International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Comparative Study of Machine Learning Models for Sentiment Analysis

Jadeja Rahul[1], Asst. Prof. Ms. Manisha Vasava[2]

[1]*Research Scholar, Institute of Information Technology, Krishna School Of Emerging Technology & Applied Research,* KPGU University, *Varnama, Vadodara, Gujarat, India*

[2]*Assistant Professor, Department of Information Technology and Engineering, Krishna School Of Emerging Technology & Applied Research,* KPGU University, *Varnama, Vadodara, Gujarat, India*

---***---

Abstract - Sentiment analysis plays a crucial role in understanding public opinions on social media platforms like Twitter. This study performs a comparative analysis of three machine learning models—Bernoulli Naïve Bayes (BernoulliNB), Linear Support Vector Classifier (LinearSVC), and Linear Regression—for binary sentiment classification (positive vs. negative) using the Sentiment140 dataset. The dataset, which contains pre-labelled tweets, is preprocessed through text normalization, tokenization, and feature extraction using the TF-IDF vectorization method.
The models are evaluated based on accuracy, precision, recall, and F1-score, along with confusion matrices and word clouds for qualitative insights. Our results indicate that model performance varies significantly depending on the feature representation and classification approach. The study highlights the strengths and weaknesses of each model and provides insights into their effectiveness for large-scale Twitter sentiment analysis. This comparative evaluation aims to guide future research in selecting appropriate machine learning models for real-world sentiment analysis applications.

Key Words: Sentiment Analysis, Twitter, Machine Learning, BernoulliNB, LinearSVC, Linear Regression, Sentiment140

1.INTRODUCTION

Social media platforms like Twitter have become powerful channels for expressing public opinion on various topics, including politics, brand perception, and social movements. Analyzing these opinions can provide valuable insights for businesses, policymakers, and researchers. Sentiment Analysis, also known as Opinion Mining, is a branch of Natural Language Processing (NLP) that focuses on determining the emotional tone behind a text—whether it is positive, negative, or neutral.

Despite the advancements in sentiment analysis, accurately classifying tweets remains a challenge due to the short length of text, use of slang and emojis, and the presence of sarcasm. Traditional sentiment analysis techniques often struggle with these complexities. While deep learning models have shown great promise, machine learning models remain widely used due to their interpretability, efficiency, and lower computational requirements.

This study aims to:

1. Compare three machine learning models—Bernoulli Naïve Bayes (BernoulliNB), Linear Support Vector Classifier (LinearSVC), and Linear Regression—for Twitter sentiment classification.
2. Evaluate these models using the Sentiment140 dataset, which contains tweets labelled as positive or negative.
3. Analyze model performance using accuracy, precision, recall, and F1-score, along with visual representations such as word clouds and confusion matrices.

2. Literature Summary

Sr No	Paper Author/ Title/Year Of Publication	Method	Data set	Limitations	Future Scope
1	K. Dave, S. Lawrence, and D.M. Pennock (2003). *Mining the peanut gallery: Opinion extraction and semantic classification of product reviews*	Opinion extraction and semantic classification using machine learning	Product reviews dataset from e-commerce platforms	Difficulty in detecting sarcasm and implicit sentiment; limited to product reviews only	Extending techniques to other domains like social media and news sentiment analysis
2	Nasukawa, Tetsuya, and Jeonghee Yi. (2003) *Sentiment Analysis: Capturing Favorability Using Natural Language Processing*	Natural Language Processing (NLP) with syntactic and semantic analysis	Customer feedback and reviews (proprietary/internal dataset)	Limited scalability and difficulty in handling complex linguistic nuances like sarcasm	Integration with machine learning for better adaptability and domain transfer
3	V. Hatzivassiloglou and K. McKeown (2004). *Predicting the semantic orientation of adjectives*	Lexicon-based approach for determining adjective polarity using co-occurrence patterns	Text corpus with labelled adjectives	Focuses only on adjectives, limiting overall sentiment analysis performance	Expanding the approach to verbs and nouns; integrating with modern NLP techniques for context-aware sentiment analysis
4	Pang B, Lee L (2008). *Opinion mining and sentiment analysis*	Machine learning techniques for sentiment classification, including Naïve Bayes, Maximum Entropy, and SVM	Movie reviews dataset and other text datasets	Limited scope due to dataset bias; difficulty in handling sarcasm and implicit sentiment	Expansion to other domains like social media; using deep learning for improved sentiment detection
5	Go A, Bhayani R, Huang L (2009). *Twitter sentiment classification using distant supervision*	Distant supervision with machine learning (Naïve Bayes, MaxEnt, and SVM) using emoticons as weak labels	Large-scale Twitter dataset (automatically labelled using emoticons)	Noisy labels due to automatic annotation; lacks deep contextual understanding	Exploring neural networks and deep learning to improve sentiment classification accuracy

6	Songbo Tan, Xueqi Cheng, Yuefen Wang, Hongbo Xu. (2009) *Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis*	Domain Adaptation for Sentiment Analysis using Naïve Bayes with an improved adaptation algorithm	Amazon product reviews and other domain-specific datasets	Struggles with highly imbalanced datasets; limited to text-based features.	Improving adaptation across more diverse domains; integrating deep learning techniques for better generalization.
7	Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R. (2011).*Sentiment analysis of Twitter*	Hybrid approach combining machine learning classifiers and feature engineering	Twitter dataset (collected tweets, labelled for sentiment)	Feature selection affects accuracy; handling negations and sarcasm remains a challenge	Exploring deep learning models; improving feature representation techniques for better accuracy
8	Alena Neviarouskaya, Helmut Prendinger, Mitsuru Ishizuka. (2011) *Secure SentiFul: A Lexicon for Sentiment Analysis*	Lexicon-based sentiment analysis using an enhanced lexicon called *SentiFul*	Manually created sentiment lexicon	Limited ability to handle context-specific sentiment variations; struggles with sarcasm and complex sentence structures.	Expansion of the lexicon with contextual understanding; integration with machine learning models for improved accuracy
9	A. Kumar and T. M. Sebastian (2012). *Machine learning assisted Sentiment Analysis*	Machine learning-based sentiment classification with automated feature selection	Various text datasets (product reviews, blogs, news articles)	Performance highly dependent on feature selection; difficulty handling mixed emotions in text	Enhancing feature engineering techniques using deep learning; improving sentiment classification in multilingual datasets
10	Bing Liu, Lei Zhang. (2012) *A Survey of Opinion Mining and Sentiment Analysis*	Survey of sentiment analysis techniques including machine learning, lexicon-based, and hybrid approaches	Not specific to any one dataset (review paper)	Lacks experimental implementation; mainly theoretical overview.	Applying reviewed methods on modern datasets, integrating deep learning, and addressing multilingual sentiment analysis.

11	Neethu, M. S., & Rajasree, R. (2013). *Sentiment analysis in Twitter using machine learning techniques*	Machine learning-based sentiment classification using feature extraction (e.g., bag-of-words) and classifiers (e.g., SVM, Naïve Bayes)	Tweets collected from Twitter (sample dataset, manually curated)	Limited dataset size; challenges with noisy, informal language and slang; limited handling of sarcasm and context-specific sentiments	Expanding dataset size; incorporating more advanced NLP and deep learning techniques; real-time sentiment analysis; improved handling of informal language and context
12	Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *International Conference on Information Technology and Quantitative Management*.	Focus on text preprocessing techniques (Stopword removal, stemming, tokenization)	IMDb Movie Reviews Dataset	Limited evaluation on diverse datasets; primarily tested on movie reviews	Applying preprocessing techniques on real-time datasets like Twitter for improved accuracy
13	Flekova, Lucie, Oliver Ferschke, and Iryna Gurevych. (2014) *Ukpdipf: Lexical Semantic Approach to Sentiment Polarity Prediction in Twitter Data*	Lexical Semantic Approach	SemEval 2014 Twitter Dataset	Struggles with domain-specific sentiment variations and context-dependent meanings	Improvement using deep learning and word embeddings for better sentiment context understanding
14	A. Sarlan, C. Nadam, S. Basri. (2014) *Twitter Sentiment Analysis*	Natural Language Processing (NLP) and Data Mining	Twitter Dataset	Struggles with sarcasm, negations, and context-based sentiment interpretation	Enhancing sentiment detection with deep learning and contextual understanding
15	Geetika Gautam, Divakar Yadav. (2014) *Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis*	Machine Learning (Naïve Bayes, SVM) combined with semantic analysis	Twitter Dataset	Limited accuracy due to informal language and short text; struggles with sarcasm and context-based sentiment.	Integration of deep learning techniques like LSTMs and BERT; improving semantic understanding for better accuracy

16	Joshi R, Tekchandani R (2016). *Comparative analysis of Twitter data using supervised classifiers*	Supervised machine learning classifiers (e.g., Naïve Bayes, SVM, Decision Tree) for sentiment analysis	Twitter dataset (collected tweets, labelled manually)	Performance varies based on the classifier; difficulty in handling sarcasm, slang, and short-text sentiment ambiguity	Exploring deep learning techniques; handling sarcasm and contextual meaning better; real-time sentiment analysis
17	J. Ramteke, S. Shah, D. Godhia, A. Shaikh (2016). *Election result prediction using Twitter sentiment analysis*	Sentiment analysis using Vader lexicon and text classification algorithms	Twitter dataset related to election discussions	Challenge in accurately mapping sentiment to real-world election results; biased dataset due to bot activity	Enhancing predictive accuracy by incorporating more diverse datasets; integrating deep learning-based sentiment models
18	Devika, M. Devi, C. Sunitha, and Amal Ganesh. (2016) *Sentiment Analysis: A Comparative Study on Different Approaches*	Comparison of machine learning and lexicon-based approaches for sentiment analysis	Various benchmark datasets for sentiment classification	Inconsistent results across different datasets and domains	Combining machine learning with lexicon-based methods for improved accuracy
19	Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, *2*(1), 325-347.	Linguistic-based sentiment analysis approach, focusing on lexicons, discourse structure, and modifiers	Various linguistic datasets reviewed	Challenges in capturing sentiment nuances like sarcasm and context-dependent meanings	Integration of linguistic insights with machine learning for improved sentiment analysis accuracy
20	Trupthi M, Pabboju S, Narasimha G (2017). *Sentiment analysis on Twitter using Streaming API*	Real-time sentiment analysis using Twitter Streaming API and machine learning techniques	Live Twitter data stream	Challenges in real-time processing and filtering relevant tweets; difficulty handling multilingual data	Enhancing real-time processing capabilities with big data technologies; integrating deep learning for better accuracy

21	Asha S Manek, P Deepa Shenoy, M Chandra Mohan, Venugopal K R (2017). *Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier*	Aspect-based sentiment analysis using Gini Index for feature selection and SVM for classification	Large-scale movie review dataset	Computational complexity in handling large datasets; limitations in capturing deep semantic meaning	Exploring deep learning for improved aspect-based sentiment analysis; optimizing feature selection for scalability
22	R. D. Desai (2018). *Sentiment Analysis of Twitter Data*	Machine learning-based sentiment analysis using classifiers (e.g., Naïve Bayes, SVM)	Twitter dataset (collected tweets, manually labelled)	Limited dataset; challenges in detecting sarcasm and emojis' sentiment impact	Improving dataset diversity; exploring deep learning models for higher accuracy
23	Shamantha RB, Shetty SM, Rai P (2019). *Sentiment analysis using machine learning classifiers: evaluation of performance*	Evaluation of various machine learning classifiers (e.g., Naïve Bayes, SVM, Random Forest) on sentiment analysis	Twitter dataset	Variability in classifier performance; issues in sentiment detection for ambiguous or sarcastic tweets	Implementing ensemble learning and deep learning models for better accuracy
24	Bonta V, Kumaresh N, Janardhan N (2019). *A comprehensive study on Lexicon based approaches for sentiment analysis*	Lexicon-based sentiment analysis approaches (e.g., SentiWordNet, Vader, AFINN)	Various text corpora	Limited accuracy for complex sentences and sarcasm; requires domain-specific lexicons for better results	Combining lexicon-based methods with deep learning for better sentiment analysis accuracy
25	Yadav N, Kudale O, Gupta S, Rao A, Shitole A (2020). *Twitter sentiment analysis using machine learning for product evaluation*	Machine learning techniques (Naïve Bayes, SVM, Random Forest) for sentiment classification of product reviews	Twitter dataset containing product-related tweets	Challenges in dealing with fake reviews and spam; difficulty in handling sarcasm in short text	Improving fake review detection; using deep learning and transformer models for enhanced sentiment analysis accuracy

3. Methodology

3.1 Dataset: Sentiment140

The Sentiment140 dataset is widely used for training sentiment analysis models. It consists of 1.6 million tweets labelled as either positive (label: 4) or negative (label: 0). Each entry in the dataset includes:

- Tweet text
- Sentiment label
- Tweet ID
- Date of the tweet
- Username

For this study, only the text and sentiment label columns were used. To simplify binary classification, labels were mapped as follows:

- Positive → 1
- Negative → 0

3.2 Data Preprocessing

• Lowercasing: All text was converted to lowercase.

• Removing usernames, URLs, and special characters: Regex techniques were applied to eliminate noise.

• Stopword removal: Common stopwords like "the," "is," "at" were removed.

• Tokenization: Tweets were split into words for vectorization.

• Vectorization: We used TF-IDF (Term Frequency–Inverse Document Frequency) to convert text into numerical form.

3.3 Machine Learning Models

3.3.1 Bernoulli Naïve Bayes (BernoulliNB)

This probabilistic model is ideal for binary/boolean feature vectors, making it suitable for sentiment classification where presence/absence of words matters.

3.3.2 Linear Support Vector Classifier (LinearSVC)

A linear form of Support Vector Machines that performs well on text classification tasks. It aims to find a hyperplane that maximally separates the two sentiment classes.

3.3.3 Linear Regression

Although typically used for continuous prediction, Linear Regression can be adapted for binary classification by thresholding its output. It serves here as a baseline comparison.

3.4 Model Training and Evaluation

• Train-Test Split: Dataset was split into 95% training and 5% testing sets.

• Evaluation Metrics:

• Accuracy

• Precision

• Recall

• F1-score

4. Results and Discussion

4.1 Bernoulli Naïve Bayes (BernoulliNB)

	precision	recall	f1-score	support
0	0.81	0.79	0.80	39989
1	0.80	0.81	0.80	40011
accuracy			0.80	80000
macro avg	0.80	0.80	0.80	80000
weighted avg	0.80	0.80	0.80	80000

The Bernoulli Naïve Bayes classifier achieved 80% accuracy, with balanced precision and recall for both sentiment classes. This symmetry reflects the model's effectiveness in handling class distributions equally well. It particularly benefits from its assumption of binary features, which simplifies the decision-making process in short texts like tweets. However, its reliance on word presence rather than frequency may limit its depth in capturing nuanced expressions. It performed best when sentiment was expressed clearly, and struggled slightly with ambiguous tweets.

4.2 Linear Support Vector Classifier (LinearSVC)

	precision	recall	f1-score	support
0	0.82	0.82	0.82	39989
1	0.81	0.83	0.82	40011
Accuracy			0.82	80000
macro avg	0.82	0.82	0.82	80000
weighted avg	0.82	0.82	0.82	80000

The LinearSVC model achieved an accuracy of 82%, outperforming the Bernoulli Naïve Bayes model in every metric. Notably, it provided a well-balanced classification for both positive and negative sentiments, with slightly higher recall for positive tweets (0.83) suggesting a better ability to detect positive sentiment. LinearSVC is known to perform exceptionally well on high-dimensional and sparse datasets like text due to its margin-maximization property. This result reinforces its suitability for sentiment analysis on Twitter data.

4.3 Linear Regression

	precision	recall	f1-score	support
0	0.83	0.82	0.83	39989
1	0.82	0.84	0.83	40011
Accuracy			0.83	80000
macro avg	0.83	0.83	0.83	80000
weighted avg	0.83	0.83	0.83	80000

Surprisingly, the Linear Regression model produced the highest accuracy at 83%, slightly surpassing both LinearSVC and BernoulliNB. While regression is typically not ideal for classification, the post-processing (e.g., thresholding continuous outputs into binary classes) appears to have effectively handled the task. It exhibited excellent balance between precision and recall across both sentiment classes, indicating strong predictive ability and generalization on the dataset.

4.4 Comparative Performance Summary

Model	Accuracy	Precision	Recall	F1-Score
BernoulliNB	0.80	0.80	0.80	0.80
LinearSVC	0.82	0.82	0.82	0.82
Linear Regression	0.83	0.83	0.83	0.83

All three models performed well, with Linear Regression surprisingly leading across all key metrics. LinearSVC came close behind and remains a more theoretically appropriate model for classification tasks, particularly for high-dimensional data. BernoulliNB, though slightly trailing, remains lightweight and fast — ideal for deployment in low-resource settings or real-time streaming environments.

These results emphasize that while model choice is important, preprocessing, feature engineering, and proper evaluation also play critical roles in performance outcomes.

5. CONCLUSIONS

Among the three, Linear Regression achieved the highest overall performance, with an accuracy of 82.8%, followed closely by LinearSVC at 81.9%, and BernoulliNB at 80.1%. The small but consistent improvement in precision and recall across the models indicates that while simpler models like BernoulliNB are computationally efficient, more complex models like LinearSVC and Linear Regression can better capture nuanced patterns in textual sentiment.

In summary, while all three models demonstrated competent performance, the results indicate that linear regression offers the best trade-off between interpretability and predictive accuracy for binary sentiment classification in Twitter data. Future research may involve incorporating deep learning methods or contextual embeddings such as BERT to further improve performance and capture the subtleties of informal language and sarcasm commonly found on social media.

REFERENCES

1 Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews.

2 Tetsuya Nasukawa and Jeonghee Yi. (2003). Sentiment analysis: capturing favorability using natural language processing.

3 V. Hatzivassiloglou and K. McKeown (2004). Predicting the semantic orientation of adjectives.

4 Pang B, Lee L (2008). Opinion mining and sentiment analysis.

5 Go A, Bhayani R, Huang L (2009). Twitter sentiment classification using distant supervision.

6 Songbo Tan, Xueqi Cheng, Yuefen Wang, Hongbo Xu. (2009). Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis.

7 Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R. (2011).Sentiment analysis of Twitter.

8 Alena Neviarouskaya, Helmut Prendinger, Mitsuru Ishizuka. (2011). Secure SentiFul: A Lexicon for Sentiment Analysis.

9 A. Kumar and T. M. Sebastian (2012). Machine learning assisted Sentiment Analysis

10 Bing Liu, Lei Zhang. (2012) A Survey of Opinion Mining and Sentiment Analysis.

11 Neethu, M. S., & Rajasree, R. (2013). Sentiment analysis in Twitter using machine learning techniques.

12 Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. International Conference on Information Technology and Quantitative Management.

13 Flekova, Lucie, Oliver Ferschke, and Iryna Gurevych. (SemEval 2014) Ukpdipf: Lexical Semantic Approach to Sentiment Polarity Prediction in Twitter Data.

14 A. Sarlan, C. Nadam, S. Basri. (IEEE 2014) Twitter Sentiment Analysis.

15 Geetika Gautam, Divakar Yadav. (IEEE 2014) Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis.

16 Joshi R, Tekchandani R (2016). Comparative analysis of Twitter data using supervised classifiers.

17 J. Ramteke, S. Shah, D. Godhia, A. Shaikh (2016). Election result prediction using Twitter sentiment analysis.

18 Devika, M. Devi, C. Sunitha, and Amal Ganesh. (2016) Sentiment Analysis: A Comparative Study on Different Approaches.

19 Taboada, M. (2016). Sentiment analysis: An overview from linguistics. Annual Review of Linguistics, 2(1), 325-347.

20 Trupthi M, Pabboju S, Narasimha G (2017). Sentiment analysis on Twitter using Streaming API.

21 Asha S Manek, P Deepa Shenoy, M Chandra Mohan, Venugopal K R (2017). Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier.

22 R. D. Desai (2018). Sentiment Analysis of Twitter Data.

23 Shamantha RB, Shetty SM, Rai P (2019). Sentiment analysis using machine learning classifiers: evaluation of performance.

24 Bonta V, Kumaresh N, Janardhan N (2019). A comprehensive study on Lexicon based approaches for sentiment analysis.

25 Yadav N, Kudale O, Gupta S, Rao A, Shitole A (2020). Twitter sentiment analysis using machine learning for product evaluation.