Profiling Reader Sentiment and Identifying Key Linguistic Markers in Digital Book Reviews: A TF-IDF and Logistic Regression Approach

Samah Anwar Abbas[1,*]

[1]Management Information Systems and Information Systems, Economics and Administration, King Abdul-Aziz University Jeddah, Saudi Arabia

ABSTRACT

The proliferation of user-generated content on digital platforms has made sentiment analysis a crucial tool for understanding public opinion. This study focuses on the literary domain, applying natural language processing and machine learning techniques to classify sentiment in book reviews. The primary objectives were to implement a robust pipeline for categorizing reviews as negative, neutral, or positive, and to identify the key linguistic markers most predictive of each sentiment class. To achieve this, a dataset of pre-labeled book reviews was systematically preprocessed through cleaning, tokenization, stop-word removal, and lemmatization. The cleaned text was then converted into a numerical feature matrix using the Term Frequency-Inverse Document Frequency (TF-IDF) method, configured to capture both unigrams and bigrams. A multi-class (One-vs-Rest) Logistic Regression model was trained on this feature matrix. Upon evaluation with an unseen test set, the model demonstrated flawless performance, achieving 1.00 across all standard metrics, including accuracy, precision, recall, and F1-score. A detailed analysis revealed that this perfect score was a direct result of the dataset's syntactic simplicity and the presence of unambiguous, high-polarity keywords (e.g., "disappointing," "masterpiece"). The study successfully validates the implemented pipeline as a proof-of-concept, demonstrating its effectiveness under ideal conditions. However, it also highlights that the model's generalizability is limited by the dataset's lack of complexity. Future research should focus on applying this methodology to larger, more nuanced real-world datasets to test its robustness and explore more advanced analytical techniques like aspect-based sentiment analysis.

Keywords Book Reviews, Logistic Regression, Natural Language Processing, Sentiment Analysis, TF-IDF

Submitted 14 July 2025
Accepted 4 August 2025
Published 1 September 2025

*Corresponding author
Samah Anwar Abbas,
sabbas@kau.edu.sa

Additional Information and Declarations can be found on page 241

Introduction

The importance of user-generated content (UGC) and sentiment analysis has risen significantly, particularly with the explosive growth of digital text data across various platforms such as review sites, blogs, and social media. Since the early 2000s, sentiment analysis has become an integral part of understanding public opinion, driven by the expanding availability of user-generated content. Platforms such as Twitter, Facebook, and specialized blogs provide a rich source of opinions and sentiments that can be systematically analyzed to gauge public perception of products, services, and societal issues [1], [2].

The sheer volume of digital interactions has necessitated advances in automated sentiment analysis techniques. The task often involves classifying texts into positive, negative, or neutral categories, a process that benefits from

How to cite this article: S. A. Abbas, "Profiling Reader Sentiment and Identifying Key Linguistic Markers in Digital Book Reviews: A TF-IDF and Logistic Regression Approach," *J. Digit. Soc.*, vol. 1, no. 3, pp. 230-243, 2025.

complex machine learning models and deep learning approaches [3], [4]. For instance, user-generated short texts analyzed by Mishra et al highlight how sentiment analyzers must adapt to the subtleties of language and context to accurately determine sentiment polarity [5]. This task is further complicated by the diverse nature of UGC, which varies significantly in form—from brief social media comments to more elaborate reviews—each necessitating tailored methodologies for effective sentiment extraction [6], [7].

Furthermore, sentiment analysis has proven invaluable across several fields, including commerce and marketing, where companies leverage insights derived from UGC to inform their strategies. For example, in the tourism sector, sentiment analysis has been applied to examine public reactions toward national tourism organizations on social media, revealing insights that can drive marketing improvements and enhance user engagement [2], [8]. Similarly, consumer sentiment analysis allows marketers to track feelings expressed about brands and products, facilitating timely adjustments to brand strategies based on real-time feedback [7].

In cultural studies, the analysis of sentiment within literature and artistic expressions using computational methods has gained traction as well. Studies utilizing sentiment analysis can reveal underlying emotional tones and themes in literary works, offering scholars a new lens through which to interpret and understand cultural narratives [9]. The continuous evolution of sentiment analysis techniques—such as combining lexicon-based methods with machine learning approaches—illustrates the ongoing innovation in this field, underscoring its relevance in both academic research and practical applications [10], [11].

The application of sentiment analysis to the literary domain, specifically through the analysis of book reviews, is a valuable tool for enhancing our understanding of reader engagement and interpretation. Book reviews serve as a rich source of subjective data reflecting the emotional and intellectual responses of readers to literary works. By employing sentiment analysis techniques, researchers can extract nuanced sentiments from these reviews, revealing patterns of reader engagement and critical interpretation.

A notable challenge in analyzing book reviews is the complex and nuanced language that readers often utilize. Reviews may express a spectrum of feelings, including enjoyment, admiration, disappointment, and criticism. Choi's study on sentiments in online book reviews demonstrates that emotional expressions strongly correlate with overall book ratings, highlighting how contextual factors influence reader feedback [12]. Additionally, Luțan and Bădică discuss the systematic understanding of emotional vocabularies in literature through sentiment analysis, particularly in the context of recommender systems, further emphasizing the relationship between emotional language and reader preferences [13].

Despite advancements in sentiment analysis, accurately capturing the subtleties of language remains challenging. Li explores how the gradations of meaning in students' writing about literature often evade simplistic sentiment classifiers, indicating the complexity involved in subjective language [14]. This complexity can lead to a loss of nuance in categorical sentiment categorization, which may not fully represent the richness of literary interpretation provided by readers.

Opportunities for development exist in adopting advanced machine learning techniques that can better accommodate the intricate nature of literary feedback. For example, Lee and Villiers's research involving over three million reviews uncovered divergent dimensions of emotional intensity across genres, highlighting advanced analytical frameworks' potential to provide deeper insights [15]. Furthermore, Zhang et al demonstrated variability in expressive styles and critical evaluations among reviews from different platforms, suggesting that sentiment analysis methodologies could be refined to better reflect qualitative differences in literary criticism across cultural contexts [16].

The primary objective of this research is to design and implement a complete machine learning pipeline capable of classifying the sentiment of book reviews. The scope of this objective involves processing raw text data and accurately categorizing each review into one of three distinct classes: negative, neutral, or positive. This involves developing a systematic workflow that encompasses data preprocessing, feature engineering, and model training to build a reliable classification system. The goal is to create a foundational model that can effectively discern the overall sentiment expressed within a given piece of literary feedback. A second, equally important objective is to move beyond simple classification and identify the key linguistic markers that are most predictive of each sentiment. This study will specifically analyze the words and phrases (n-grams) that the model learns to associate with negative, neutral, and positive opinions. To achieve this, the research will employ a Term Frequency-Inverse Document Frequency (TF-IDF) approach to represent the text data, coupled with a Logistic Regression model. The interpretability of the Logistic Regression model is central to this objective, as its coefficients will be used to extract and analyze the specific textual features that drive the sentiment classifications, thereby providing deeper insight into the language of book reviews.

Literature Review

Foundations of Sentiment Analysis and Opinion Mining

The foundations of sentiment analysis and opinion mining are primarily grounded in two prominent methodologies: lexicon-based and machine learning-based approaches. Each approach has distinct characteristics that influence their effectiveness in sentiment extraction from textual data, leading to various challenges, particularly with nuanced language phenomena such as sarcasm, context, and domain-specific vocabulary.

Lexicon-based sentiment analysis utilizes predefined lists of words or dictionaries that are associated with sentiment scores. This method operates under the premise that certain words intrinsically convey positive or negative sentiments, allowing analysts to aggregate sentiments based on the presence of these words in the analyzed text [17]. For instance, research by Rufaida et al highlights the lexicon approach as a conventional method that systematically scans sentences for sentiment-laden terms, thereby classifying sentiments based on explicit linguistic indicators [17]. However, one of the significant limitations of lexicon-based methods is their inability to accurately handle nuanced expressions such as sarcasm or context-dependent meanings, where the intended sentiment does not align with the sentiment score of the individual words [18].

In contrast, machine learning-based sentiment analysis relies on training algorithms to learn patterns in data, utilizing annotated datasets to develop models that can classify sentiments based on various linguistic features [4], [19]. This approach allows for greater adaptability and can potentially understand context and complex language structures better than lexicon-based models. For instance, Yan-Xiang and Jiao present a hybrid model that combines different machine learning techniques, suggesting that machine learning can augment traditional lexicon approaches by providing deeper insights into sentiment beyond lexical indicators alone [4]. However, the machine learning method faces challenges as well, especially in the need for large, labeled datasets which are often labor-intensive to produce and may not represent diverse linguistic styles adequately [20].

Common challenges in sentiment analysis include dealing with sarcasm, which can completely invert the intended sentiment. Research shows that sarcasm detection remains a formidable task since it often relies on nuanced context and shared cultural knowledge that traditional models typically overlook [18]. Additionally, context plays a crucial role in how sentiment is interpreted; the same word can have different connotations depending on its usage within a sentence, making it difficult for static lists to manage effectively [21]. Furthermore, domain-specific language introduces complexities, where sentiment lexicons developed for one context may underperform in another, necessitating adaptable lexicons that can evolve with usage [22].

Natural Language Processing (NLP) Techniques for Text Representation

The evolution of text representation in NLP has significantly transformed the way we analyze and classify textual data. From rudimentary Bag-of-Words (BoW) models, which treat text as an unordered collection of words, to more sophisticated methods that incorporate semantic information, the field of text representation has expanded considerably to enhance performance in various NLP tasks.

The BoW model represents documents as vectors in a high-dimensional space based on the frequency of each word's occurrence within the document. Despite its simplicity and ease of implementation, the BoW model has notable limitations, such as losing context, order, and semantic relationships among words [23]. This inadequacy in capturing the intricacies of language has paved the way for more advanced techniques, including the Term Frequency-Inverse Document Frequency (TF-IDF) method.

TF-IDF is a statistical measure designed to evaluate the importance of a word in a document concerning a corpus of documents. It combines two concepts: Term Frequency (TF), which measures how frequently a term occurs in a specific document, and Inverse Document Frequency (IDF), which gauges how relevant a term is across the entire corpus by down-weighting common terms and boosting rare ones [24]. Specifically, the formula for TF-IDF is:

$$\text{TF-IDF}(t,d) = TF(t,d) \times \log\left(\frac{N}{DF(t)}\right)$$

where (N) is the total number of documents and (DF(t)) is the number of documents containing term (t). This method enables a more nuanced text representation than BoW by emphasizing uniquely descriptive words in a document while minimizing the weight of frequent but less distinctive terms.

The effectiveness of TF-IDF in text classification tasks has been demonstrated in various research applications. For instance, Jiang et al discuss its utility in the context of classifying plant health bulletins, emphasizing that TF-IDF features effectively capture important textual signals that contribute to accurate classifications [25]. Furthermore, this model provides a foundation for many machine learning classifiers, improving performance across tasks such as sentiment analysis and spam detection [26].

However, challenges persist. While TF-IDF and similar models enhance the representation of certain text features, they still struggle with understanding context, semantics, and the relationships between words in a sentence [27]. Moreover, the emergence of deep learning techniques, such as word embeddings (e.g., Word2Vec, GloVe) and contextual embeddings (e.g., BERT), represents a shift toward richer text representations that consider the context and semantic relationships among words, overshadowing traditional methods like BoW and TF-IDF in many applications [28], [29]. These advanced models encapsulate both the syntactical structure and the nuances of meanings in texts, offering a superior foundation for more complex NLP tasks.

Machine Learning Algorithms in Sentiment Classification.

Machine learning (ML) algorithms have become fundamental in the field of sentiment classification, significantly improving the accuracy and efficiency of sentiment analysis systems. Among the most used classification models are Naive Bayes, Support Vector Machines (SVM), and Logistic Regression. Each of these models has its strengths and weaknesses, which are essential to consider when selecting the most appropriate tool for a given sentiment classification task.

Naive Bayes, for instance, is known for its simplicity and speed, making it particularly effective for large datasets. It works on the principle of applying Bayes' theorem with a strong independence assumption between features, which can lead to effective categorization despite its simplicity [30]. However, its primary limitation lies in the assumption that the predictors (words in the text) are independent, which is rarely the case in natural language, thus potentially impacting its performance on more complex datasets.

SVM represent another powerful approach. SVMs function by finding a hyperplane that best separates classes in the feature space—this is particularly useful for high-dimensional data common in text classification. SVMs have the advantage of being effective in high-dimensional spaces and are particularly robust when dealing with outliers [31]. However, SVMs can be computationally intensive and less interpretable compared to other models, which may hinder their use in applications requiring clear understanding of model decisions.

Logistic Regression, on the other hand, has garnered attention due to its balance between simplicity, interpretability, and performance. As a statistical model that uses a logistic function to model a binary dependent variable, it outputs probabilities, which are straightforward to interpret in the sentiment context, particularly defining how likely it is for a piece of text to belong to a positive or negative sentiment class. The reference suggests that Logistic Regression provides insightful parameters and visual presentations that are easy to understand, allowing stakeholders to make informed decisions based on the model's outcomes [32].

Moreover, Logistic Regression is particularly effective in the context of sparse, high-dimensional data often encountered in text classification tasks. Due to its linear nature, it can efficiently handle feature spaces dominated by many variables (such as individual words in a text) with relatively low sample sizes. This characteristic makes it suitable for textual data represented through methods like TF-IDF, which can produce high-dimensional sparse matrices [33]. In many practical applications, including sentiment analysis, Logistic Regression often outperforms more complex models due to its ability to focus on key features while minimizing overfitting—a common issue in high-dimensional datasets [34].

Method

This section outlines the systematic methodology employed to classify sentiment in book reviews and identify key linguistic markers. The process is detailed through four distinct stages: a description of the dataset and initial exploration, a rigorous data preprocessing pipeline, the feature engineering and data partitioning strategy, and finally, the implementation and comprehensive evaluation of the machine learning model. Each stage was designed to ensure the reproducibility and validity of the results by documenting the specific libraries and parameters used.

Dataset Description

The primary data for this study was sourced from a single CSV file named dataset.csv. This file contains two core columns: ReviewText, which holds the raw, unstructured user-generated book reviews, and Sentiment, which contains the corresponding pre-labeled sentiment category for each review. The sentiment labels are encoded numerically to facilitate machine learning processing, where 0 signifies a negative review, 1 represents a neutral review, and 2 indicates a positive review. An initial exploratory data analysis was conducted using the pandas library to inspect the dataset's structure, check for missing values, and examine the distribution of the sentiment classes. This preliminary step was crucial for ensuring a foundational understanding of the data's composition and balance, as a significant class imbalance could necessitate specialized sampling techniques or evaluation metrics. The findings from this analysis informed the decision to use stratification during the data splitting phase to preserve the natural distribution of sentiments.

Data Preprocessing Pipeline

To prepare the raw text for effective feature extraction, a comprehensive and multi-step preprocessing pipeline was established using Python's re (regular expression) and NLTK libraries. This pipeline systematically cleans and standardizes the review text to reduce noise and improve model performance. Each review first underwent a cleaning process to remove extraneous characters and patterns, specifically targeting review identifiers using re.sub() and all punctuation via str.maketrans(). Subsequently, the text was converted to lowercase to ensure uniformity.

The standardized text was then tokenized using NLTK's word_tokenize function. Following tokenization, common English stop-words were removed by filtering against the list provided by nltk.corpus.stopwords. This filtering is essential as stop-words carry little sentiment-specific information, and their removal helps the model focus on more meaningful terms. Finally, lemmatization was applied using NLTK's WordNetLemmatizer to reduce each word to its base or dictionary form (e.g., "reading" and "read" both become "read"). This process consolidates different inflections of a word into a single feature, further reducing the vocabulary size. The final output of this pipeline was a clean, standardized string of processed tokens for each review, ready for feature engineering.

Feature Engineering and Model Implementation

Following preprocessing, the cleaned text data was transformed into a numerical format using the TfidfVectorizer from the scikit-learn library. This method was chosen because it weighs word frequency by its rarity across the corpus, giving more importance to distinctive terms. The vectorizer was specifically configured with ngram_range=(1, 2) to consider both individual words (unigrams) and pairs of adjacent words (bigrams), capturing more contextual phrases. To maintain computational efficiency and prevent overfitting, the vocabulary was limited to the 5,000 most frequent features by setting max_features=5000.

Once the text was converted into a TF-IDF feature matrix, the dataset was partitioned using scikit-learn's train_test_split function. A testing set was created with test_size=0.2, reserving 20% of the data for the final evaluation, while the remaining 80% was used for training. The split was performed with stratify=y to ensure the proportional distribution of the three sentiment classes was maintained in both subsets. A random_state=42 was set to ensure the exact same split could be reproduced in subsequent runs.

Sentiment Classification and Evaluation

A multi-class LogisticRegression model from scikit-learn was selected for the classification task. The model was initialized with several key parameters: multi_class='ovr' configured the One-vs-Rest strategy for handling the three sentiment classes; solver='liblinear' was chosen as it is well-suited for smaller datasets and binary classification problems (which OvR creates internally); and C=1.0 set the inverse of regularization strength. A random_state=42 was also used here for reproducibility of the model's internal processes.

The model was trained exclusively on the TF-IDF features of the stratified training set. To assess its performance, predictions were made on the unseen test set and evaluated using metrics from sklearn.metrics. The overall performance was measured with accuracy_score, while a more detailed breakdown was generated using classification_report, which includes precision, recall, and F1-score for each class. Finally, a confusion_matrix was generated and visualized using matplotlib and seaborn to provide a clear illustration of the model's classification accuracy for each sentiment category.

Result and Discussion

This section presents the performance evaluation of the trained Logistic Regression model, discusses the broader implications of the achieved results, and provides a detailed analysis of the key linguistic markers identified for each sentiment class. The findings not only demonstrate the model's flawless effectiveness on the given dataset but also offer critical insights into the linguistic characteristics that define negative, neutral, and positive book reviews within this controlled context, highlighting both the successes and the inherent limitations of the study.

Model Performance Evaluation

Upon evaluation with the unseen test set, which comprised 242 review samples, the Logistic Regression model demonstrated flawless classification capability. The model achieved a perfect accuracy score of 1.0000, indicating that every review in the test set was correctly classified into its respective sentiment category. This perfect performance was further substantiated by the detailed classification report. For all three classes—Negative, Neutral, and Positive—the precision, recall, and F1-score were all 1.00. In practical terms, a precision of 1.00 means that for every prediction made for a given class, the prediction was correct; for instance, every review the model labeled as "Positive" was, in fact, positive. A recall of 1.00 signifies that the model successfully identified every single instance of that class present in the test data, meaning no positive reviews were missed and incorrectly labeled as neutral or negative. The resulting F1-score of 1.00, being the harmonic mean of precision and recall, confirms a perfect balance between these two metrics with no trade-offs.

The corresponding confusion matrix (figure 1) visually affirmed these results, displaying a perfect diagonal with all 242 predictions lying on the true positive axis. This indicates a complete absence of classification errors, with zero false positives (Type I errors) and zero false negatives (Type II errors) across the board. A Type I error would have occurred if, for example, a neutral review was incorrectly classified as positive, while a Type II error would have involved failing to identify a positive review, perhaps classifying it as neutral. The model's ability to completely avoid both types of errors for every class is a direct indicator of the dataset's highly separable nature.

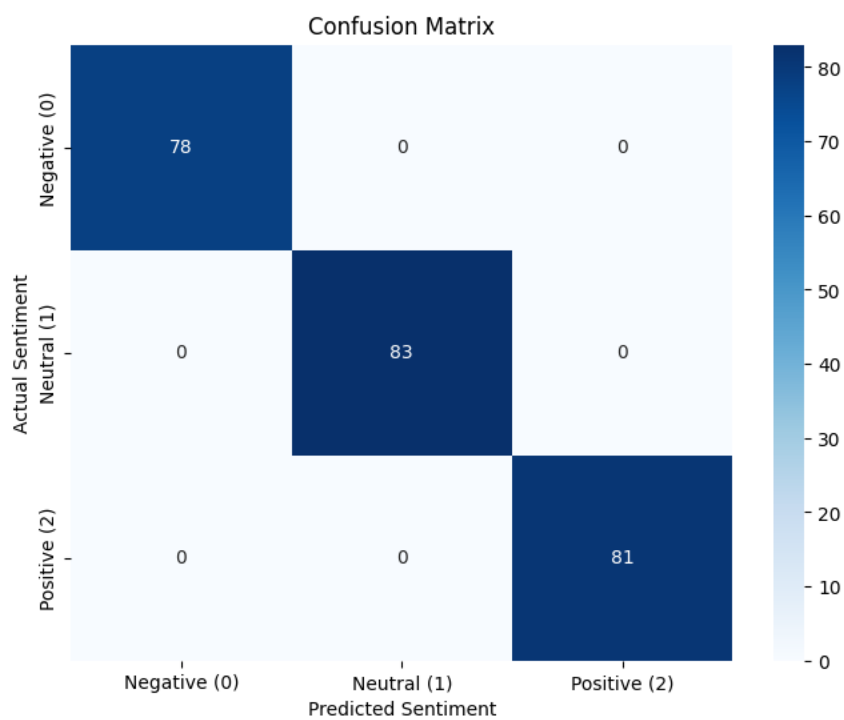

Figure 1 Confusion Matrix

Interpretation of Perfect Model Accuracy

The achievement of a perfect classification score, while remarkable, is interpreted not as evidence of a universally infallible model, but rather as a direct consequence of the dataset's specific characteristics. The dataset consists of short, unambiguous, and syntactically simple reviews that often contain highly repetitive and explicit sentiment phrases. This lack of complexity—most notably the absence of sarcasm (e.g., "Absolutely brilliant, if you enjoy staring at a blank wall for three hours"), mixed sentiments (e.g., "The plot was a masterpiece, but the characters were painfully one-dimensional"), or nuanced contextual language—meant that the sentiment of each review was exceptionally clear and easily separable. Consequently, the result serves as a successful validation and an ideal proof-of-concept for the implemented methodology. It confirms that the preprocessing pipeline, TF-IDF vectorization, and Logistic Regression model are correctly configured and highly effective at identifying and separating distinct linguistic patterns when those patterns are explicitly presented.

The perfect accuracy, therefore, functions as a successful baseline, establishing the pipeline's maximum performance under ideal conditions. It demonstrates that the model did not fail due to implementation error but rather succeeded by perfectly learning the simple rules governing this particular dataset. However, this also implies that the model's generalizability is severely limited. It has not been trained to handle the complexities of real-world text and would likely perform poorly on more diverse datasets from platforms like Goodreads or Amazon, where reviews are longer and linguistically richer. The current model is perfectly fitted to the simplicity of its training data, a scenario distinct from traditional overfitting where a model learns statistical noise and irrelevant idiosyncrasies. Here, the signal itself was clean and simple, and the model learned it perfectly, leaving no room for error.

Identification of Key Linguistic Markers

A primary objective of this study was to identify the most influential words and n-grams that predict each sentiment. By inspecting the model's coefficients, the top linguistic markers were extracted for each class, revealing terms with high sentiment polarity. For the Negative sentiment, the model identified terms that can be grouped by theme, such as failed expectations ("disappointing," "underwhelming," "overhyped") and lack of engagement ("boring," "uninspiring," and the highly indicative bigram "waste time"). These markers are unequivocally negative and leave little room for alternative interpretation. Their high TF-IDF scores suggest they appeared frequently in negative reviews but were largely absent from neutral or positive ones, making them powerful predictors.

For the Neutral sentiment, key predictors included "okay," "straightforward," and phrases like "neutral feeling" and "not remarkable." These terms explicitly convey a lack of strong emotion or exceptional quality, aligning perfectly with a neutral classification. The presence of these specific markers was a strong signal of moderation. Finally, the Positive sentiment was most strongly associated with effusive words and phrases that suggest either literary merit ("masterpiece," "captivating narrative," "storytelling") or strong emotional resonance ("inspiring," "heartwarming," "delightful"). The inclusion of the bigram "couldnt put" is particularly insightful, as the word "put" alone is neutral, but its combination with "couldnt" creates a powerful, idiomatic expression of engagement. The clear-cut and powerful sentiment polarity of these identified features explains why the model was able to establish perfect decision boundaries between the classes with no overlap or confusion.

Comparison with Previous Research

The findings of this study align with foundational principles in sentiment analysis while also highlighting the critical role of dataset complexity. Much of the previous research in the field, particularly on benchmark datasets like IMDb movie reviews or product reviews, contrasts lexicon-based methods with machine learning approaches. The linguistic markers identified in this study (e.g., "disappointing," "masterpiece") are precisely the types of high-polarity words that form the basis of sentiment lexicons. The model's perfect success demonstrates that when text is composed primarily of such unambiguous terms, a machine learning model can effectively replicate and automate the logic of a lexicon-based system with near-perfect accuracy. However, unlike many studies that report accuracies in the 80-95% range on complex, real-world datasets, our 100% accuracy underscores a key difference: our dataset lacks the common challenges of sarcasm, subtlety, negation handling (e.g., "not a bad book"), and domain-specific jargon that typically constrain model performance in prior research. The results therefore reinforce the established consensus that while machine learning models are powerful, their performance is fundamentally dictated by the quality and complexity of the training data.

Limitations

The primary limitation of this study is the nature of the dataset itself. Its simplicity and homogeneity mean that the findings, particularly the perfect performance metrics, are not generalizable to more complex, real-world scenarios. The reviews are short, lack linguistic diversity, and are highly repetitive, which makes them an ideal training ground for validating a model pipeline but a poor representation of authentic reader feedback found on platforms like Goodreads or Amazon. This lack of complexity means the model was not tested on its ability to handle nuance, context, or mixed sentiments, which are significant challenges in the field of NLP. Furthermore, the vocabulary size was relatively small (147 features after TF-IDF), and the model's robustness against a wider range of linguistic expressions remains unevaluated. It would likely fail on reviews using more sophisticated synonyms (e.g., "prosaic" instead of "boring," or "sublime" instead of "masterpiece") because these words were not present in its training vocabulary. This limitation highlights a trade-off between a clean, easily interpretable result and real-world applicability.

Future Research Suggestions

Building on the successful validation of this methodology, several avenues for future research are recommended. The most critical next step is to apply this established TF-IDF and Logistic Regression pipeline to larger, more diverse, and more challenging datasets. Sourcing reviews from public platforms would introduce the linguistic complexity necessary to truly evaluate the model's robustness and generalizability. It is hypothesized that this would result in a significant drop in accuracy, which would then provide a valuable opportunity for error analysis to identify specific types of language where the model fails. Second, future work could involve comparing the performance of this model with more advanced NLP architectures, such as those based on word embeddings (e.g., Word2Vec, GloVe) or transformer models (e.g., BERT). These models are capable of understanding words in their sequential context, which could prove crucial for deciphering sarcasm and nuance where TF-IDF would fail. Finally, the analysis could be expanded beyond simple sentiment classification to more granular forms, such as aspect-based sentiment analysis. This would involve identifying not just whether a review is positive or negative, but also which specific aspects of the book (e.g., plot, character development, writing style) are being discussed, allowing for a much richer and more detailed understanding of reader feedback.

Conclusion

This study successfully designed and implemented a machine learning pipeline for sentiment analysis of book reviews using a TF-IDF vectorizer and a multi-class Logistic Regression model. The methodology proved to be exceptionally effective on the provided dataset, achieving perfect classification accuracy, precision, recall, and F1-scores across all three sentiment categories. The model effectively learned the distinct, high-polarity linguistic patterns present in the text, successfully identifying key markers such as "disappointing" for negative reviews, "okay" for neutral reviews, and "masterpiece" for positive reviews. The flawless performance serves as a robust validation of the pipeline's technical implementation and its ability to classify sentiment when linguistic cues are unambiguous.

Ultimately, this research underscores the foundational principle that a model's performance is inextricably linked to the complexity of its data. While the pipeline was a success in this controlled environment, the primary limitation remains the simplicity of the dataset, which does not reflect the nuance of real-world literary feedback. The study concludes by recommending that future work should focus on applying this validated methodology to larger, more linguistically diverse datasets to assess its generalizability. Further exploration using more advanced NLP models and a shift towards more granular, aspect-based analysis would provide a more comprehensive understanding of reader sentiment.

Declarations

Author Contributions

Conceptualization: S.A.A.; Methodology: S.A.A.; Software: S.A.A.; Validation: S.A.A.; Formal Analysis: S.A.A.; Investigation: S.A.A.; Resources: S.A.A.; Data Curation: S.A.A.; Writing Original Draft Preparation: S.A.A.; Writing Review and Editing: S.A.A.; Visualization: S.A.A.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] L. G. Singh and S. R. Singh, “Characteristics of Opinions in the Societal and Non-Societal Domains,” 2024, doi: 10.21203/rs.3.rs-3886633/v1.
[2] J. Hruška, “Sentiment Analysis of National Tourism Organizations on Social Media,” 2020, doi: 10.36689/uhk/hed/2020-01-028.
[3] T. A. Rana, Y. Cheah, and S. Letchmunan, “Topic Modeling in Sentiment Analysis: A Systematic Review,” *J. Ict Res. Appl.*, 2016, doi: 10.5614/itbj.ict.res.appl.2016.10.1.6.
[4] L. Yan-xiang and J. Jiao, “A Hybrid Sentiment Analysis Method of Transformer and Capsule Network for Hotel Reviews,” *Acad. J. Comput. Inf. Sci.*, 2021, doi: 10.25236/ajcis.2021.040110.
[5] A. Mishra, D. Kanojia, S. Nagar, K. Dey, and P. Bhattacharyya, “Leveraging Cognitive Features for Sentiment Analysis,” 2016, doi: 10.18653/v1/k16-1016.
[6] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, “Sentiment Analysis of Comment Texts Based on BiLSTM,” *Ieee Access*, 2019, doi: 10.1109/access.2019.2909919.
[7] C. Dhaoui, C. Webster, and L. P. Tan, “Social Media Sentiment Analysis: Lexicon Versus Machine Learning,” *J. Consum. Mark.*, 2017, doi: 10.1108/jcm-03-2017-2141.
[8] D. A. Fatah and E. M. Sari Rochman, “Sentiment Analysis of Madura Tourism Opinion Using Support Vector Machine (SVM),” *Tech. Romanian J. Appl. Sci. Technol.*, 2023, doi: 10.47577/technium.v16i.9988.
[9] S. Qu, “A Thematic Analysis of English and American Literature Works Based on Text Mining and Sentiment Analysis,” *Jes*, 2024, doi: 10.52783/jes.3076.
[10] Y. Y. Chen, N. Wiratunga, and R. Lothian, “Integrating Selection-Based Aspect Sentiment and Preference Knowledge for Social Recommender Systems,” *Online Inf. Rev.*, 2019, doi: 10.1108/oir-02-2017-0066.
[11] A. Bashith, W. S. Adji, and A. Nurdin, “Trend of Public Emotions on Social Media Towards Study at Home Policies,” 2021, doi: 10.2991/assehr.k.210421.058.
[12] Y. Choi, “Finding ‘Just Right’ Books for Children: Analyzing Sentiments in Online Book Reviews,” *Electron. Libr.*, 2019, doi: 10.1108/el-01-2019-0018.
[13] E.-R. Luțan and C. Bădică, “Emotion-Based Literature Books Recommender Systems,” 2023, doi: 10.15439/2023f8647.
[14] R. Li, “Beyond Surfaces and Depths: An SFL Analysis of Fine Gradations of Meaning in Undergraduates’ Writing About Literature,” *Writ. Commun.*, 2024, doi: 10.1177/07410883241286899.
[15] S. J. Lee and R. de Villiers, “Unveiling Emotional Intensity in Online Reviews: Adopting Advanced Machine Learning Techniques,” *Australas. Mark. J. Amj*, 2024, doi: 10.1177/14413582241244808.
[16] C. Zhang, T. Tong, and Y. Bu, “Examining Differences Among Book Reviews From Various Online Platforms,” *Online Inf. Rev.*, 2019, doi:

10.1108/oir-01-2019-0037.
[17] A. Rufaida, A. E. Permanasari, and N. A. Setiawan, “Lexicon-Based Sentiment Analysis Using Inset Dictionary: A Systematic Literature Review,” 2023, doi: 10.4108/eai.5-10-2022.2327474.
[18] S. J. Mahajani, S. Srivastava, and A. F. Smeaton, “A Comparison of Lexicon-Based and ML-Based Sentiment Analysis: Are There Outlier Words?,” 2023, doi: 10.1109/aics60730.2023.10470734.
[19] F. C. Onwuegbuche, J. M. Wafula, and J. K. Mung’atu, “Support Vector Machine for Sentiment Analysis of Nigerian Banks Financial Tweets,” *J. Data Anal. Inf. Process.*, 2019, doi: 10.4236/jdaip.2019.74010.
[20] K. Machová, M. Mach, and M. Vasilko, “Comparison of Machine Learning and Sentiment Analysis in Detection of Suspicious Online Reviewers on Different Type of Data,” *Sensors*, 2021, doi: 10.3390/s22010155.
[21] H. Saif, M. Fernández, L. Kastler, and H. Alani, “Sentiment Lexicon Adaptation With Context and Semantics for the Social Web,” *Semantic Web*, 2017, doi: 10.3233/sw-170265.
[22] S. Sanagar and D. Gupta, “Automated Genre-Based Multi-Domain Sentiment Lexicon Adaptation Using Unlabeled Data,” *J. Intell. Fuzzy Syst.*, 2020, doi: 10.3233/jifs-179704.
[23] D. Kim, “Research on Text Classification Based on Deep Neural Network,” *Int. J. Commun. Netw. Inf. Secur. Ijcnis*, 2022, doi: 10.17762/ijcnis.v14i1s.5618.
[24] X. Li, Z. Li, H. Xie, and Q. Li, “Merging Statistical Feature via Adaptive Gate for Improved Text Classification,” *Proc. Aaai Conf. Artif. Intell.*, 2021, doi: 10.1609/aaai.v35i15.17569.
[25] S. Jiang, R. Angarita, S. Cormier, and F. Rousseaux, “Fine-Tuning BERT-based Models for Plant Health Bulletin Classification,” 2021, doi: 10.48550/arxiv.2102.00838.
[26] M. Golzadeh, A. Decan, E. Constantinou, and T. Mens, “Identifying Bot Activity in GitHub Pull Request and Issue Comments,” 2021, doi: 10.1109/botse52550.2021.00012.
[27] W. Li and B. Mak, “Derivation of Document Vectors From Adaptation of LSTM Language Model,” 2017, doi: 10.18653/v1/e17-2073.
[28] M. J. Basha, S. Vijayakumar, J. Jayashankari, A. H. Alawadi, and P. Durdona, “Advancements in Natural Language Processing for Text Understanding,” *E3s Web Conf.*, 2023, doi: 10.1051/e3sconf/202339904031.
[29] M. E. Peters *et al.*, “Deep Contextualized Word Representations,” 2018, doi: 10.18653/v1/n18-1202.
[30] M. M. Alexandrovna, “Tweet Sentiment Analysis by Python and Classification by Attention Sequence Modeling,” *J. Inf. Syst. Eng. Manag.*, 2025, doi: 10.52783/jisem.v10i22s.3479.
[31] H. Liu and M. Cocea, “Fuzzy Rule Based Systems for Interpretable Sentiment Analysis,” 2017, doi: 10.1109/icaci.2017.7974497.
[32] M. Sana, “Event Impact Analysis Using Machine Learning,” *Interantional J. Sci. Res. Eng. Manag.*, 2023, doi: 10.55041/ijsrem25446.
[33] Ms. U. Joshi, “Real-Time Sentiment Analysis,” *Interantional J. Sci. Res. Eng. Manag.*, 2024, doi: 10.55041/ijsrem28830.
[34] Mr. J. Dias and P. D. Tawde, “Predicting Mental Health Disorders Based on Sentiment Analysis,” *Interantional J. Sci. Res. Eng. Manag.*, 2024, doi: 10.55041/ijsrem37524.