

Comparative Analysis of Sentiment Analysis Models: RoBERTa vs. VADER

Amal Krishna P¹, Balendu M S² and Rahul Varma U³

¹⁻³Department of Computer Science and Applications, Amrita Vishwa Vidyapeetham, Kollam, India
Email: amalkrismf@gmail.com, msbalu2424250@gmail.com², u.rahulvarma@gmail.com

Abstract—Opinions are statements that reflect people's perception or sentiment. In this paper, we have done sentimental analysis using two major NLP models - VADER and RoBERTa using sentiment analysis. The study begins with an overview of RoBERTa regarding its pre-training and optimization enabling dynamic data collection. Similarly, the legal status of VADER is explained in terms of being based on dozens of opinions and predetermined rules. We are testing these models using two different dataset customers for evaluating the accuracy and precision of each model in predicting the sentiment of text. The results are analyzed and compared to know the problems and weakness of the models. These tests include training and refinement of RoBERTa's emotional data list model which uses transformer model to process input sequences, while VADER relies on pre-generated lexicon text. Different methods are used to compare the performance of two models and identify their strengths and weaknesses in different situations.

Index Terms— VADER, RoBERTa, Lexicon, NLP, Sentimental analysis, Transformer model.

I. INTRODUCTION

Sentimental Analysis is a process of extracting and also analyzing the emotions of users' tone towards texts like customer feedback or random tweets for determining whether it is positive, negative or neutral. Customer feedback is important information for a business or any other companies because it provides valuable insights or opinions on products. So it can be used for improvement or betterment of the products based on the users feedback.

This study provides insight into the comparison between the VADER and RoBERTa models in emotional assessment. Among the many sentimental analyses available today, we are using two major models: Vader and Roberta. The Vader model is a dictionary and rule-based convention that is widely used because of its simplicity and speed. Roberta is a reference to the Robustly Optimized BERT method, which represents a cutting-edge deep learning model known for its performance in various NLP tasks helping researchers and practitioners select the appropriate model based on their specific needs and limitations. analytical thinking and its applications in today's modern information world. In coming sections we will dive deeper into each model's architecture, advantages, and limitations in sentimental analysis.

II. LITERATURE SURVEY

From the research paper [1] Extracting text is an important step in analyzing text. Journals are usually in PDF

format, which is a partial document. The Journal is divided into several sections such as Introduction, Procedure, Experimental setup, Results, and Evaluation.

The main purpose of partial inference is to find a representative of the data containing information about the entire collection. Various methods for extracting partitions from research data include statistical methods, NLP, machine learning, and others. In this article, we present an investigation of various methods to extract from PDF files.

Research paper [2] shows that Social media such as Twitter and Facebook are seen as places where emotions are created in today's world. The insights from those tweets can provide policymakers and others with insights to better recommend and enforce policies.

The data collected from twitter during November 9 to December 3 were considered for analysis. Data sets are preprocessed to clean up the data and enable analysis. Small set of 5,000 tweets was analyzed using their different techniques such as support vector machines, pure Bayesian classifiers, and decision trees.

Sentiment analysis is an invaluable tool that can be used to understand web sentiment and help make more informed decisions. Opinions can be positive, negative, or objective, and the way they give weight to concepts and expressions is important in determining the accuracy of the distribution of opinions.

Their work [3] is to improve demand classification by modifying the sentiment results returned by SentiWordNet based on the content of the word associated with the increased vocabulary.

Using digital camera product reviews from Amazon and eBay, they test their classification hypothesis and show that our method improves prediction accuracy.

The article [4] presents a new methodology for understanding the thoughts and feelings behind the comments of users of various e-commerce websites. This system has an advantage over the existing star rating system by providing users with better and more descriptive results.

They show us that an important part of human behavior is always knowing what other people think about the next event. With the continuous development of e-commerce sites and personal blogs, various technologies are used to understand and share ideas.

In the research paper [5] works were done within constraints as part of the joint project on Sentiment Analysis of Indian Languages (SAIL 2015). The task is to divide the tweet data into three pole groups such as positive, negative and neutral. A trilingual twitter dataset is provided in Hindi, Bengali and Tamil for education.

Their method uses binary features, statistical features generated by SentiWordNet, and word availability (binary features). Due to the sparsity of the generated features, the input features are mapped to a random Fourier feature space for separation, and regular least squares are used for linear distribution.

Among the test tweets in Tamil, positive tweets were detected more than any other bipolar group. Due to the lack of specific words and ways of thinking, tweets in the middle are less well known and also lead to misclassification of the tri-pole group. This encourages us to carry out our research in this area in a planned manner.

Sentiment analysis paper [6] of Tamil online news to classify features of famous events, people and places. It becomes the main source of information, from private information to public opinion. People announce events related to their lives, express their views on politics and religion, express support or opposition to social issues, discuss products, education, entertainment and more.

III. PROPOSED SYSTEM

Sensitivity analysis, also known as sentiment mining, is the process of determining sentiment or sentiment behind a document. Here are the general steps for analyzing emotions:

A. Here are the steps to determine the sentiment of a text:

A. Data Collection

The appropriate amount of Data used in this paper is an arranged set of product reviews collected from <http://kaggle.com>. The database spans 10 years and contains a total of ~500,000 reviews as of October 2012. Reviews include materials and reference materials, ratings, and literature reviews. Dataset distribution of Amazon review is given in Figure 2.

B. Data processing

Clean and pre-write data to remove noise and irrelevant data. These steps usually include removing special characters, punctuation, and static words, and normalizing the text (eg, converting it to lowercase).

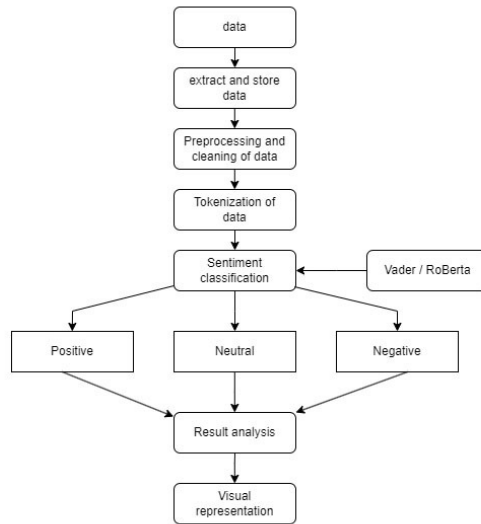


Figure 1. Architecture



Figure 2.Dataset Distribution

C. Tokenization and Feature extraction

In this process we split the data into different words or tokens. So that it will create a well structured data representation. After that we will identify the main features or features that will be used to represent the text's point of view. This may include bag representations, n-grams or more advanced techniques such as word embedding.

D. Sentiment classification and model training

Training the emotional behavior using appropriate machine learning or natural language processing (NLP). This involves labeling a dataset with a label (positive, negative, neutral) and using it to train the model to predict the sentiment of given text. Then we will split the data for training and testing. Using the model we train the data and evaluate its performance using test data. Evaluation parameters usually include accuracy, precision, recall, and F1 score.

E. Post processing and Analysis

Analyzing the prediction and interpreting the results. This may include collecting sentiment scores, visualizing sentiment, or gaining insights from sentiment results.

IV. ALGORITHMS

A. VADER

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a dictionary and sentiment management system that uses a list of terms to determine the sentiment of a text. It is a simple model that requires no prior training, making it easy to use and operate. However, it does not work well with hard and fine-grained text.

B. RoBERTa

- Optimization Methods for Pre-Training NLP Systems
- RoBERTa (Robustly Optimized BERT Pre-Training Method) method used for pre-training language processing (NLP) which enhances the bidirectional encoder represented by Transformers (BERT).

C. Transformers

Roberta model is based on transformer architecture which is famous for its great performance for doing NLP tasks.

Transformers in NLP is a new architecture designed to solve periodic tasks while easily sustaining long-term success. It relies solely on itself to calculate the representation of its components and components, without using parallel RNNs or networks. The model is explained in Figure 3.

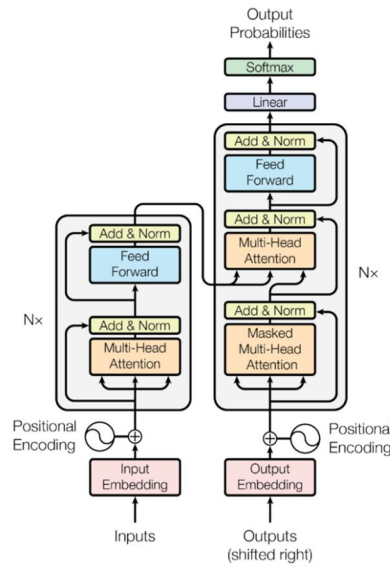


Figure 3. Transformer Architecture

V. RESULTS AND DISCUSSIONS

In Figure. 5 We can see the positive and negative results obtained from the Vader and roBERTa models. As it turns out, roBERTa is more confident in the results. Both positive and negative outcomes are competing for one side. It was more contagious than Vader. The score indicates a score from 1 to 5 given by the dataset, which can contain texts containing positive or negative emotions, in comparison each score can have several emotions, i.e. score 1 is the lowest score for negative emotions, but articles can contain negative or positive emotions and feelings.

As the graph shows, Roberta is more confident in the outcome. In the graph, positive and negative values are on one side. Compared to Vader, it is more spread in Figure. 4. The more congested one side of the graph is, the more confident the model is in its predictions. In the dataset, there is a label called stars, where the user gives 1 to 5 stars for a review. Although each star review contains a mixed text review of positive or negative sentiment, any of the scores can contain many sentiments. For example, score 1, which is the lowest score, indicates a negative sentiment, but the text can contain both negative and positive sentiment. Which here is accurately distinguished. This is activated by RoBERTa with the help of transformers and fine tuning.

In the given Table 1, We ran a sentimental prediction using a labeled dataset that contains random tweets from people. From these prediction results, We can clearly see that if the sentence is positive, Roberta positive values are way more accurate than the Vader compound values. As in the negative sentence we can see that roberta negative values are much more accurate than vader compound values. In the table, we can see that for the text 'Uh oh, I am sunburned, the vader compound value is 0, which indicates neutral. In the case of Roberta, it clearly shows that the text has a negative sentiment, as the value of 0.745 is negative. "Uh oh" is interpreted as a neutral interjection or an expression of surprise rather than a strong indicator of negative sentiment. Additionally, the phrase itself doesn't inherently convey strong positive or negative emotions. While sunburn can be uncomfortable or painful, the statement itself doesn't contain explicit emotional language. Based on our research

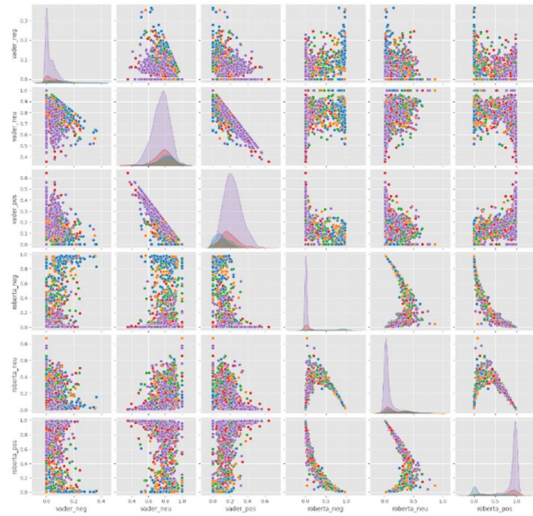


Figure 4. Sentiment Distribution

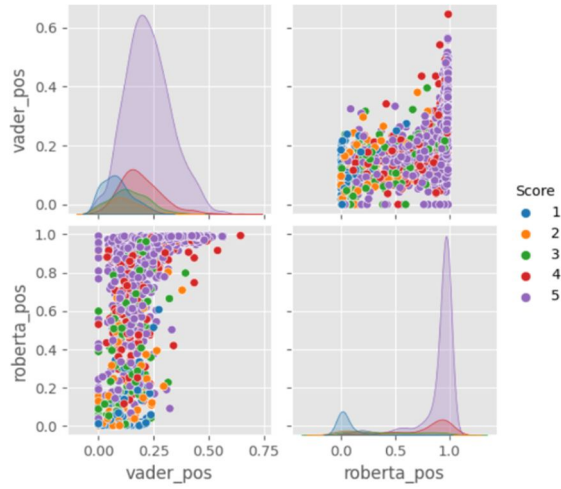


Figure 5. Roberta Accuracy

TABLE I

	Sentiment Prediction				
	Text	Sentiment	roberta_pos	roberta_neg	Vader_comp
1	That's very funny.	positive	0.964	0.003	0.755
2	Uh oh, I am sunburned	negative	0.028	0.745	0
3	I'm starving!!	negative	0.961	0.961	0.635
4	Sick. Feel like jello	negative	0.005	0.948	0.202
5	Aww, congrats. ;)	positive	0.985	0.013	0.681

and sentiment prediction, many studies and results have shown that transformer models like Roberta often outperform rule-based approaches like VADER in sentiment analysis tasks.

VI. CONCLUSION

In this research article, we explored and analyzed the differences between vader and roberta using sentimental analysis. With some research and analysis, we have identified some key issues, challenges and opportunities in the field. We evaluated and compared the models to solve the research question.

The choice of VADER and RoBERTa depends on many factors, such as the specific needs of the analysis project, availability of resources and the best balance between accuracy and flexibility. It's worth trying both methods and evaluating their effectiveness on your particular data to determine the most appropriate option.

However, it is important to remember that there are no models that are perfect. While VADER is helpful, it relies on predefined notions and assumptions that may prevent it from adapting to new domains or evolving language use. It does have limitations and may not always capture the full context or intent behind a statement. They can sometimes produce results that may seem odd or contrary to human interpretation. Although RoBERTa is powerful, it requires a lot of computational and training data to achieve optimum performance. Additionally, both models may have problems understanding emotions based on sarcasm or subtle content.

ACKNOWLEDGMENT

We want to thank our project guide, Rahul Varma U [Faculty Assistant], for his constant support and guidance throughout the project. His valuable suggestions and feedback have been instrumental in shaping this project and also grateful to Ms. Remya Rajesh and Ms.Kavitha K R, [Assistant Professors], for providing us with the necessary resources and infrastructure to carry out this project.

REFERENCES

- [1] K. Jayaram and Sangeeta, K., "A review: Information extraction techniques from research papers", in IEEE International Conference on Innovative Mechanisms for Industry Applications, ICIMIA 2017 - Proceedings, 2017, pp. 56-59.
- [2] Dhanya N. M. and Harish, U. C., "Sentiment analysis of twitter data on demonetization using machine learning techniques", Lecture Notes in Computational Vision and Biomechanics, vol. 28, pp. 227-237, 2018.
- [3] Jasmine Bhaskar, Sruthi, K., and Prof. Prema Nedungadi, "Enhanced sentiment analysis of informal textual communication in social media by considering objective words and intensifiers", in IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), 2014, Jaipur, 2014
- [4] A. C. Jishag, Rakshesh, V., Mohan, S., N. Varma, V., Shabu, V., Nair, L. S., and Menon, M., "Automated Review Analyzing System Using Sentiment Analysis", Advances in Intelligent Systems and Computing, vol. 904, pp. 329-338, 2019.
- [5] Sachin Kumar S, Premjith B, M Anand Kumar, KP Soman, AMRITA-CEN-NLP@SAIL2015: Sentiment analysis in Indian language using regularized least square approach with randomized feature learning, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, Volume 9468, Hyderabad India, p.671-683 (2015) (Scopus)
- [6] R. G., K. Prasath, N., and Rajasundari, T., "Sentimental Analysis on Tamil News Feed", Journal of Advanced Research in Dynamical and Control Systems, vol. 9, pp. 397-399, 2017.
- [7] Boluki, Ali, et al. Evaluating the Effectiveness of Pre-Trained Language Models in Predicting the Helpfulness of Online Product Reviews. arXiv, 19 Feb. 2023. arXiv.org, <http://arxiv.org/abs/2302.10199>
- [8] Dai, Junqi, et al. Does Syntax Matter? A Strong Baseline for Aspect-Based Sentiment Analysis with RoBERTa. arXiv, 11 Apr. 2021. arXiv.org, <http://arxiv.org/abs/2104.04986>.
- [9] Handoyo, Alif Tri, et al. "Sarcasm Detection in Twitter -- Performance Impact While Using Data Augmentation: Word Embeddings." INTERNATIONAL JOURNAL of FUZZY LOGIC and INTELLIGENT SYSTEMS, vol. 22, no. 4, Dec. 2022, pp. 401-13. arXiv.org, <https://doi.org/10.5391/IJFIS.2022.22.4.401>.