalanced datasets,” in Proc. 2025 IEEE 16th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), Yorktown Heights, NY, USA, 2025, pp. 0014–0020, doi: 10.1109/UEMCON67449.2025.11267657.
- [5] M. C. Wuisang and G. Pangestu, “Aspect-based sentiment analysis in Indonesian hotel reviews: Machine learning versus large language models,” in Proc. 2025 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2025, pp. 444–449, doi: 10.1109/iSemantic67418.2025.11292470.
- [6] A. N. Fadzal, A. F. Maliki, N. Fadzal, W. M. I. Wan Mohammad, and A. N. Fadzal, “Sentiment analysis of airline reviews using Multinomial Naïve Bayes algorithm,” in Proc. 2025 6th International Conference on Artificial Intelligence and Data Sciences (AiDAS), West Java, Indonesia, 2025, pp. 234–239, doi: 10.1109/AiDAS67696.2025.11213697.
- [7] C.-H. Chu, C.-A. Wang, Y.-C. Chang, Y.-W. Wu, Y.-L. Hsieh, and W.-L. Hsu, “Sentiment analysis on Chinese movie review with distributed keyword vector representation,” in Proc. 2016 Conference on Technologies and Applications of Artificial Intelligence (TAAI), Hsinchu, Taiwan, 2016, pp. 84–89, doi: 10.1109/TAAI.2016.7880169.
- [8] S. Sushma, S. K. Nayak, and M. V. Krishna, “A comprehensive review of sentiment analysis: Trends, challenges, and future directions,” in Proc. 2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI), Tirunelveli, India, 2024, pp. 1175–1181, doi: 10.1109/ICDICI62993.2024.10810919.

[9] F. Liu, H. Xiao, and R. Liu, “Research on customer review text sentiment analysis based on CNN-LSTM networks,” in Proc. 2024 7th Asia Conference on Cognitive Engineering and Intelligent Interaction (CEII), Singapore, 2024, pp. 183–190, doi: 10.1109/CEII65291.2024.00044.

[10] T. N. P. Nalini, N. Sathiyendran, and R. S., “Consumer sentiment analysis towards Lluvia bike products using machine learning approaches,” in Proc. 2025 First International Conference of Advances in Engineering and Computing Technologies for Sustainable Development (AECTSD), Coimbatore, India, 2025, pp. 1–6, doi: 10.1109/AECTSD65988.2025.11411514.