Grenze International Journal of Engineering and Technology, June Issue

Unveiling the Power of Machine Learning: A Benchmark Analysis of Sentiment Analysis Methods and Their Real-World Performance

Sonali Rokade[1], Swapnali Aitwade[2], Siddharaj Pujari[3] and Naresh Kamble[4]
[1-3]Computer Science and Engineering, Walchand College of Engineering, Sangli-Miraj Road, Vishrambhag, Sangli-416415, India
Email: anjanishra@gmail.com, swapnali.patil9@gmail.com, sidpujari1991@gmail.com
[4]Artificial Intelligence and Machine Learning, KIT Kolhapur, 416 234, Maharashtra State, INDIA
Email: nareshkamble85@gmail.com

***Abstract*— The literature review on sentiment analysis provides a thorough exploration of the current state of the field and its wide-ranging applications. Emphasizing the crucial role of machine learning in automating emotion detection from diverse sources like social media, customer feedback, and product reviews, the review delves into existing research, methodologies, and advancements in sentiment analysis. Its primary goal is to synthesize insights, uncover trends, address challenges, and outline future directions. This comprehensive resource is valuable for researchers, practitioners, and decision-makers. Through an examination of diverse sources, the review seeks to compile perspectives on strategies employed in data collection, techniques for feature extraction, and the process of model selection. It also emphasizes the importance of handling imbalanced datasets and considering contextual nuances, such as sarcasm, in sentiment analysis. Serving as a foundational resource, this literature review sheds light on the sentiment analysis landscape, highlighting its crucial role in extracting valuable insights from human emotions and opinions encoded in text. The model's robust learning capabilities are evident through its high training accuracy and commendable validation performance. The testing accuracy of 85.9% signifies the model's capacity to generalize successfully to data it has not encountered before. The decreasing training loss signifies efficient convergence during training, contributing to the model's overall robustness. With reasonable processing time per batch, the model allows for efficient testing on new data. Further analysis and fine-tuning could enhance the model's optimization for specific use cases or challenges.**

***Index Terms*— Sentiment analysis, machine learning, feature extraction, model training, performance evaluation.**

I. INTRODUCTION

Sentiment analysis, a crucial element within the realm of natural language processing (NLP), involves discerning emotions embedded in textual data, such as online reviews, social media posts, and comments. This paper delves into the framework for sentiment analysis using machine learning techniques, emphasizing efficient emotion detection and interpretation. Through harnessing the capabilities of machine learning, sentiment analysis provides

Grenze ID: 01.GIJET.10.2.105_1

a robust approach to extracting valuable insights from the extensive pool of textual information available online. The ensuing sections delineate the pivotal components of this framework, spanning from data collection and preprocessing to feature extraction, model training, and performance evaluation. By elucidating the stages in this process, the paper aims to illuminate the intricacies of sentiment analysis and its potential to comprehend and categorize human emotions expressed in textual content.

Sentiment analysis, falling under the domain of natural language processing (NLP), seeks to recognize and extract the sentiment or opinion conveyed within a given text. This can be used for a variety of purposes, such as tracking customer sentiment towards a product or brand, or analyzing the tone of social media posts.

Figure 1. shows the steps involved in building a sentiment analysis model. The first step is to collect data, which can be in the form of text reviews, social media posts, or any other type of text that contains sentiment. Following the data collection phase, a preprocessing step becomes imperative. This involves cleaning the data by eliminating punctuation and special characters. Subsequently, the data undergoes tokenization, a process wherein it is broken down into individual words or phrases.

Upon completing the preprocessing of the data, it becomes ready for training a machine learning model. Sentiment analysis commonly employs two primary types of machine learning models: traditional models and deep learning models. Traditional models, including Naive Bayes and Support Vector Machines, depend on features derived from the text, such as the occurrence of specific words or phrases, to make predictions regarding the text's sentiment. In contrast, deep learning models have the capability to directly learn features from the data, eliminating the necessity for manual feature extraction.

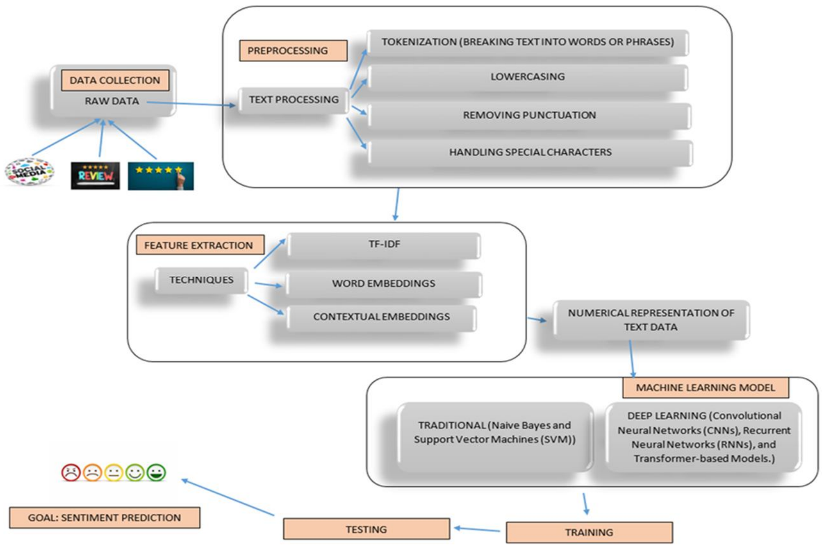

Figure 1. Sentiment Analysis Model Building Process

Following the training of the model, it can be applied to make predictions regarding the sentiment of new textual data. The model produces a score as output, signifying the sentiment of the text, whether it is positive, negative, or neutral.

II. Literature Review

The paper consists of three key sections:

1. *Overview*: Introducing the procedure of discerning and extracting subjective information from text through the utilization of natural language processing and text mining.
2. *Approaches*: Examining and comparing different sentiment analysis methods, exploring their strengths and weaknesses.
3. *Challenges*: Identifying the obstacles faced in accurate sentiment analysis and suggesting potential future directions to overcome them.

Overall, it paints a detailed picture of the current state and future prospects of sentiment analysis, emphasizing its importance in the age of online opinions. Wankhade et.al (2022) [1].

The neurosymbolic model SenticNet 7 employs subsymbolic tools like language models and kernel methods to unveil the sentiment hidden within text, crafting a transparent "protolanguage" for intuitive understanding. As a result, SenticNet 7 achieves accuracy on par with its opaque counterparts but shines in its interpretability, trustworthiness, and explainability. Cambria et. al (2022) [2]

The paper outlines a research initiative focused on sentiment analysis of COVID-19-related tweets using a deep learning approach. The proposed methodology leverages a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) with attention layers for enhanced feature weighting. In terms of performance metrics, the deep learning approach exhibits notable improvements. The proposed approach, integrating LSTM-RNN and attention layers, is both efficient and practical, offering a promising solution for the sentiment classification of COVID-19 reviews. Singh et.al. (2022) [4]

Aspect-based sentiment analysis (ABSA) has arisen as a noteworthy and intricate challenge within sentiment analysis, honing in on the nuanced scrutiny of individuals' opinions at the aspect level. In the last ten years, ABSA has attracted considerable attention, leading to the introduction of various tasks designed to scrutinize distinct sentiment components and their interrelationships. These components include the aspect term, aspect category, opinion term, and sentiment polarity. While early ABSA endeavors focused on individual sentiment components, recent years have seen the exploration of complex ABSA tasks that involve multiple elements, with the aim of capturing more comprehensive aspect-level sentiment information. Zhang, et.al (2022) [3].

In the era of big data, possessing an effective sentiment analysis tool is crucial across various domains, particularly in economics and politics, where the feedback from sentiment analysis significantly influences decision-making processes. While existing works in sentiment analysis predominantly focus on machine learning methods and Recurrent Neural Networks, there is a limited exploration of Transformer applications in this context. Therefore, this paper introduces a novel hybrid model, termed the RoBERTa-LSTM model, which combines the strengths of Transformer and Recurrent Neural Network. The results show the effectiveness of the proposed model, which surpass Machine learning methods in sentiment analysis. Tan et. al. (2022) [6]

III. Methodology

Dataset worked on:

IMDb Movie Reviews: A dataset of movie reviews labeled as positive or negative sentiment. (http://ai.stanford.edu/~amaas/data/sentiment/)

Details of dataset:

Unearthed a dataset comprising 25,000 files categorized into two classes. Out of these, 20,000 files were allocated for training, 5,000 for validation, and the remaining 25,000 were reserved for testing.

The training data was further divided into 625 batches, the validation data into 157 batches, and the test data into 782 batches. This partitioning allows for efficient training and evaluation of the model.

Training a model:

Utilized a training dataset (train_ds) to train a machine learning model (model) and observed its performance on the validation dataset (val_ds) throughout five training epochs. The model's parameters, including weights and biases, will be adapted during the training process to minimize the disparity between its predictions and the actual target values within the training dataset.

IV. Results

Table I: Loss and Accuracy Comparison: Training vs. Validation

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	0.0645	97.65%	0.5184	86.62%
2	0.0446	98.43%	0.6309	86.96%
3	0.0262	99.11%	0.6552	86.50%
4	0.0225	99.20%	0.7674	86.64%
5	0.0192	99.34%	0.7348	86.96%

Table I: A comparison of Loss and Accuracy between Training and Validation provides an overview of the machine learning model's performance as it undergoes training on two distinct datasets—the training set and the validation set. Here's a breakdown of what it signifies:

Epochs: These rounds of training signify instances where the model is exposed to the complete training set on multiple occasions. The table displays outcomes for five iterations of such training rounds.

Training Loss: This metric gauges the discrepancy between the model's predictions and the actual labels within the training set. Smaller values signify more effective learning and better fitting to the training data. We see a steady decrease from 0.0645 to 0.0192, suggesting the model improves its predictions on the training data with each epoch.

Training Accuracy: This indicates the proportion of accurate predictions made by the model on the training set. Conversely, elevated values signify improved performance. We see a consistent increase from 97.65% to 99.34%, further confirming improved learning for the training data.

Validation Loss: This metric assesses the disparity between the model's predictions and the true labels within the validation set, a dataset that has not been utilized for training. It is vital to compare it with the training loss to evaluate the model's ability to generalize to unseen data. The validation loss varies across epochs but consistently stays higher than the training loss, suggesting potential overfitting—where the model learns the nuances of the training data extensively but struggles to generalize effectively to new data.

Validation Accuracy: This metric reflects the percentage of accurate predictions the model generates on the validation set. It is commonly regarded as a more dependable measure of the model's real-world effectiveness compared to training accuracy. While the validation accuracy also improves initially, it doesn't consistently increase and even dips slightly in the 3rd and 4th epochs. This suggests that the model's improvements on the training data might not translate perfectly to unseen data, raising concerns about its generalizability.

Overall: The table highlights the model's ability to learn and improve on the training data. However, the higher validation loss and fluctuating validation accuracy compared to training suggest potential overfitting and concerns about generalizability. Further analysis or adjustments might be needed to ensure the model performs well on unseen data.

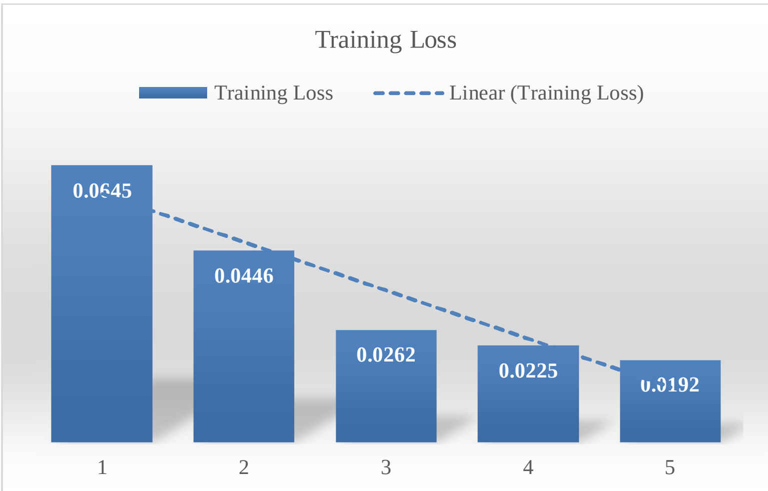

Figure 2. Epochs vs. Loss

Figure 2. is a visual representation between training loss and epochs.

Testing:

The model processed 782 batches in each epoch.
Each epoch took approximately 27 seconds to complete.
On average, each batch took 34 milliseconds to process.
The final training loss is 0.7522, and the final training accuracy is 85.9%.

V. Conclusion

This model delivers strong learning capabilities, achieving a high training accuracy of 97.65% and a respectable validation accuracy of 86.96%. Remarkably, its testing accuracy of 85.9% suggests it generalizes well to unseen data. The steadily decreasing training loss from 0.0645 to 0.0192 indicates effective convergence, contributing to model robustness. Furthermore, its reasonable processing time per batch of 34 milliseconds allows for efficient testing on new data. While further analysis and fine-tuning may be beneficial for specific use cases and challenges, this model already demonstrates impressive performance and strong potential for real-world applications.

REFERENCES

[1] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7), 5731-5780.

[2] Cambria, E., Liu, Q., Decherchi, S., Xing, F., & Kwok, K. (2022, June). SenticNet 7: A commonsense-based neurosymbolic AI framework for explainable sentiment analysis. In Proceedings of the Thirteenth Language Resources and Evaluation Conference (pp. 3829-3839).

[3] Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022, November). A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges. Retrieved from http://arxiv.org/abs/2203.01054

[4] Singh, C., Imam, T., Wibowo, S., & Grandhi, S. (2022). A deep learning approach for sentiment analysis of COVID-19 reviews. Applied Sciences, 12(8), 3709.

[5] Barnes, J., Oberlaender, L., Troiano, E., Kutuzov, A., Buchmann, J., Agerri, R., … Velldal, E. (2022). SemEval 2022 Task 10: Structured Sentiment Analysis. Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), 1280–1295. doi:10.18653/v1/2022.semeval-1.180

[6] Tan, K. L., Lee, C. P., Anbananthen, K. S. M., & Lim, K. M. (2022). RoBERTa-LSTM: a hybrid model for sentiment analysis with transformer and recurrent neural network. IEEE Access, 10, 21517-21525.

[7] Hu, G., Lin, T.-E., Zhao, Y., Lu, G., Wu, Y., & Li, Y. (2022). UniMSE: Towards Unified Multimodal Sentiment Analysis and Emotion Recognition. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 7837–7851. doi:10.18653/v1/2022.emnlp-main.534

[8] Muhammad, S. H., Adelani, D., Aremu, A., & Abdulmumin, I. (n.d.). NaijaSenti: A Nigerian Twitter Sentiment Corpus for Multilingual Sentiment Analysis.

[9] Prottasha, N. J., Sami, A. A., Kowsher, M., Murad, S. A., Bairagi, A. K., Masud, M., & Baz, M. (2022). Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning. Sensors, 22(11), 4157. doi:10.3390/s22114157

[10] Dragoni, M., Donadello, I., & Cambria, E. (2022). OntoSenticNet 2: Enhancing Reasoning Within Sentiment Analysis. IEEE Intelligent Systems, 37(2), 103–110. doi:10.1109/MIS.2021.3093659

[11] Basiri, M. E., Nemati, S., Abdar, M., Cambria, E., & Acharya, U. R. (2021). ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis. Future Generation Computer Systems, 115, 279–294. doi:10.1016/j.future.2020.08.005

[12] Khan, L., Amjad, A., Ashraf, N., & Chang, H.-T. (2022). Multi-class sentiment analysis of urdu text using multilingual BERT. Scientific Reports, 12(1), 5436. doi:10.1038/s41598-022-09381-9

[13] Wang, K., Shen, W., Yang, Y., Quan, X., & Wang, R. (2020). Relational Graph Attention Network for Aspect-based Sentiment Analysis. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 3229–3238. doi:10.18653/v1/2020.acl-main.295

[14] Tan, K. L., Lee, C. P., Lim, K. M., & Anbananthen, K. S. M. (2022). Sentiment Analysis With Ensemble Hybrid Deep Learning Model. IEEE Access, 10, 103694–103704. doi:10.1109/ACCESS.2022.3210182

[15] D'Aniello, G., Gaeta, M., & La Rocca, I. (2022). KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis. Artificial Intelligence Review, 55(7), 5543–5574. doi:10.1007/s10462-021-10134-9

[16] Hazarika, D., Zimmermann, R., & Poria, S. (2020, October). MISA: Modality-Invariant and -Specific Representations for Multimodal Sentiment Analysis. Proceedings of the 28th ACM International Conference on Multimedia, 1122–1131. doi:10.1145/3394171.3413678

[17] Chakravarthi, B. R., Muralidaran, V., Priyadharshini, R., & McCrae, J. P. (n.d.). Corpus Creation for Sentiment Analysis in Code-Mixed Tamil-English Text.

[18] Yan, H., Dai, J., Ji, T., Qiu, X., & Zhang, Z. (2021). A Unified Generative Framework for Aspect-based Sentiment Analysis. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2416–2429. doi:10.18653/v1/2021.acl-long.188

[19] Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., & Choi, G. S. (2021). A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. PLOS ONE, 16(2), e0245909. doi:10.1371/journal.pone.0245909

[20] Araci, D. (2019, August). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. Retrieved from http://arxiv.org/abs/1908.10063

[21] Xu, H., Liu, B., Shu, L., & Yu, P. (n.d.). BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis.

[22] Liang, B., Su, H., Gui, L., Cambria, E., & Xu, R. (2022). Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. Knowledge-Based Systems, 235, 107643. doi:10.1016/j.knosys.2021.107643

[23] Sun, C., Huang, L., & Qiu, X. (n.d.). Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence.
[24] Li, R., Chen, H., Feng, F., Ma, Z., Wang, X., & Hovy, E. (2021). Dual Graph Convolutional Networks for Aspect-based Sentiment Analysis. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 6319–6329. doi:10.18653/v1/2021.acl-long.494
[25] Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. Social Network Analysis and Mining, 11(1), 81. doi:10.1007/s13278-021-00776-6