



RECOMMENDATION OF A SENTIMENT ANALYSIS ML MODEL FOR MARKETING INSIGHTS



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



STUDENT NAME: MUHAMMAD HAROON
STUDENT ID: 2434025
MODULE CODE: ACCA7025
TITLE: APPLICATION OF MACHINE LEARNING

Table of Contents

Abstract.....	2
1. Introduction	2
2. Related Works	3
3. Workflow Methodology.....	5
3.1. Data Acquisition and Labelling	5
3.2. Data Pre-processing	6
3.3. Feature Engineering	6
3.4. Model Training and Selection Strategy	7
3.5. Model Evaluation Metrics	7
4. Discussion and Recommendations	8
4.1. Anticipated Model Performance and Justification	8
4.2. Final Recommendation: Support Vector Machine (Linear Kernel)	9
4.3. Recommended Workflow and Deployment.....	9
4.4. Limitations and Future Directions	10
5. Conclusion	10
6. References.....	11

Abstract

The growth of user generated content especially shopper reviews is a tremendous opportunity to the marketing firm to be able to get real-time feedback on the consumer sentiment. This report is a detailed consultation on how one can apply traditional machine learning (ML) workflows to sentiment analysis with the consideration of computational efficiency and resource limitations. The report compares and contrasts the established ML algorithms including Naive Bayes, Support Vector Machines, and Logistic Regression through a critical review of the literature. It is determined that to have a balanced mode of performance, interpretability and computational economy, a Support Vector Matrix (SVM)-based model with a linear kernel, trained on a set of features based on TF-IDF vectorized unigrams and bigrams, would be the best recommendation. In the report, a full ML workflow, including the data acquisition, pre-processing, feature engineering, model training, and its strict evaluation on the basis of such measures as accuracy, precision, recall, F1-score, and the AUC-ROC curve is thoroughly described. This recommendation can be explained by the fact that it allows achieving high classification accuracy and has a comparatively simple and inexpensive operational pipeline that corresponds to the need of the client to use a classic, non-deep learning algorithm.

1. Introduction

In the modern online market, the voice of the customer is more powerful and sound than ever. The overload of user-generated content in e-commerce websites, social media, and review sites in particular, summarizes the consumer sentiment, feelings, and experiences with the products and services. This information is a treasure trove to the marketing organisations; the capacity to use this sentiment in a systematic way to analyse and interpolate it may lead to effective strategic decision making, better management of customer relationships, improvement in product development and strengthening of brand reputation [1]. However, the amount, speed, and diversity of this data make it impossible to analyse it manually and require the implementation of automated calculation tools.

Sentiment analysis or opinion mining. The computational research of opinions, sentiments, attitudes and emotions in written language expressed by people. It is essentially a text classification problem, which normally classifies text as positive,

negative, or neutral polarities [2]. Though the more recent developments have led to the development of sophisticated deep learning architectures, such as Recurrent Neural Networks (RNN) or Transformers (e.g., BERT), they may require large computational resources, large labelled data, and substantial skills to implement and maintain [3]. Such models can be too costly and may be inefficient to a young marketing company that is trying to reduce the cost of computation and complexity of its operation.

In this report, the particular problem presented by the marketing company is to prescribe a standard machine learning algorithm and workflow of sentiment analysis of shopper review. The essence is to find a model that would strike a good tradeoff between predictive power and computational simplicity. It will critically examine the landmarks of conventional ML techniques, provide a rationale as to why a particular algorithm has been chosen, and outline a well-developed, end-to-end process of sentiment classifier development, evaluation, and deployment. The recommendation will be careful not to need neural networks and deep learning but only highly established, interpretable, resource-efficient models that have proved to be effective in the academic literature and industrial practice [4]. The following paragraphs will examine related work, describe the proposed methodology and give a final discussion and a recommendation based on a critical analysis of the available evidence.

2. Related Works

Sentiment analysis by the use of machine learning has been the focus of several studies spanning more than 20 years. Before the development of deep learning, the classical supervised learning algorithms dominated the field and made use of the statistical features derived on the base of the text. The effectiveness of these models is very reliant on the quality of the engineering of the features that convert raw text into a numerical format that could be processed by algorithms.

One of the most relevant works in the field is that of Ogutu [5], who introduced the use of machine learning to sentiment analysis of movie reviews. Their experiment considered three of the classic classifiers; Naive Bayes (NB), Maximum Entropy (MaxEnt) and Support Vector Machines (SVM). Their results were significant and it was established that SVM performed significantly better than NB and MaxEnt on this task. This made SVM a model of text classification to be used as a benchmark and it has retained this to a great

extent. The paper emphasized that simple unigrams (single words) as features might be very effective in sentiment analysis, refuting the assumption of the time that more extensive linguistic analysis was required.

Naive Bayes classifier is a simple and fast classifier still popular as a baseline because it is surprisingly useful on numerous text classification problems. It is based on the Bayes theorem but with the naive assumption of independence between features. Other studies, including those by Abbass et al. [6], have examined NB variations, i.e. Multinomial and Bernoulli Naive Bayes, on sentiment tasks. Although this can be much less accurate than SVM, NB is computationally very inexpensive, and does not need as much data to get reasonable performance which makes it a contender in resource-constrained settings.

Another popular model is the Logistic Regression (LR) that is valued due to its probabilistic response and the inherent interpretability. In contrast to NB, LR does not rely on feature independence and is more capable of depicting more complex feature-to-feature relationships. Compared to LR, the studies which compare LR and SVM including those by Shobayo et al. [7] in the case of sentiment analysis in social media tend to be competitive with the difference in the performance determined by the particular dataset and feature set. LR offers probability estimates that are well-calibrated and this can be very essential in business decisions where the levels of confidence matter.

Other than the algorithm, feature representation is an important aspect. The most basic model is the Bag-of-Words (BoW) model, where words are not ordered, but their frequency is counted. It is refined by its Term Frequency-Inverse Document Frequency (TF-IDF) which values words based on their significance in a document compared with the rest of the corpus, which has the effect of down-ranking words that are common but less informative [8]. The combination of n-grammes (sequences of n consecutive words, e.g., bigrams) may be used to retrieve certain phrasal and contextual information, e.g. not good, lost in a unigram model. The studies have always indicated that the use of bigrams can yield a significant increase in the accuracy of sentiment classification [9].

Moreover, the necessity of a strict data pre-processing pipeline is pointed out throughout the literature. This involves such steps as tokenization, lowercasing, deleting stop-words (such common words as the, is), and stemming/lemmatization (cutting words down to

their root form) [10]. These procedures are useful in minimizing noise and dimensionality, which results in stronger and generalizable models.

In brief, the literature unanimously confirms that the classical ML-based methods, and specifically SVM, NB, and LR, with attentive feature engineering (such as TF-IDF and n-grammes) and careful pre-processing, can be used to carry out proper sentiment analysis quite effectively. Their main benefit compared to deep learning models in this regard is that they are less expensive in terms of computation, both training and inference, have shorter development times, and are sometimes better interpretable hence are best suited to the requirements of the given marketing company.

3. Workflow Methodology

An effective machine learning project has a specific set of features, which are identified not only by the algorithm type, but also by a logical and repeatable workflow. The suggested flow of creating the sentiment analysis model is implemented in the form of a pipeline, where the steps of data collection, pre-processing, feature engineering, model training and selection, and overall evaluation are arranged. Such systematic method guarantees the methodological rigor and allows developing an effective and precise model.

3.1. Data Acquisition and Labelling

The initial and the most important measure is to obtain a labelled data of shopper reviews. In case of a supervised learning task, both the quality and quantity of the labelled data are the most important.

- Source: The company may use the historical data of reviews that have been manually rated (e.g., 1-5 stars). Alternatively, publicly accessible datasets, including the Amazon Product Review dataset or the Yelp Dataset may be utilised when developing a preliminary model and performing benchmarking [11].
- Labelling Scheme: A well-defined labelling scheme should be created. To be brief and efficient, it is suggested to use a ternary system: Positive, Negative and Neutral. An example can be given of a 5-star system that 1-2 stars can be reflected to Negative, 3 stars to Neutral, and 4-5 star to Positive. This strategy helps curb subjectivity that is likely to be encountered with finer-grained scales.

3.2. Data Pre-processing

Unstructured raw text data is noisy. It requires a uniform set of pre-processing tools to clean the data before it is ready to be extracted into a feature.

- Tokenization: Division of the content of each review into single words or tokens.
- Lowercasing: Making all the text lowercase so that there is consistency in it (e.g. great and Great are the same token).
- Purge of Noise: This involves the removal of HTML tags, URLs, punctuations and non-alphabetic characters. It is also possible to delete user handles (e.g. @username) and hashtag symbols, but the text of a hashtag can be kept.
- Stop-word Removal: Removing high-frequency and low-information words (e.g. then, and, is, etc.) by using a standard stop-word list. This minimizes the feature space dimensions.
- Lemmatization: It means to reduce words to their basic word form (lemma). In the case of running, one can say it is running, better turns into good. Lemmatization is better than the more vulgar stemming because it forms linguistically more correct roots [12].

3.3. Feature Engineering

This stage converts the purged text to a numerical feature vector which devolves to processing by ML algorithms. The proposed feature set is the combination of:

- TF-IDF Vectorization: TF-IDF features are generated in the pre-processed tokens into a matrix. The approach indicates the significance of a word to a document in a collection. It is effective in down-weighting words that are presented in all documents (e.g., "product," "buy") that are not as discriminative as to sentiment.
- N-gramme Range: The TF-IDF vectorizer should be set to extract the unigrams (1-grammes), but also the bigrams (2-grammes) to obtain the individual meanings of words as well as any short-range contextual phrases. This enables this model to acquire the following features such as excellent quality (positive) and stopped working (negative) among others which are pivotal in the accurate identification of sentiment [13]. The vocabulary can be managed through a minimum document frequency which will eliminate extremely rare words.

3.4. Model Training and Selection Strategy

No a priori algorithm will be selected but a comparative analysis of the top candidate will be done to make a recommendation based on the data. The models that can be proposed to be evaluated are:

- Multinomial Naive Bayes (MNB): It is selected on the basis of its speed, simplicity, and efficiency with TF-IDF features.
- Support Vector Machine (SVM) using Linear Kernel: Selected due to its proven state of the art when dealing with high dimensional and sparse text data [14].
- Logistic Regression (LR): Selected due to its probabilistic interpretation and good performance which is frequently comparable to SVM.

The following protocol will be followed in the process of training:

- Data Splitting: The complete labelled dataset will be divided in a training (e.g. 70 percent) and a held-out test (e.g. 30 percent) set. The test set will not be touched at all in the model development and tuning phase, and would act as the final and objective performance measure.
- Cross-Validation and Hyperparameter Optimization: A k-fold cross-validation (e.g., k=5 or 10) loop will be run on the training set to optimize the hyperparameters of each of the models. As an example, regularization parameter C of SVM and LR is essential to avoid overfitting. Tuning with cross-validation can be used to ensure that a model is useful on unseen data and the optimal set of parameters is chosen in each algorithm [15].
- Imbalance in the Classes: When the dataset has a high imbalance (many positive reviews than negative ones), methods like modifying the weights of the classes in the models (e.g., `class_weight=balanced` in scikit-learn) will be utilised when training. This makes sure that the model does not favor the majority class.

3.5. Model Evaluation Metrics

Correctness is an inadequate measure in itself, particularly when there are skewed samples. It is proposed to use a multi-faceted evaluation strategy:

- Confusion Matrix: This is a table that has a finer breakdown of the true positives and true negatives, false positives and false negatives as per class.
- Precision, Recall, and F1-Score: These statistics provide more of an idea.

- Precision (to the Positive class): How many of all reviews that were predicted to be positive were really positive? High accuracy is essential when the cost of a false positive action (e.g. marketing a falsely positive review) is high.
- This is in memory (on the Positive class): How many of the actually positive reviews was the model, correctly, identifying? The concept of high recall is crucial when a false negative (the cost of a missed positive sentiment) is expensive.
- F1-Score: A single count that is a harmonic mean of precision and recall, and which gives the balance between the two.
- Accuracy of overall: The percentage of the total correct predictions. This works well as a summary statistic where the distribution of the classes is not very skewed.
- Area Under Rectangle Curve (AUC-ROC): It is used to assess a model in the way it can differentiate the classes in all classification levels. An increase in AUC proves to be superior performance.

4. Discussion and Recommendations

This section is the synthesis of the outcomes that can be expected basing on the critical analysis of the literature and on the suggested workflow and gives the final, justified offer to the marketing company.

4.1. Anticipated Model Performance and Justification

The comparative assessment described in the methodology is likely to bring the results that are in accordance with the existing literature. The actual performance is data-dependent, and a general hierarchy will tend to appear in that Linear SVM and Logistic Regression will probably have the best, and most similar, accuracy and F1-scores, which Multinomial Naive Bayes will not. The difference between the performances of SVM/LR and NB can be often explained by the fact that the latter is based on the naive independence of features when by definition in text, words depend on each other.

A marginal improvement in performance is not the only matter on which the final model has been selected. It entails a compromise between performance, computational cost and interpretability.

- Multinomial Naive Bayes: It may be the fastest to train and least resource-intensive, though with lower expected accuracy, who is not the ideal choice to use as a primary model, though it is a great baseline.

- **Logistic Regression:** This and the most desired feature are a good performance as well as the fact that it produces probabilities that are well-calibrated (e.g., "This review is 92% likely to be positive"). It is a considerable business strength as it can enable the marketing group to gauge the sentiment level. Its learning period is usually effective.
- **Support Vector Machine (Linear Kernel):** The Support Vector Machine is the one that always performs well in the text classification. It has especially an advantage in high-dimensional spaces and is resistant to overfitting, particularly when regularization is done correctly.

4.2. Final Recommendation: Support Vector Machine (Linear Kernel)

The Linear Support Vector Machine is recommended as the optimal algorithm for the marketing company's sentiment analysis task. The justification is threefold:

- **Proven High Performance:** The body of research, from Saxena et al. [16] to contemporary studies, consistently demonstrates that linear SVM is a top-tier performer for sentiment analysis with TF-IDF features. It is highly likely to deliver the classification accuracy required for actionable business insights.
- **Computational Efficiency:** A linear SVM is significantly more efficient to train and run for prediction than non-linear kernels or deep learning models. Once trained, the classification of new reviews is a simple dot product operation, making it suitable for deployment on systems with limited resources. This directly addresses the client's core requirement to minimize computational cost [17].
- **Effective Regularization:** The C parameter in SVM provides a straightforward mechanism to control the trade-off between maximizing the margin and minimizing classification error on the training data. A well-tuned C value effectively prevents overfitting, ensuring the model generalizes robustly to new, unseen shopper reviews.

4.3. Recommended Workflow and Deployment

All the workflow outlined in Section 3 would have to be put into practice. The model is also to be trained using the reduced and aggregated data, and the ultimate hyperparameters are chosen through cross-validation. The withheld test set will give the ultimate performance indicators to test the model prior to its implementation.

To implement it, a pre-processing and feature extraction pipeline that is automated should be constructed. The new review must undergo the same tokenization, lemmatization, and TF-IDF transformation procedures when it comes to the new review [18]. The transformed feature vector is then used in prediction by the saved SVM model to predict the sentiment. It can be containerized (e.g. with Docker) and deployed into a low-cost cloud service or even on-premise hardware.

4.4. Limitations and Future Directions

Although the suggested conventional ML method is the best in the described limitations, it is not without restrictions. The quality of the feature engineering limits the performance of the model. It is unable to comprehend long-range dependencies, sarcasm and intricate linguistic nuances as more sophisticated (and expensive) Transformer models can [19]. Moreover, the model can need some retraining every couple of days with new data to follow the changing language and slangs.

With the expansion of the company and the available resources, the future may have:

- New Embeddings: Word2Vec and GloVe are pre-trained word vectors which may be a better feature to use and may capture more information about the semantics [20].
- Aspect-Based Sentiment Analysis: Beyond the sentiment on a document level, to the sentiment on a feature of the product (e.g. battery life vs. screen quality).
- Ensemble Methods: This consists of a combination of predictions of SVM, LR, and other models in a possible effort to obtain a slight performance increase [21].

5. Conclusion

In this report, a full consultation has been given on how to implement a classical machine learning workflow of sentiment analysis of shopper reviews. The literature review revealed that classical models are still very efficient and are particularly effective where computational efficiency and resource limitation are the main issues. The comprehensive procedure of workflow was introduced, and the essential steps of data preparation, the feature engineering using TF-IDF and n-grammes, and the strict model evaluation plan were presented.

The fundamental conclusion of this analysis is that it is highly recommended to use a Linear Support Vector Machine (SVM) model. This suggestion is supported by the fact that the algorithm has a history of high accuracy, it can be efficiently used in training and

in the deployment process, and that it is resistant to overfitting. With the suggested workflow, the marketing company is able to create an effective sentiment analysis system that not only makes it easier to use, but also economical enough, so that they can learn to use customer feedback as an engine of informed marketing and business development and expansion, without the baggage of complicated deep learning systems.

6. References

- [1] T. A. Rainy, M. A. Rahman, and A. J. Mou, "CUSTOMER RELATIONSHIP MANAGEMENT AND DATA-DRIVEN DECISION-MAKING IN MODERN ENTERPRISES: A SYSTEMATIC LITERATURE REVIEW," *American Journal of Advanced Technology and Engineering Solutions*, vol. 4, no. 04, pp. 57–82, 2024.
- [2] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current State of Text Sentiment Analysis from Opinion to Emotion Mining," *ACM Comput. Surv.*, vol. 50, no. 2, pp. 1–33, Mar. 2018, doi: 10.1145/3057270.
- [3] I. D. Mienye and T. G. Swart, "A comprehensive review of deep learning: Architectures, recent advances, and applications," *Information*, vol. 15, no. 12, p. 755, 2024.
- [4] M. J. Ahmed, A. Mozo, and A. Karamchandani, "A survey on graph neural networks, machine learning and deep learning techniques for time series applications in industry," *PeerJ Computer Science*, vol. 11, p. e3097, 2025.
- [5] R. V. A. Ogutu, "Ensemble Model for Target Sentiment Analysis Based On Naïve Bayes and Support Vector Machine for Product Review Classification," PhD Thesis, JKUAT-COPAS, 2023. Accessed: Oct. 15, 2025. [Online]. Available: <http://ir.jkuat.ac.ke/handle/123456789/6117>
- [6] M. Abbas, K. A. Memon, A. A. Jamali, S. Memon, and A. Ahmed, "Multinomial Naive Bayes classification model for sentiment analysis," *IJCSNS Int. J. Comput. Sci. Netw. Secur*, vol. 19, no. 3, p. 62, 2019.
- [7] O. Shobayo, S. Adeyemi-Longe, O. Popoola, and B. Ogunleye, "Innovative sentiment analysis and prediction of stock price using FinBERT, GPT-4 and logistic regression: A data-driven approach," *Big Data and Cognitive Computing*, vol. 8, no. 11, p. 143, 2024.
- [8] M. K. Kim, A. Zouaq, and S. M. Kim, "Automatic detection of expert models: The exploration of expert modeling methods applicable to technology-based assessment and instruction," *Computers & Education*, vol. 101, pp. 55–69, 2016.
- [9] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach," *Expert systems with applications*, vol. 57, pp. 117–126, 2016.
- [10] S. Manda and B. V. Vardhan, "A Systematic Literature Review on Sentimental Analysis using Machine Learning for Preprocessing, Feature Extraction and Classification," 2024, Accessed: Oct. 15, 2025. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5061012
- [11] N. Asghar, "Yelp Dataset Challenge: Review Rating Prediction," May 17, 2016, *arXiv*: arXiv:1605.05362. doi: 10.48550/arXiv.1605.05362.

- [12] V. K. Pant, R. Sharma, and S. Kundu, "An overview of stemming and lemmatization techniques," *Advances in Networks, Intelligence and Computing*, pp. 308–321, 2024.
- [13] G. Kaur and A. Sharma, "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis," *J Big Data*, vol. 10, no. 1, p. 5, Jan. 2023, doi: 10.1186/s40537-022-00680-6.
- [14] N. Cristianini and B. Scholkopf, "Support vector machines and kernel methods: the new generation of learning machines," *Ai Magazine*, vol. 23, no. 3, pp. 31–31, 2002.
- [15] O. A. Montesinos López, A. Montesinos López, and J. Crossa, "Overfitting, Model Tuning, and Evaluation of Prediction Performance," in *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, Cham: Springer International Publishing, 2022, pp. 109–139. doi: 10.1007/978-3-030-89010-0_4.
- [16] R. R. Saxena, "Examining Reactions about COVID-19 Vaccines: A Systematic Review of Studies Utilizing Deep Learning for Sentiment Analysis," *Authorea Preprints*, 2024, Accessed: Oct. 15, 2025. [Online]. Available: <https://www.techrxiv.org/doi/full/10.36227/techrxiv.171262869.98334885>
- [17] A. Julia, E. Sundararajan, and Z. Othman, "Cloud computing service composition: A systematic literature review," *Expert systems with applications*, vol. 41, no. 8, pp. 3809–3824, 2014.
- [18] A. Tabassum and R. R. Patil, "A survey on text pre-processing & feature extraction techniques in natural language processing," *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 06, pp. 4864–4867, 2020.
- [19] E. Hashmi and S. Y. Yayilgan, "Multi-class hate speech detection in the Norwegian language using FAST-RNN and multilingual fine-tuned transformers," *Complex Intell. Syst.*, vol. 10, no. 3, pp. 4535–4556, Jun. 2024, doi: 10.1007/s40747-024-01392-5.
- [20] P. Kumar, S. Harinisubashree, and A. Helen Shalini, "Analysis of Market Trends with Sentimental Analysis using Word2Vec and GloVe," in *2025 International Conference on Visual Analytics and Data Visualization (ICVADV)*, IEEE, 2025, pp. 125–130. Accessed: Oct. 15, 2025. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10960977/>
- [21] M.-W. Huang, C.-W. Chen, W.-C. Lin, S.-W. Ke, and C.-F. Tsai, "SVM and SVM ensembles in breast cancer prediction," *PloS one*, vol. 12, no. 1, p. e0161501, 2017.