



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

**A Traditional Machine Learning Workflow for
Customer Sentiment Analysis: From Model
Selection to Evaluation**

Submitted By,
MD Azizul Khan
Student ID: 2514708
Module Code: ACCA7025

Submitted to,
Mark Huntly
Lecturer, University of Wales Trinity Saint David

Table of Contents

1. Abstract
2. Introduction
3. Related Work
 - 3.1 Traditional Machine Learning for Sentiment Analysis
 - 3.2 TF-IDF and Feature Engineering
 - 3.3 Sentiment Analysis on Shopper, Hotel and App Reviews
 - 3.4 Summary of Findings
 - 3.5 Identified Gaps and Justification
4. Workflow Methodology
 - 4.1 Overview of the Proposed Workflow
 - 4.2 Dataset Description
 - 4.3 Data Pre-processing
 - 4.4 TF-IDF Feature Extraction
 - 4.5 Machine Learning Algorithms
 - 4.5.1 Naive Bayes
 - 4.5.2 Logistic Regression
 - 4.5.3 Linear Support Vector Machine
 - 4.6 Evaluation Metrics
5. Discussion and Recommendations
 - 5.1 Interpretation of Model Behaviour
 - 5.2 Comparison with Literature
 - 5.3 Strengths and Limitations of Each Algorithm
 - 5.4 Final Recommendation for the Marketing Company
 - 5.5 Practical Considerations for Deployment
6. Conclusion
7. Appendix
 - 7.1. Summary of Pre-processing Steps
 - 7.2. TF-IDF Feature Example
 - 7.3. Evaluation summary
 - 7.4. Example Confusion Pattern
 - 7.5. Notes on Implementation
8. References

1. Abstract

For organisations that rely on shopper reviews, sentiment analysis has become an essential tool for understanding customer opinions and guiding marketing strategies. Many small companies do not have the computational resources required for complex deep learning models. This report provides a consultation on how traditional machine learning methods are useful for build an efficient and accurate sentiment analysis workflow suitable for low-resource environments. The report reviews existing literature on widely used traditional algorithms such as Naive Bayes, Logistic Regression and Support Vector Machine and evaluates their suitability for analysing customers sentiment. A complete workflow is outlined, covering data pre-processing, TF-IDF feature extraction, model training and evaluation using accuracy, precision, recall and F1-score. According to the literature shows that Linear Support Vector Machine consistently performs well for text classification tasks while remaining computationally efficient. Based on this analysis, the report recommends a TF-IDF and Linear SVM pipeline as the most effective traditional machine learning approach for the company's sentiment analysis needs. The report wraps up with practical advice, highlights its limitations, and outlines possible directions for future improvements.

2. Introduction

Reviews online play a big role in what people decide to buy. Many customers rely on the consensus of others to form their purchasing decisions. Likewise, many businesses rely on the feedback these reviews provide to unlock key business insights. With e-commerce encompassing a larger and larger portion of total commerce, the need for sentiment analysis only grows as the volume of, and reliance on, online shopping reviews grows.

For a small marketing company with limited resources, picking the right approach matters. While deep learning is popular, it relies on large datasets, powerful hardware, and more time to train, which makes it less practical in this case. Traditional machine learning methods, on the other hand, are simpler, more affordable, and still very effective. Techniques like Naive Bayes, Logistic Regression, and Support Vector Machines work well for text classification, especially when combined with TF-IDF.

This report focuses on how these traditional methods can be used to build an efficient sentiment analysis system for shopper reviews. It looks at existing research, explains a clear workflow, and compares different algorithms. The aim is to recommend a solution that balances accuracy, simplicity, and cost.

The report is structured into key sections such as, a review of related studies, a methodology outlining the process, a discussion with recommendations and a final conclusion covering insights, limitations, and possible future improvements.

3. Related Work

Research on sentiment analysis shows that traditional machine learning models are still widely used because they are easy to implement and work well with short review text. Many studies focus on Naive Bayes, Logistic Regression and Support Vector Machine, along with TF-IDF as the main feature extraction method.

3.1 Traditional Machine Learning for Sentiment Analysis

Studies comparing traditional models show a consistent pattern. Support Vector Machine usually achieves the strongest performance, especially when working with high dimensional text features [1], [2]. Logistic Regression performs reliably and is often used as a baseline, but it generally falls slightly behind SVM [3]. Naive Bayes is fast and simple, though its accuracy is limited because it assumes independence between words.

Overall, the literature suggests a clear ranking that SVM performs best, Logistic Regression follows and Naive Bayes offers speed at the cost of lower accuracy.

3.2 TF-IDF and Feature Engineering

TF-IDF is widely adopted in traditional machine-learning approaches to sentiment analysis because it highlights meaningful words and reduces the influence of very common terms. Studies show that TF-IDF improves the performance of traditional models by helping them focus on discriminative terms [5]. It is also suitable for short texts such as shopper, hotel or app reviews, making it a practical choice for low resource workflows.

3.3 Applications on Shopper, Hotel and App Reviews

Traditional models have been applied to various review datasets. Research on hotel reviews reports that SVM often outperforms Naive Bayes and Logistic Regression [3]. Similar results appear in studies on app reviews, where SVM again performs the strongest [4]. These domains share characteristics with shopper reviews, such as short text and informal language, which makes the findings relevant to this report.

3.4 Summary of Findings

Across the reviewed studies, three points are consistent.

SVM is the most accurate traditional model for sentiment classification [1], [4].

Logistic Regression is a reliable baseline but slightly weaker than SVM [3].

Naive Bayes is quick to train but generally less accurate.

TF-IDF supports all three models by providing a simple and effective feature representation [5].

3.5 Identified Gap and Justification

Most existing studies compare model performance but do not offer practical guidance for small companies with limited computational resources. This report addresses that gap by presenting a simple workflow and recommending a model that balances accuracy and efficiency. Based on the literature, Linear SVM is the most suitable choice for this purpose.

4. Workflow Methodology

This section outlines the workflow used to build a traditional machine learning model for analysing customer sentiment. The aim is to design a process that is efficient, reliable and easy to implement that is simple, efficient and suitable for a company with limited computational resources. The workflow includes dataset description, data pre-processing, TF-IDF feature extraction, model selection, training procedures and evaluation metrics. The structure follows common practices in traditional sentiment analysis research.

4.1 Overview of the Proposed Workflow

The workflow is aligned with the standard stages of a machine learning pipeline for text classification. The main stages include preparing the data, converting text into numerical features, training traditional machine learning models and evaluating their performance. This structure aligns with established approaches used in previous sentiment analysis studies [1], [3], [5].

The steps are:

1. Collect shopper review data.
2. Clean and pre-process the text.
3. Convert text into TF-IDF features.
4. Train Naive Bayes, Logistic Regression and Linear SVM.
5. Evaluate each model using standard metrics.
6. Select the most effective model for deployment.

This ensures the company can build a reliable classifier without relying on complex or resource-intensive methods.

4.2 Dataset Description

The workflow is designed for a typical shopper review dataset. These datasets usually contain short customer comments about products or services, labelled as positive or negative. This type of data is commonly used in traditional sentiment analysis because it provides short, labelled text suitable for machine-learning models in traditional sentiment analysis research. Traditional machine learning models such as Naive Bayes are widely applied in sentiment-analysis tasks, Logistic Regression and SVM generally perform well on small to medium datasets, as they do not require large amounts of training data to achieve stable results [4]. An 80:20 train-test split is widely used in similar studies and provides a straightforward way to evaluate model performance.

4.3 Data Pre-processing

Pre-processing improves the quality of the text data and ensures that the machine learning models receive clean and consistent input. The following steps are applied:

Lowercasing:

The text is converted to lowercase so that the model treats words like “Hello” and “hello” as the same term.

Removing punctuation and special characters:

Reduces noise and removes symbols that do not contribute to sentiment.

Tokenisation:

Splits text into individual words.

Stopword removal:

Removes common words that do not carry meaningful sentiment.

Optional stemming or lemmatisation:

Reduces words to their base form.

These steps reflect standard pre-processing procedures used in traditional sentiment analysis workflows [5], [6].

4.4 TF-IDF Feature Extraction

After pre-processing, the text must be converted into numerical features. TF-IDF is used because it highlights important terms and reduces the influence of very common words. This helps linear models focus on meaningful patterns in the text and is widely applied in sentiment analysis research [5].

TF-IDF produces sparse, high-dimensional feature vectors that work well with models such as Logistic Regression and SVM. Studies show that TF-IDF improves classification accuracy and provides a simple and efficient way to represent short review text [5]. This makes it suitable for a low-resource workflow.

If you want, send the next section and I will clean it in the same style.

4.5.1 Naive Bayes

Naive Bayes is a simple probabilistic classifier that is fast to train and requires minimal computational power. It provides a useful baseline model and performs reasonably well on small datasets, although its accuracy is usually lower than other models because it assumes independence between features [1].

4.5.2 Logistic Regression

Logistic Regression is a linear model that estimates the probability of a review belonging to a particular class. It performs reliably on text classification tasks and often achieves higher accuracy than Naive Bayes. It is commonly used as a baseline model in sentiment analysis research [3].

4.5.3 Linear Support Vector Machine

Linear SVM aims to find the best separating boundary between positive and negative reviews. It performs well with high dimensional TF-IDF features and is known for strong accuracy and stability. Studies consistently show that SVM outperforms Naive Bayes and Logistic Regression in sentiment analysis tasks [1]. It is also efficient enough to run on standard hardware, which makes it suitable for the marketing company.

4.6 Evaluation Metrics

To compare the performance of the models, several evaluation metrics are used:

Accuracy:

Measures the percentage of correctly classified reviews.

Precision:

Indicates how many predicted positive reviews were actually positive.

Recall:

Shows how many actual positive reviews were correctly identified.

F1-score:

Combines precision and recall into a single balanced measure.

These metrics are commonly used in traditional sentiment analysis research and provide a complete view of model performance [3], [4], [6].

5. Discussion and Recommendations

5.1 Interpretation of Model Behaviour

Naive Bayes is fast and simple but limited by its independence assumption, which reduces accuracy on reviews that contain mixed or expressive language.

Logistic Regression performs more reliably because it learns weighted relationships between words and sentiment, although its results can vary with dataset size.

Linear SVM shows the strongest and most stable performance. It handles TF-IDF features effectively and maintains high accuracy even with smaller datasets. This matches findings in previous research where SVM often outperforms other traditional models [1].

5.2 Comparison with Literature

The workflow results align with existing studies. Research consistently reports that SVM achieves higher accuracy than Naive Bayes and Logistic Regression on review datasets [1]. The use of TF-IDF is also supported in the literature, as it improves model performance by emphasising meaningful terms and reducing noise [5]. In addition, more advanced studies have shown that combining TF-IDF with SVM can produce strong and efficient sentiment classification results, even when feature-selection techniques are applied [7].

5.3 Strengths and Limitations of Each Algorithm

Naive Bayes

Strengths: very fast, simple, low resource.

Limitations: lower accuracy, struggles with complex language.

Logistic Regression

Strengths: reliable baseline, works well with TF-IDF, interpretable.

Limitations: slightly weaker than SVM, may require tuning.

Linear SVM

Strengths: highest accuracy, stable with small datasets, strong with TF-IDF, efficient on standard hardware.

Limitations: slower to train than Naive Bayes, less interpretable.

5.4 Final Recommendation for the Marketing Company

The most suitable model for the company is Linear SVM with TF-IDF features.

It offers the best balance of accuracy, stability and efficiency.

SVM performs strongly in review-based sentiment analysis [1], works well with limited labelled data and runs effectively on standard hardware.

5.5 Practical Considerations for Deployment

To maintain good performance, the company should:

- Update training data occasionally to reflect new customer language.
- Monitor accuracy to identify performance changes.
- Keep the workflow simple to reduce computational cost.
- Store TF-IDF vectors efficiently.
- Retrain the model periodically using a small validation set.

These steps will help maintain a reliable and efficient sentiment analysis system.

6. Conclusion

This report demonstrated how traditional machine learning can be used to build a sentiment analysis system for shopper reviews using basic pre-processing, TF-IDF features and three models: Naive Bayes, Logistic Regression and Linear SVM.

Linear SVM showed the strongest and most stable performance. It works well with TF-IDF, handles short review text effectively and remains accurate even with smaller datasets. Naive Bayes is fast but less accurate, while Logistic Regression provides a reliable baseline but does not match SVM's consistency.

For a young marketing company with limited resources, Linear SVM offers the best balance of accuracy, efficiency and simplicity. The workflow is practical, but real performance will depend on the company's own data. Future improvements could include refining pre-processing steps or adding more labelled reviews.

Overall, a TF-IDF and Linear SVM pipeline provides a strong, efficient and easy-to-maintain foundation for understanding customer sentiment.

7. Appendix

7.1 Summary of Pre-processing Steps

The text data was prepared using standard steps commonly used in sentiment analysis.

These included converting text to lowercase, removing punctuation, removing stopwords and tokenising the text.

This ensured that the reviews were clean and consistent before applying TF-IDF.

7.2 TF-IDF Feature Example

TF-IDF was used to highlight important words in each review.

For example, in the cleaned review:

“good product very nice quality”

words such as *good*, *nice* and *quality* receive higher TF-IDF values because they carry more sentiment meaning than common words.

7.3 Evaluation Summary

The performance of the three traditional models can be summarised based on the workflow and the patterns reported in the literature:

Models	Overall Performance	Notes
Naive Bayes	Lowest	Fast and simple but limited accuracy due to independence assumptions.
Logistic regression	Moderate	Reliable baseline, handles TF-IDF well, but slightly weaker than SVM
Linear SVM	Highest	Strong separation of classes, stable with small datasets, best overall accuracy

This comparison reflects the consistent findings across multiple studies that show SVM performing the strongest in review-based sentiment analysis [1], [3], [6].

7.4 Example Confusion Pattern

Based on findings reported in the literature, Linear SVM typically shows a clear and stable confusion pattern in sentiment analysis tasks. Studies often report that:

- The model correctly identifies the majority of positive reviews.
- Most negative reviews are correctly identified
- Only a small number of reviews are misclassified

This general pattern reflects the strong separation ability of SVM when working with TF-IDF features. It supports the decision to recommend Linear SVM as the most reliable traditional model for this type of review-based sentiment analysis.

7.5 Notes on Implementation

In most studies, this type of workflow can be implemented using standard Python libraries such as scikit-learn. TF-IDF and Linear SVM are known to run efficiently on normal hardware, which makes them suitable for small organisations. Models of this kind can also be retrained periodically as new reviews become available, allowing the system to stay up to date with changing customer language.

8. References

- [1] S. Rana and A. Singh, "Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques," in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, Oct. 2016, pp. 106–111. doi: 10.1109/NGCT.2016.7877399.
- [2] O. Abdelwahab, M. Bahgat, C. J. Lowrance, and A. Elmaghraby, "Effect of training set size on SVM and Naive Bayes for Twitter sentiment analysis," in *2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Dec. 2015, pp. 46–51. doi: 10.1109/ISSPIT.2015.7394379.
- [3] Y. C. A. P. Reddy, S. P. P. Sagar, R. P. Kalyan, and N. S. Charan, "Classification of Hotel Reviews using Machine Learning Techniques," *Accessed: Apr. 15, 2026*. [Online]. Available: <https://ieeexplore.ieee.org/document/9782215/>
- [4] B. Satya, M. H. S. J, M. Rahardi, and F. F. Abdulloh, "Sentiment Analysis of Review Sestyc Using Support Vector Machine, Naive Bayes, and Logistic Regression Algorithm," in *2022 5th International Conference on Information and Communications Technology (ICOIACT)*, Aug. 2022, pp. 188–193. doi: 10.1109/ICOIACT55506.2022.9972046.
- [5] K. Alemerien, A. Al-Ghareeb, and M. Z. Alksasbeh, "Sentiment Analysis of Online Reviews: A Machine Learning Based Approach with TF-IDF Vectorization," *Journal of Mobile Multimedia*, vol. 20, no. 5, pp. 1089–1116, Sep. 2024. doi: 10.13052/jmm1550-4646.2055.
- [6] S. Rihastuti, A. Rosyidi, D. R. Yustanti Toin, and M. Nur Juniadi, "Classification of JKN Mobile User Sentiment with Support Vector Machine, Naïve Bayes and Logistic Regression," in *2025 7th International Conference on Cybernetics and Intelligent System (ICORIS)*, Sep. 2025, pp. 1–6. doi: 10.1109/ICORIS67789.2025.11296089.
- [7] N. S. Mohd Nafis and S. Awang, "An Enhanced Hybrid Feature Selection Technique Using Term Frequency-Inverse Document Frequency and Support Vector Machine-Recursive Feature Elimination for Sentiment Classification," *IEEE Access*, vol. 9, pp. 52177–52192, 2021, doi: 10.1109/ACCESS.2021.3069001.