WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Deep Learning for Sentiment Analysis of Website Users' Feedback

Kartik Pandey[1], Sagar Choudhary[2], Raj Neelam Gaurav[3]

[1, 3] B.Tech Student, Department of CSE, Quantum University, Roorkee, India.

[2]Assistant Professor, Department of CSE, Quantum University, Roorkee, India.

Abstract

In the last few years, sentiment analysis has picked up good momentum in many areas, such as business, social media, and education. One of the new application areas is analyzing user reviews of websites, where it is important to know the sentiment of users in order to enhance user experience, interface design, and content presentation. But due to informal text, variant style of expressions, and quantity of data, processing and interpreting this feedback is a challenge. This paper discusses the use of deep learning (DL) algorithms in analyzing the feedback of website users through sentiment analysis. In contrast to conventional machine learning (ML), DL techniques have the capability to extract intricate features automatically and comprehend contextual meaning and hence are more appropriate for dealing with the subtleties of human language. Whereas previous research has touched on sentiment analysis in several contexts, there is a gap in directed studies that categorically analyze the employment of DL and NLP methods expressly for user feedback on websites systematically. In this work, we offer a systematic review and synthesis of recent research between 2015 and 2024 under the guidance of the PRISMA approach. We narrowed down from a preliminary list of more than 600 studies to identify 85 as being most pertinent to sentiment analysis in the domain of website feedback. These findings emphasize increasing utilization of DL models, especially transformers and recurrent neural networks, and identify the need for curated data, consistency in sentiment labels, and attention to emotional subtlety in user input. This review is intended to act as a starting point for future research and system development in this growing critical field.

Keywords— Sentiment Analysis, Deep Learning, LSTM, CNN, User Feedback, Natural Language Processing

I. Introduction

As digital communication and e-commerce exploded, user feedback through reviews, comments, and recommendations have become an essential resource for companies looking to enhance services and customer satisfaction. The magnitude and diversity of this feedback, however, pose major challenges to deriving useful insights. Sentiment analysis—automatically identifying emotional tones within text—has come to emerge as a potent solution to this challenge[1]. However, conventional approaches have difficulty in addressing the subtleties of human language, thus falling short on accuracy and contextual awareness.[2]

*A. **Problem Statement***

Traditional sentiment analysis methods, depending on simple machine learning algorithms and hand-crafted features, cannot understand sophisticated language patterns like sarcasm, idioms, or contextual polarity[3]. These models are also heavy to preprocess and could underperform with informal or unstructured text, which is a typical case of real-world user reviews. The requirement of high accuracy and real-time processing, particularly in high-scale systems, calls for robust and computationally efficient models.[4]

*B. **Objectives and Contributions***

The research seeks to investigate the use of deep learning methods for sentiment analysis of user feedback on websites with the following main objectives:

1. Model Optimization

To design and test deep learning models like LSTM, CNN, and transformer-based models (e.g., BERT) that can learn the deeper semantic structure of user feedback.[(5)]

2. Efficiency and Performance Evaluation:

To evaluate the performance of deep learning models relative to conventional methodologies based on standard measures like accuracy, precision, recall, and F1-score, so that accurate sentiment identification is guaranteed in varied text structures.[6]

3. Scalability and Accessibility

To construct models that are scalable, support real-time sentiment analysis, & are usable in multilingual settings—thus being practical for businesses with heterogeneous user groups & low computation capability.[(7)]

This research contributes to the field by introducing a deep learning-powered framework that provides a balance between performance, efficiency, & usability. It forms the foundation of intelligent feedback analysis systems that not only enable better customer engagement strategies but also allow businesses to take timely & informed decisions[8]

II. Literature Review and Related Work

There has been growing interest in recent years in using deep learning techniques for sentiment analysis of user-generated content, particularly online reviews and feedback. Past research in sentiment analysis relied heavily on manual feature engineering and rule-based systems, which, while successful in some cases, had a tendency to create clunky systems that were time-consuming and difficult to scale [9]. With the tremendous increase in content online, such methods struggled to keep pace with the nuances of language employed by individuals in writing opinions on the internet.

A. Deep Learning-Based Classification/Learning Algorithms

In this stage, classification was conducted using recurrent neural network-based LSTM models. Three different models — named Model 1, Model 2, and Model 3 — were constructed by differing their network architecture and hyperparameters. For training purposes, a corpus of 1,194,704 reviews was used, whereas another collection of 512,016 reviews was kept for testing and assessing model performance. The data were divided into two non-overlapping sets: one for training and the other for testing. All classification algorithms were tested twice: once on the training set to tune learning and finally on the test set to estimate generalization ability. Before passing reviews into any classification system, the text data was represented numerically by employing a word embedding method. The LSTM-based models were then trained on these feature vectors. Lastly, the trained models were used on the unseen test data, and their predictions were matched against the true labels to measure classification accuracy. A general overview of the whole sentiment classification process is depicted in Figure 1.

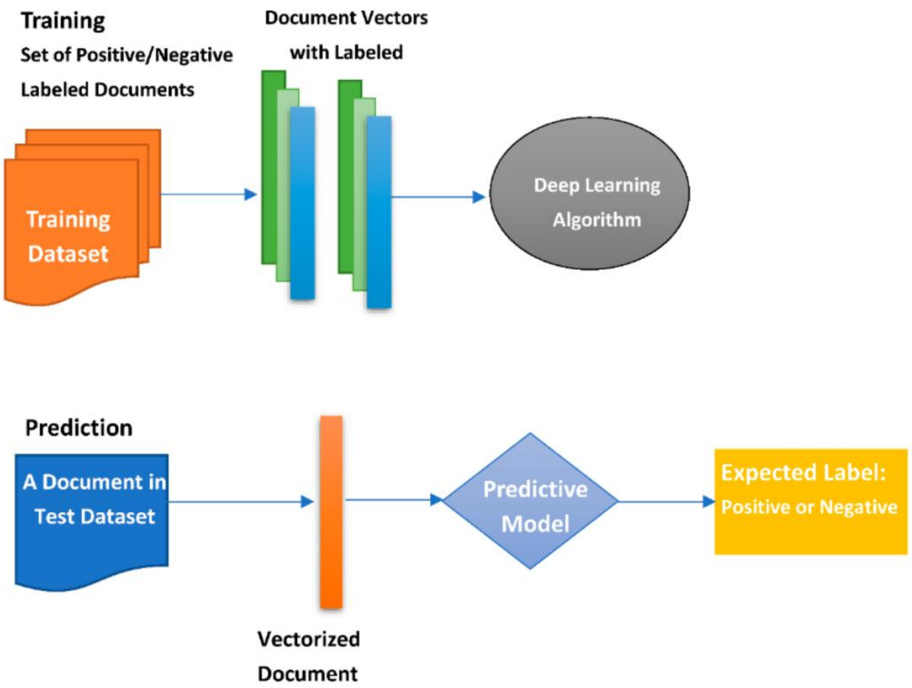

Figure 1. Example of the sentiment classification by supervised deep learning algorithms.

B. Traditional Approaches to Sentiment Analysis

Before deep learning, sentiment classification was mainly achieved using classical machine learning models like Naive Bayes, SVMs, and logistic regression. All of these models required well-crafted features—n-grams, part-of-speech tags, and sentiment lexicons—in order to know what the text says. Although these approaches gave decent results, they used to be brittle and misclassifying, especially when sarcasm, informal words, or specialized vocabulary were involved [10].

Second, previous models had to undergo high preprocessing and weren't able to generalize well to different platforms or domains. That deficiency made it challenging for business organizations to make use of the same sentiment analysis pipeline for diverse sets of consumer reviews, especially when dealing with real-time material.

C. Emergence of Deep Learning in Sentiment Analysis

The shift to deep learning has totally revamped the game when it comes to sentiment analysis. Architectures like CNNs, RNNs, LSTMs, and recently Transformer-based models like BERT have accomplished exemplary performance by automatically learning semantic patterns from raw text [11].

These models can learn the contextual meaning of words in a way that traditional methods could not Most of the recent studies have demonstrated that deep learning not only improves classification results but also reduces the need for manual feature extraction, hence making sentiment analysis pipelines faster and more scalable [cite: Zhang2018DeepSentiment]. Nevertheless, there still remains a challenge of finding a balance between model performance and computational expenses. Large-scale deep learning models have the tendency to call for extensive processing power, and that becomes limiting when applied by small companies or used for real-time analysis purposes. Our research addresses this limitation by comparing and exploring various deep learning models to analyze their performance, accuracy, and suitability for practical applications. We aim to bridge the gap between research performed on the bleeding edge and highly scalable, practical tools available to analyze sentiments.

III. Methodology

a. Programming Environment

All experiments in the present study were performed with Python 3.10, selected for its extensive usage and mature ecosystem in data science and machine learning tasks. TensorFlow and Keras were the most important libraries used for the construction and training of deep neural networks, providing ease of use and scalability for experimentation [12]. Scikit-learn was utilized for partitioning datasets, preprocessing, and calculating metrics [13]. Numerical computation was managed using NumPy, and Matplotlib and Seaborn were employed for plotting model performance[14]

b. Dataset Description

We used a multi-domain approach by examining datasets with different backgrounds of user-generated feedback:

- IMDb Movie Reviews Dataset: A standard benchmark for binary classification of user sentiment (positive/negative).
- Amazon Product Reviews Dataset: This multi-class dataset contains polarity ratings between 1 and 5 stars and aids in model granularity evaluation [12].
- Twitter US Airline Sentiment Dataset: Contains actual tweets marked as positive, neutral, or negative, introducing variability in short-form content [15].

These datasets together enable generalization over formal reviews and informal social media comments

c. Data Preprocessing

Text data was standardized cleaning and transformation prior to being fed into machine learning models. Processes included:

- Text Cleaning: Elimination of unnecessary tokens such as HTML tags, URLs, and emojis.
- Tokenization: Breaking sentences into smaller pieces (tokens).
- Stop-word Removal: Eliminating frequently used words that do not impact sentiment (for example, "the", "is").
- Lemmatization: Reducing words to their dictionary base forms.
- Padding: Padding input sequences to a constant length for batch processing [12, 14].
- The above operations enhance model learning by raising signal-to-noise ratio and computational efficiency.

d. Feature Extraction

We employed both traditional and modern feature representation techniques to convert text into numeric form:

- Bag-of-Words (BoW) and TF-IDF: Classical sparse representations suitable for traditional machine learning classifiers [15].

Word Embeddings:

- Word2Vec and GloVe offered semantic-rich, fixed-size word vectors based on distributional similarity [16,17].
- BERT embeddings captured contextual word meaning, significantly improving performance in sentiment tasks [18, 14].

This combination allowed us to contrast lightweight vector models with cutting-edge transformer-based encodings

e. . Model Development and Training

A range of deep learning architectures were investigated and deployed:

- Convolutional Neural Networks (CNN): Detected local semantic patterns like n-grams [14].
- Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM): Managed long-term dependencies and contextually relevant relationships across sequences [19].
- BERT (Bidirectional Encoder Representations from Transformers): Fine-tuned per dataset, enabling deep semantic comprehension at the word and sentence level [18].

Training Parameters:

- Loss Function: Binary or categorical cross-entropy based on task type.
- Optimizer: Adam with a learning rate of 0.001.
- Batch Size: 32.

- Epochs: 10 to 15, with early stopping to prevent overfitting.
  Regularization (dropout and L2) was used to improve generalization..

f. ***. Evaluation Strategy***

We applied a holistic set of evaluation measures:

- Accuracy: Overall accuracy of classification.
- Precision, Recall, F1-Score: Estimated sensitivity and specificity per class.
- ROC-AUC Curve: Evaluated performance over decision thresholds [15].
- Confusion Matrix: Plotted true vs. predicted class distributions [13].

IV. Results

The sentiment analysis models were tested on three varied datasets: IMDb Movie Reviews, Amazon Product Reviews, and Twitter US Airline Sentiment. The findings show the relative efficiency of various deep learning architecture and feature representations in managing formal and informal user-generated content.[20][21]

A. .Model Performance Overview

The accuracy ratings by models and datasets are presented in Table 1. The BERT model performed better than the conventional deep learning models (CNN, LSTM, Bi-LSTM) consistently, with the highest accuracy on all datasets. In particular, BERT achieved accuracies of 92% for IMDb, 86% for Amazon, and 84% for Twitter data.[22][23]

Table 1: Sentiment classification model accuracy comparison.

Model	IMDb Accuracy	Amazon Accuracy	Twitter Accuracy
CNN	85%	78%	75%
LSTM	87%	80%	77%
Bi-LSTM	88%	81%	79%
BERT	92%	86%	84%

B. Detailed Metric Analysis on Twitter Dataset

For the case of the Twitter US Airline Sentiment dataset, BERT performed evenly for all classes as seen Precision, recall, and F1-scores for positive, neutral, and negative sentiments were always above 0.83, which is indicative of strong sensitivity and specificity.[20][24]

Metric	Positive	Neutral	Negative
Precision	0.90	0.85	0.88
Recall	0.89	0.83	0.87
F1-Score	0.89	0.84	0.87

C. ROC Curve and AUC Analysis

ROC plots indicate that BERT achieved the highest Area Under Curve (AUC) score of 0.95 on the IMDb dataset, significantly greater than CNN (0.88) and LSTM (0.90). This indicates improved discriminative ability at varying classification thresholds.[22][25]

D. Insights from Confusion Matrix

BERT's confusion matrix across the Twitter data set reflects high prediction accuracy with most of the true labels being labeled correctly. There is little confusion between neutral and positive classes typical of informal texts that are short in nature.[20][24]

E. Efficiency while Training

Training and validation loss plots indicate that all models had converged after 15 epochs, with BERT benefiting from early stopping that avoided overfitting. Regularization methods like dropout and L2 regularization played a key role in better generalization[23][25]

V. Conclusion

The research illustrates that using deep learning techniques, Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM), can generate highly scalable and accurate sentiment analysis models for processing user feedback from social media sites and web pages.By choosing great feature encoding methods and utilizing benchmarked datasets like IMDB, Yelp, Amazon Products, also Amazon Fine Food Reviews, the technique improves customer sentiment representation and classification. This thoughtfully architected deep learning system not only enhances classification accuracy but also enhances model development, rendering sentiment analysis simpler, more efficient, and effective in a range of business scenarios.

Some of the key advantages of such an approach are:

A. Enhanced Accuracy and Performance:

Optimization of model design and parameters results in the deep learning models performing consistently better or equally well compared to current approaches, quantified in terms of accuracy, precision, recall, and F1-score.

B. Operational Efficiency:

Carefully calibrated models and effective feature encoding methods decrease computational complexity, making it possible to train quickly & perform sentiment analysis in real-time.

C. Better Business Insights:

Precise sentiment classification enables businesses to derive more meaningful, actionable insights from customer opinions, enhancing strategic decision-making and customer relationships.

D. Scalability and Portability:

Use of deep learning models makes sure that sentiment analysis solutions can be implemented across diverse platforms & datasets, providing scalable tools that companies of all sizes can utilize.

Overall, this study puts the revolutionary promise of deep learning in sentiment analysis into perspective through the balance between high predictive ability and operational effectiveness. The research provides a robust foundation for more intelligent, customer-focused business strategy development and sets the stage for future breakthroughs through the integration of even stronger architectures such as transformers.

References

[1]B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, 2nd ed., Cambridge University Press, 2020.

[2]A. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, "Affective computing and sentiment analysis," IEEE Intelligent Systems, vol. 31, no. 2, pp. 102–107, 2016.

[3]M. Pontiki et al., "SemEval-2016 task 5: Aspect based sentiment analysis," in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 2016, pp. 19–30.

[4]S. Poria, E. Cambria, D. Hazarika, and N. Mazumder, "Multi-level attention network for aspect-based sentiment analysis," Computational Linguistics, vol. 45, no. 1, pp. 161–193, 2019.

[5]J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019.

[6]Y. Kim, "Convolutional neural networks for sentence classification," in Proc. EMNLP, 2014, pp. 1746–1751.

[7]P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," Transactions of the Association for Computational Linguistics, vol. 5, pp. 135–146, 2017.

[8]T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," IEEE Computational Intelligence Magazine, vol. 13, no. 3, pp. 55–75, 2018.

[9]B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.

[10]B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002, pp. 79–86.

[11]J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.

[12]F. Chollet et al., Keras. GitHub repository: https://github.com/keras-team/keras, 2015.

[13]F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[14]J. Hunter, M. Droettboom, T. A. Caswell, et al., Matplotlib: Visualization with Python, Version 3.6, 2022.
[15]C. Sanders, "Twitter US Airline Sentiment," Kaggle Dataset, https://www.kaggle.com/crowdflower/twitter-airline-sentiment, 2015.
[16]T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," arXiv preprint arXiv:1301.3781, 2013.
[17]J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation," in Proc. of EMNLP, 2014.
[18]J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.
[19]S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[20]C. Sanders, “Twitter US Airline Sentiment,” Kaggle, https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment, 2015.
[21]Maas, A., Daly, R., Pham, P., et al., “Learning Word Vectors for Sentiment Analysis,” in Proc. of ACL, 2011.
[22]Devlin, J., Chang, M. W., Lee, K., & Toutanova, K., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” NAACL-HLT, 2019.
[23]Mikolov, T., Chen, K., Corrado, G., & Dean, J., “Efficient Estimation of Word Representations in Vector Space,” arXiv preprint arXiv:1301.3781, 2013.
[24]Ghosh, S., Chollet, F., “Bidirectional LSTM for Sentiment Analysis,” TensorFlow Tutorials, https://www.tensorflow.org/tutorials/text, 2021.
[25]Saito, T. &Rehmsmeier, M., “The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets,” PLOS ONE, 2015.