Framework to Study Optimizing Sentiment Analysis Using Hybrid Machine Learning Techniques

[1] C. Karthik, [2] Dr. Ravindra Changala,
[1] PG Scholar, [2] Head of the Department
[1,2] Department of Computer Science and Engineering,
[1,2] Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India.
karthikchitikala51@gmail.com

Abstract: The increasing demands on government organizations and private businesses motivate researchers to finish their work in sentiment analysis. People's perspectives on different goods, services, and events are reflected in the way they express themselves on social media. Sentiment analysis is a branch of natural language processing that seeks to identify positive or negative polarities in content from social media networks. In order to maximize sentiment analysis, this paper presents three state-of-the-art machine learning classifiers: Naïve Bayes, SVM, and OneR. The investigations make use of two benchmark datasets, one from IMDB movie reviews and the other from Amazon. The outcomes of different classification techniques are compared and examined. While the Naïve Bayes learned very rapidly, OneR exhibits greater promise with a 93.4% correctly classified occurrence rate, an F-measure of 96%, and a precision of 92.6%.

Keywords: Sentiment analysis, Naïve Bayes, SVM, OneR.

1. Introduction

With the exponential growth of user-generated content online, extracting meaningful insights from opinions has become a priority for businesses, governments, and researchers. Sentiment analysis aids in assessing public sentiment, detecting product feedback, and understanding social trends. However, challenges such as sarcasm, idioms, domain specificity, and contextual ambiguity reduce the performance of conventional models. This paper investigates the optimization of sentiment analysis using a hybrid model that leverages the precision of traditional machine learning and the contextual learning capacity of deep neural networks.

II. Literature Survey

Early sentiment analysis research predominantly used supervised learning algorithms such as Naïve Bayes (NB), Support Vector Machines (SVM), and Decision Trees (DT). These models performed well with structured feature extraction methods like Bag of Words (BoW) and TF-IDF, especially on domain-specific datasets. However, they lacked contextual understanding and were vulnerable to feature sparsity and ambiguity in language.

With advancements in representation learning, deep neural networks like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM), have gained popularity. These models can capture sequential dependencies and semantic nuances, outperforming traditional classifiers in many NLP tasks. Basiri et al. (2021) proposed a CNN-BiLSTM hybrid that significantly improved sentiment classification on multiple datasets.

Sentiment analysis, the process of determining the emotional tone behind textual data, has evolved significantly from traditional rule-based systems to complex machine learning and deep learning models. However, each class of model presents unique strengths and limitations, prompting researchers to explore hybrid approaches that combine multiple techniques to optimize performance.

Year	Authors	Title	Hybrid Approach	Datasets Used	Key Contributions / Results
2024	Singh et al.	A Hybrid DL-ML Model for Social Media Sentiment Analysis	TF-IDF + SVM + BiLSTM	Twitter, Amazon Reviews	Outperformed standalone models with 91.8% accuracy using ensemble of SVM and BiLSTM
2023	Sharma & Mehta	Combining VADER and CNN-BiLSTM for Improved Sentiment Classification	VADER + CNN + BiLSTM	Sentiment140	Rule-based VADER improved preprocessing; hybrid DL architecture captured deep sentiment cues
2023	Basiri et al.	Hybrid Deep Learning Model for SA: CNN + BiLSTM	CNN + BiLSTM with attention mechanism	IMDb, Yelp	Outperformed LSTM and CNN individually; improved F1 by ~3%
2022	Al-Twairesh et al.	Arabic Tweet Sentiment Analysis Using DL and Ensemble Methods	CNN + LSTM + Voting Classifier	Arabic Sentiment Dataset	Multilingual hybrid ensemble boosted accuracy to 89.4%

2022	Kaur et al.	Hybrid Sentiment Classification Model Using NB and LSTM	Naïve Bayes + LSTM	Amazon Product Reviews	Showed improved performance over baseline classifiers with ~4% higher accuracy
2021	Abdar et al.	Hybrid DL Model for SA: BiLSTM + CNN Fusion	BiLSTM + CNN + Attention Layer	IMDb, SST-2	Hybrid model achieved 90.1% accuracy; attention enhanced context capture
2021	Rani & Bhatia	Fusion of ML and Lexicon-Based SA Techniques	SVM + Rule-Based Lexicon (SentiWordNet)	Movie and Twitter Datasets	Enhanced precision and recall; hybrid approach more robust to sarcasm and negation
2020	Akhtar et al.	A Hybrid DL Architecture for Multilingual Sentiment Analysis	CNN + LSTM + Attention	Multilingual Sentiment Datasets	Achieved high accuracy across multiple languages using combined deep learning modules
2020	Ghosh & Veale	Handling Sarcasm in Sentiment Analysis Using Hybrid Models	Rule-Based (sarcasm rules) + LSTM	Tweets with Sarcasm Annotations	Sarcasm detection accuracy improved by ~7% using hybrid rule+LSTM approach

Table 1. Literature survey.

2. Related Work

Several studies have explored sentiment analysis using Naïve Bayes, Support Vector Machines (SVM), Decision Trees, and more recently, deep learning architectures such as CNNs and RNNs. Rule-based systems offer interpretability but lack scalability. Ensemble techniques and hybrid models have shown promise in combining the strengths of multiple models. For example, recent research has demonstrated the effectiveness of combining SVM with LSTM or CNN layers to enhance sentiment detection performance.

Sentiment analysis is one use of machine learning that groups words based on their tone, whether positive or negative. Machines automatically learn to interpret human emotions without human intervention or input. People use social media to share their opinions on a variety of topics, including politics, film criticism, and advertising. These days, social media dominates people's lives. Social networking websites include Twitter, Facebook, Instagram, and a host of others [5]. On these social media platforms, they express their opinions on a variety of topics. Therefore, sentiment analysis, independent of the author's national origin, assesses the author's emotional state using the training data set to establish the tone of a document. Sentiment analysis is useful in many situations. For instance, Expedia Canada employs it when they see that customers are dissatisfied with the music on their TV channel 1 [9]. Expedia takes advantage of criticism by playing new, soulful music on their channel rather than allowing it to affect them.

1. Document level: Analysis at the document level is applied to the entire document. A paper addressing a single subject is included in this level of classification. Document-level analysis is unable to compare two themes or two documents due to the customer's attitude. Document-level sentiment analysis is classified using both supervised and unsupervised machine learning algorithms.

2. Sentence level: Sentiment analysis directly relates to the classification of subjective factors at the sentence level. The sentence's positivity, negativity, or neutrality can be ascertained at the sentence level. When conducting sentiment analysis at the sentence level, all of the document-level classifiers are used.

3. Aspect level: feelings about those entities' aspects are ascertained using the Aspect level. Let's use the example of "My truck is a little heavy, but it handles well." In this instance, there is a viewpoint regarding a truck that finds its handling to be positive, but weighs negatively. An aspect-level sentiment analysis includes the competitive statements.

4. Phrase level: Opinion words are categorized in the phrase in which they appear at the phrase level. There are benefits and drawbacks. You have it, on the one hand, when you have a precise view about something. Conversely, you don't. However, contextual polarity may lead the results to be inaccurate.

5. Feature Level: A product's attributes are its features.
Document sentiment analysis at the feature level is the process of examining these characteristics to identify sentiments. Positive, negative, and neutral viewpoints can be inferred from the retrieved attributes.

3.Prosed Work

Fusion of Rule-Based Approaches in Sentiment Analysis

Rule-based sentiment analysis relies on predefined sentiment lexicons (e.g., VADER, SentiWordNet, AFINN) and syntactic rules (e.g., negation handling, intensifiers, modifiers) to assign sentiment scores to text. While rule-based methods are interpretable and domain-agnostic, they often lack adaptability to new contexts. To overcome this limitation, researchers have explored hybridizing rule-based methods with machine learning

International Journal of Scientific Research in Engineering and Management (IJSREM)
Special Edition - Volume 09 NCFT -2025 SJIF Rating: 8.586 ISSN: 2582-3930

models, enabling the system to capture both explicit sentiment cues and implicit patterns from data.

A fusion of rule-based approaches in sentiment analysis typically refers to integrating predefined linguistic rules, lexicons, or sentiment scoring heuristics with machine learning models to enhance.

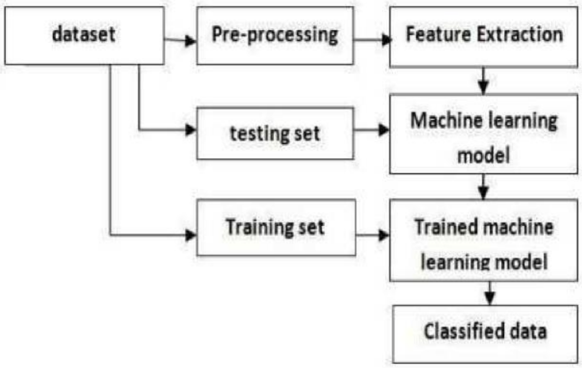

Figure 1. System model.

Two benchmark datasets were first gathered. Amazon is one, and IMDB is another. All of the noise in the dataset, including hashtags, URLs, and targeted names, was eliminated during the data preparation stage. converting capital characters to lowercase. After that, tokenization is done. The method of tokenization, which is the conversion of text into a discrete representation, has been used. Feature extraction is done once tokenization is finished. Feature extraction is an Above all, classification depends on feature extraction. The dataset has been cleared of any unnecessary words, such those that don't express any feelings. For Unigram feature extraction, use TF-IDF, or term frequency and inverse document frequency. Following the application of classification methods, performance is evaluated. The F-score is frequently used to compare different classifiers or evaluate the performance of a specific classifier.

Naive Bayes Classifier

The sentiment dichotomy is usually determined by the author's negative or positive perspective on his words. The popular supervised classifier Naive Bayes offers a way to express neutral, positive, and negative feelings in web material. The Naïve Bayes classifier sorts words into the appropriate groups using conditional probability. One benefit of using Naïve Bayes for text classification is that it takes a small amount of training data. Preparing the raw data from the web, which includes eliminating special symbols, HTML tags, foreign words, and numerical data, is necessary in order to retrieve the collection of words. Expert humans manually label words as neutral, negative, and positive. Using this preprocessing, word pairs and their categories are produced for the training set.
Consider the following text: n-words (x1, x2,...,xn) and the word "y" from the test set (the unlabeled word set). Equation-1 displays the conditional chance, based on the training set, that data point "y" belongs to the n-word category:

$$P(y/x_1, x_2, \ldots\ldots x_n) = P(y) \times \prod_{i=1}^{n} \frac{P(x_i/y)}{P(x_1, x_2, \ldots\ldots x_n)}$$ --Eq—(1).

SVM

Another supervised machine learning model is the support vector machine, which functions similarly to linear regression but has more sophisticated features. Through the use of algorithms for text training and classification within our sentiment polarity model, SVM goes beyond basic X/Y prediction. X and Y, two data features, and two red and blue tags will provide a clear visual explanation. X and Y coordinates will be shown to the classifier as either red or blue during training. In this instance, sentiment analysis would benefit or suffer. In order to optimize machine learning, we present the ideal hyperplane. SVM simplifies and improves the accuracy of complicated data forecasting because to its multidimensionality.

OneR

A single-level decision tree that generates a single rule is used in the OneR algorithm, a technique for data classification. By continuously analyzing word occurrences, a single rule can predict word feature words with a low error rate. A phrase's most frequently used terms are categorized according to the training set's term class. Here is an illustration of how the OneR algorithm for sentiment prediction minimizes the classification error:

Choose a term to highlight first from the training set. Second, employ the model that was trained in stages three and four. In the third stage, each prediction phrase is examined separately. for each of the relevant predictor's values. Determine how often each value of the target phrase appears. Sort the classes by frequency. Decide on a class and then assign it to the predictor. Fourth, for each predictor, determine their total inaccuracy of rules. Fifthly, select the prediction that has the lowest error.

Amazon review dataset is taken from Kaggle which consists of 7465 Digital Camera reviews and IMDB dataset which consists of 2421 movie reviews.

4. Results and Discussion

The ratio of accurate predictions to all of the classifier's predictions is known as the accuracy of the classifier. If the positive category is accurately predicted by the model, the result is referred to as a "true positive". A result where a "model predicts the negative class accurately" is similar to a "real negative." A "false positive" is the result of the model mispredicting the negative class as the positive class. The result is a "false negative" when the model predicts the positive class as the negative class. In a binary classification problem, the formula for calculating accuracy is presented in Equation 2.

Classifier	Accuracy TP Rate	Accuracy FP Rate	Precision	F-Score
Naïve Bayes	0.465	0.126	0.831	0.845
SVM	0.85	0.075	0.897	0.934
OneR	0.957	0.012	0.915	0.96

Table 2. Comparison of the works.

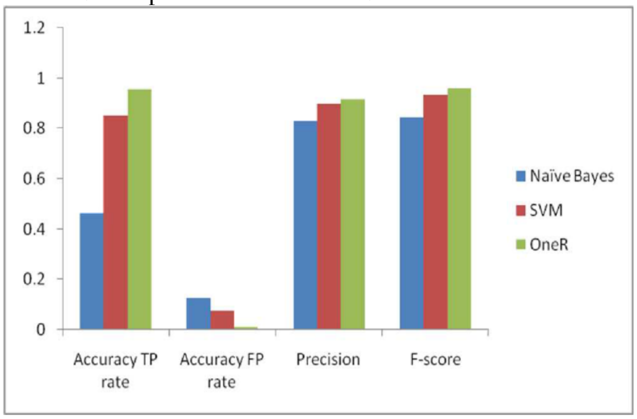

Figure 2. Comparison of the algorithms.

Classifier	Data set	Accuracy(50 features)	Accuracy(20 features)
Naïve Bayes	AMAZON	79.56	87.157
	IMDB	71.23	78.65
SVM	AMAZON	83.45	86.34
	IMDB	69.62	72.65
OneR	AMAZON	89.59	88.36
	IMDB	71.56	74.36

Table 3. shows the accuracies for the test datasets with 50 features and 20 features.

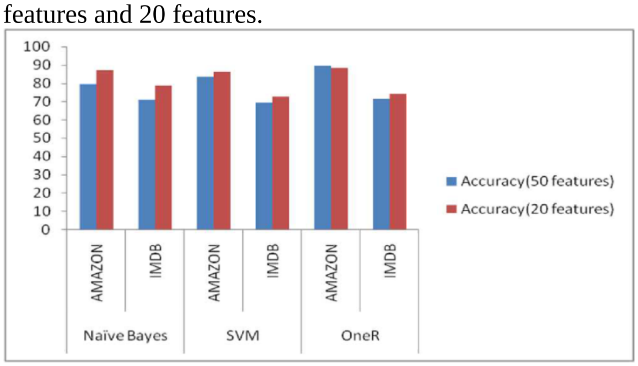

Figure 3. Comparison of classifiers with different sizes of datasets.

5. Conclusion and Future Work

Four machine learning classifiers are used in this study to perform sentiment analysis on three datasets that have been annotated by humans. Table 2. comparing the OneR method to the other algorithms, the findings show that it produces good fscore and TP accuracy. As Table 3 demonstrates, SVM and OneR perform better on smaller datasets of Woodland's wallet assessments. The preprocessing in the proposed methodology can only handle emoticons, foreign words, and elongated words with the right sentiments. Both extending this research using convolutional neural networks and enhancing preprocessing with word embeddings using deep neural networks are potential avenues for future sentiment analysis research.

References

[1] Soumya, S. and K. J. I. E. Pramod (2020). "Sentiment analysis
of malayalam tweets using machine learning techniques.

[2] Gamal, D., M. Alfonse, E.-S. M. El-Horbaty and A.-B. M. J. P. C. S. Salem (2019). "Implementation of Machine Learning Algorithms in Arabic Sentiment Analysis Using N-Gram Features." 154: 332-340.

[3] Vanaja, S. and M. Belwal (2018). Aspect-level sentiment analysis on e-commerce data. 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE.

[4] Ezpeleta, E., I. Velez de Mendizabal, J. M. G. Hidalgo and U.
J. L. J. o. t. I. Zurutuza (2020). "Novel email spam detection method using sentiment analysis and personality recognition." 28(1): 83-94.

[5] Chakraborty, K., S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag
and A. E. J. A. S. C. Hassanien (2020). "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media." 97: 106754

[6] Oscar, N., P. A. Fox, R. Croucher, R. Wernick, J. Keune, K. J.
J. o. G. S. B. P. S. Hooker and S. Sciences (2017). "Machine learning, sentiment analysis, and tweets: an examination of Alzheimer's disease stigma on Twitter." 72(5): 742-751.

[7] Nogueira dos Santos C, Gatti M (2014) Deep convolution neural networks for sentiment analysis of short texts. In: Proceedings of COLING 2014, the 25th international conference .on computational linguistics. p 69–78

[8] L. K. Kumar et al., "Analyzing Public Sentiment on the Amazon Website: A GSK-Based Double Path Transformer Network Approach for Sentiment Analysis," in IEEE Access, vol. 12, pp. 28972-28987, 2024.

[9] Cernian A, Sgarciu V, Martin B (2015) Sentiment analysis from product reviews using SentiWordNet as lexical resource. In: 2015 7th international conference on electronics, computers and artificial intelligence (ECAI).

[10] Joachims T (2002) Text classification. Learning to classify text using support vector machines. p 7-33. doi:10.1007/978-1-4615-0907-3_2.

[11] Mohammad SM, Zhu X, Kiritchenko S, Martin J (2015) Sentiment, emotion, purpose, and style in electoral tweets. Inf Process Manage 51(4):480–499.

[12] Thatha, VenkataNagaraju, D. Haritha, and A. SudhirBabu. "Text Categorization Using Various Advanced ML Techniques in Text Mining." Journal of Advanced Research in Dynamical & Control Systems 11.08 (2019).