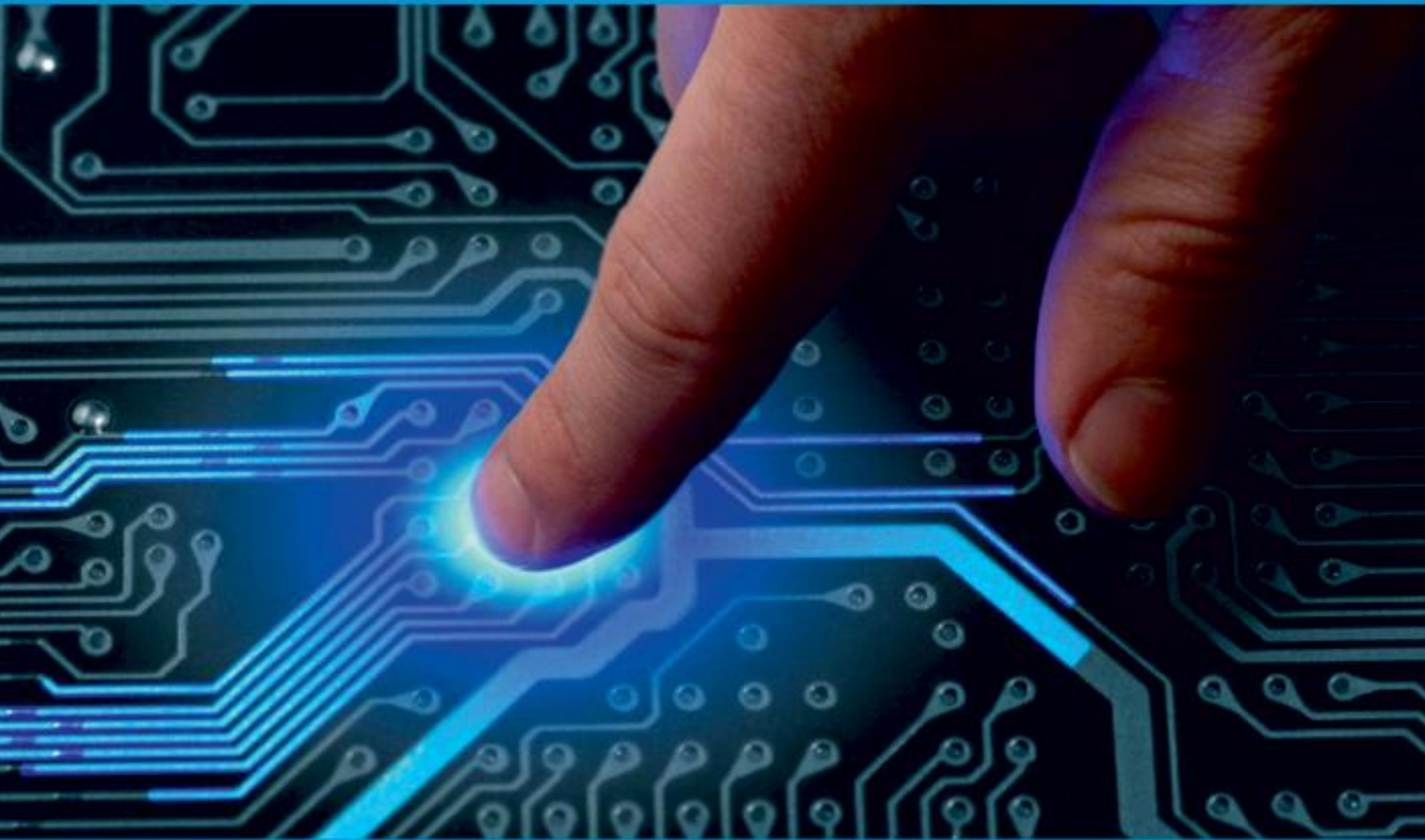
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Special Issue 1, March 2024

1st International Conference on Machine Learning, Optimization and Data Science

Organized by

Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur, India

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

Comparison of Sentiment Analysis using VADER and RoBERTa

Zohaib Hasan[1], Abhishek Singh[2], Rajender Kumar Yadav[3]

Assistant Professor, Dept. of C.S.E., BGIEM, Jabalpur, India [1],

Assistant Professor, Dept. of C.S.E., BGIEM, Jabalpur, India [2]

Assistant Professor, Dept. of E.C.E., BGIEM, Jabalpur, India[3]

ABSTRACT: Analysing human emotions about a product or an incident is called sentiment analysis. In the current world where there are uncountable number of platforms where people free express their feelings, there is an overload ofsentiment based data available. It has become a challenge to gather this data and get some useful information out of it. Sentiment analysis has become an important domain even before the time when various sophisticated techniques of machine learning and deep learning were available. By analyzing social media, product reviews, and customer support interactions, companies can promptly respond to emerging trends, manage their brand reputation, and enhance customer satisfaction. This paper makes an effort to analyze the sentiments of people about a product in Amazon. This paper first uses very basic machine model called VADER (Valence Aware Dictionary and sEntiment Reasoner) and then a more sophisticated Deep Learning approach called RoBERTa (Robustly optimized Bidirectional Encoder Representation from Transformers approach).

KEYWORDS: Sentiment analysis; Machine Learning; Deep Learning; Natural Language Processing; VADER, RoBERTa

I. INTRODUCTION

Sentiment analysis has become an important skill for machines just because of the overload of human comments and reviews on social media platforms like YouTube, Instagram, X etc. and ecommerce platforms like Amazon, Flipkart, eBay etc. Detecting correct human sentiments through text is important to gain deeper insights of customer opinions, preferences, and feedback, empowering them to make informed decisions on the basis of data. Machines are expected to do this analysis and prediction task due to two main reasons. Firstly, analysing so much data is very difficult for human and secondly advancements in various deep learning techniques and loads of data to help them.

A prevalent approach involves categorizing the sentiment of a text based on user satisfaction, dissatisfaction, or neutrality. The polarity may exhibit variations in labeling or employ varying degrees, typically ranging from positive to negative, to represent the emotional tone of a text, spanning from a positive to a negative state. [1]. The main problem is that so many words in English cannot be labelled as positive or negative. Unsupervised learning methods are required to do this task. Human way of writing English is tricky. Some sentences are framed in such a way that only humans with vast experience with the language can understand. A machine model can interpret some keywords as negative, positive or neutral easily but to understand the relationship between words of the same sentence a more complex mechanism is needed. For example, following review on a PC mouse is sarcastically negative:
"*Really this mouse is amazing, sometimes the cursor moves even when I am not touching it*"

Natural Language Processing (NLP) constitutes a subset within the domains of computer science, artificial intelligence and linguistics. Primary objective is to facilitate the comprehension of human communication, encompassing speech and text, by computational systems. NLP encompasses a wide range of techniques to analyse human language. NLP can be used in any of the fields like sentiment analysis, summarization, keyword extraction and tokenization. Some benefits of using NLP are it can be used to analyse both structured and unstructured data, improving customer experience, reducing cost of employing people or bots to analyse loads of used responses and target marketing. Although, the analysis of favorable and unfavorable sentiments is a task that needs a lot of intelligence and deep insight of the textual data [1]. There are several literature and research available that shows that sentiment/emotion analysis can be performed using Traditional machine learning approaches like Naïve Bayes [2], SVM (Support Vector Machines) [9], Logistic Regression [3], Decision trees and random forests and KNN [2] (K-nearest neighbours). There are must sophisticated and quicker analysis methods of Deep Learning like "RNN (Recurrent Neural Network)" [6], "LSTM (Long Short-term Memory)" [5], "GRU (Gated Recurrent Unit)" [7], "CNN (Convolution Neural Network)" [4] and "BERT based models" [8].

II. RELATED WORK

In paper [2], two supervised machine learning algorithms, namely "K-Nearest Neighbour" (K-NN) and "Naïve Bayes", are introduced for the analysis of hotel and movie reviews. The study includes a comparative evaluation of their overall accuracy, precision, and recall values. The findings indicate that, in the context of movie reviews, Naïve Bayes outperformed K-NN with an accuracy of 80.12% compared to 61.81%. However, for hotel reviews, both algorithms exhibited relatively lower and comparable accuracies, with Naïve Bayes at 51.84% and K-NN at 46.31%.
In [3], machine learning techniques like Support Vector Classifier, Stochastic Gradient Descent, Naive Bayes and Logistic Regression are used to analyse sentiments in Twitter reviews in 3 languages. The accuracy obtained are 69%, 71%, 77% and 81% respectively. A dataset of 5295 tweets are used for training and 100 tweets for testing.

This paper [4] uses a framework of Word2vec + CNN. Word2vec is used to transform words into vectors. To categorize sentences into distinct sentiment labels, Convolutional Neural Network (CNN In the paper [1], favorability of a subject is captured from a document. Part of Speech (POS) tagging is done using Markov-model-based tagger. After POS tagging shallow parsing is used to identify the phrase boundaries & local dependencies. After obtaining results from the shallow parser, syntactic dependencies among the phrases are identified. Finally a sentiment polarity has been assigned to be either +1 (positive) or -1 (negative). With benchmark corpus accuracy obtained is 94.3% and with open corpus accuracy was 94.5%. This work is believed to be one of the first on Sentiment Analysis using natural language processing. It became the basis of research for many future researchers.

) is employed. This study incorporates pooling layers, Parametric Rectified Linear Unit (PReLU) layers, convolutional layers and dropout layers within the CNN framework. Notably, in the CNN architecture, a significant portion of the neural network training duration is dedicated to the convolution process. In this technique the test accuracy is just 45%.
LSTM (Long Short Term Memory) used in [5] to analyse IMDB movie reviews and Amazon product reviews. LSTM networks have demonstrated effectiveness in the classification, processing, and prediction of data based on statistical information. Raw text is first tokenized, these tokenized words are then converted into real-valued vectors, and these vectors are then given to the LSTM network. Softmax activation function is then used that scales numbers/logits into probabilities.

The research outlined in reference [6] employs the Recurrent Neural Network (RNN) algorithm for sentiment classification, leveraging its inherent temporal dimension. This temporal characteristic sets RNN apart from feedforward neural networks. The study focuses on a binary classification, distinguishing between positive and negative sentiments. The Traveloka website dataset serves as the basis for training and testing the model, with subsequent experimentation. The outcomes reveal that the model yields exceptional performance, achieving an accuracy rate of approximately 91.9%.

III. PROPOSED METHODOLOGY

The work proposed here is divided into following phases:
1. A dataset of Amazon reviews is imported from Kaggle.
2. Data is cleaned, preprocessed and initial analysis is done.
3. Basics of natural language processing is used from NLTK library to tokenize and apply POS to the comments.
4. "VADER" models and "RoBERTa" models are used for sentiment analysis and prediction.
5. Comparative analysis of both the techniques is performed on basis performance and accuracy.

Amazon fine food dataset is acquired from Kaggle. The dataset contains metadata like Product ID, User ID, Profile Name, etc. and data useful to the problem under discussion like Product Rating out of 5 and Text Review. To speed up the analysis only 500 out 50,000 data is used, although the machine model is pre-trained with the complete dataset. The reviews are also plotted against the rating to check the trend of reviews Figure 1.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

Fig 1 Count of review vs Stars

The Natural Language Toolkit in Python is used to perform basic data pre-processing. 'punkt' is used to tokenize reviewsentences. Using an unsupervised approach, this tokenizer divides a text into a list of sentences in order to build a model for words that begin sentences, collocations, and abbreviations. Before it can be used, it needs to be trained on a sizable collection of target language plaintext. The "averaged perceptron tagger" is then used to tag these tokens with Part of Speech. The frequent tags used to tag parts of speeches are displayed in Table 1.

Table 1 Example POS tags and their meaning

Part-of-speech tag	Description
UKW	Unknown word
IN	Preposition or subordinating conjunction
JJ	Adjective
MD	Modal
NN	Noun
NNP	Proper noun
PRP	Pronoun
RB	Adverb
SYM	Symbol, including all types of punctuation
UH	Interjection
VB	Verb

The tagged words are then fed to “maximum entropy Named Entity Chunker”. It has two English entity chunkers that have already been trained using the ACE corpus. A named entity is something like the Chief Minister, Narendra Modi or Gujarat. Words like store, walk or saw do not come under named entities. Chunk is a substring of text that cannot overlap another chunk. Consider the following sentence:

“Narendra Modi was appointed Chief Minister of Gujarat”

For the sentence above, *Chief Minister* and *Minister* will never be part of a chunk because these words will overlap. So the process of Named Entity chunking is identifying chunks in the text that are named entities.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

In the second phase, “VADER" sentiment analysis is applied. VADER is a sentiment analysis tool that is primarily tailored to the opinions shared on social media. It is built on rules and a vocabulary. The tool is open-source and free.

VADER additionally considers the degree modifier and word order [10]. VADER uses a combination of sentiment lexicon, which is a set of lexical properties (words, for example) that are often classified as positive or negative depending on their semantic orientation. VADER determines the degree of positivity or negativity in a sentiment in addition to providing the positivity and negativity score. VADER sentiment analyser gives polarity scores between 0 and 1 is given to positive, negative and neutral. Any of these traits that gets a score close to 1 will decide that the sentiment is positive, negative or neutral. VADER technique does not always give a score very close to 1 even if the statement is clearly positive or negative. Also this model gets easily confused.
In the third and last phase, “RoBERTa” model is used. Words in the sentence are divided into tokens and converted into numerical representation. Model that is specifically designed for sequence classification is initialized. This model is pre-trained with Twitter data. The NLP pipeline for sentiment analysis is ready to accept text input, tokenize them and pass them through the model to predict sentiments. A function is then created to take sentiment text, tokenize them, apply model, extract scores, convert them into probabilities using softmax activation function.

IV. PSEUDO CODE

Pseudo code for VADER Prediction

```
# Read the dataset from a CSV file
df = read_csv('dataset/Reviews.csv')

# Download the VADER lexicon from NLTK
nltk.download('vader_lexicon')

# Import the SentimentIntensityAnalyzer class from NLTK
from nltk.sentiment import SentimentIntensityAnalyzer

# Import tqdm for showing progress bars
from tqdm.notebook import tqdm

# Initialize a SentimentIntensityAnalyzer object
sia = SentimentIntensityAnalyzer()

# Create an empty dictionary to store sentiment scores
result = {}

# Loops over each row in the dataset
for each row in tqdm(df.iterrows(), total=len(df)):
    # Get the text and ID of the current row
    text = row['Text']
    myid = row['Id']

    # Calculate the polarity scores using VADER for the current text
    scores = sia.polarity_scores(text)

    # Store the polarity scores in the result dictionary with ID as the key
    result[myid] = scores

# Convert the result dictionary into a DataFrame
vaders = pd.DataFrame(result).T

# Rename the index of the DataFrame to 'Id'
vaders = vaders.rename_axis('Id')

# Merge the VADER scores DataFrame with the original DataFrame based on the 'Id' column
vaders = vaders.merge(df, how='left', on='Id')
```

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

Pseudo Code for RoBERTa prediction

```
# Import necessary libraries
import transformers
from transformers import AutoTokenizer, AutoModelForSequenceClassification
from scipy.special import softmax

# Define the pre-trained model identifier
MODEL = "cardiffnlp/twitter-roberta-base-sentiment"

# Initialize tokenizer using the pre-trained model
tokenizer = AutoTokenizer.from_pretrained(MODEL)

# Initialize model for sequence classification using the pre-trained model
model = AutoModelForSequenceClassification.from_pretrained(MODEL)

# Define a function to calculate sentiment scores using the RoBERTa model
function roberta_polarity_scores(example):
    # Tokenize input text using tokenizer

       encoded_text = tokenizer(example, return_tensors='pt')

       # Pass the tokenized text through the pre-trained model
       output = model(**encoded_text)

       # Extract the raw scores from the model output
       scores = output[0][0].detach().numpy()

       # Apply softmax activation to convert raw scores into probabilities
       scores = softmax(scores)

       # Create a dictionary to store sentiment scores
       scores_dict = {
           'roberta_neg': scores[0],
           'roberta_neu': scores[1],
           'roberta_pos': scores[2]
       }

       # Return the dictionary of sentiment scores
       return scores_dict
```

V. SIMULATION RESULTS

Fig 2 shows a bar plot where x-axis is 5 star rating and y-axis is the score that VADER model assigned. The figure shows that positive score is high if the rating is 5, negative score is high of the rating is low as expected. Neutral on the other hand has a flat bar plot for every star rating. Table 2 has the VADER and RoBERTa scores for the review below that is positive without any doubt.

> *"I have bought several of the Vitality canned dog food products and have found them all to be of good quality. The product looks more like a stew than a processed meat and it smells better. My Labrador is finicky and she appreciates this product better than most"*

Table 2 Prediction of VADER vs RoBERTa model for a Positive Review

	Positive	Negative	Neutral
VADER prediction	0.305	0.0	0.695
RoBERTa Prediction	0.94	0.009	0.05

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

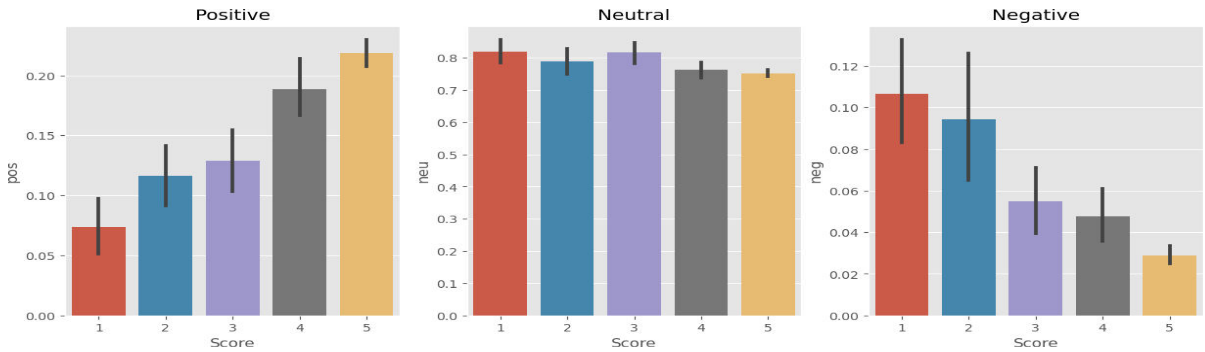


Fig 2 Positive, Neutral and Negative Scores assigned against rating score

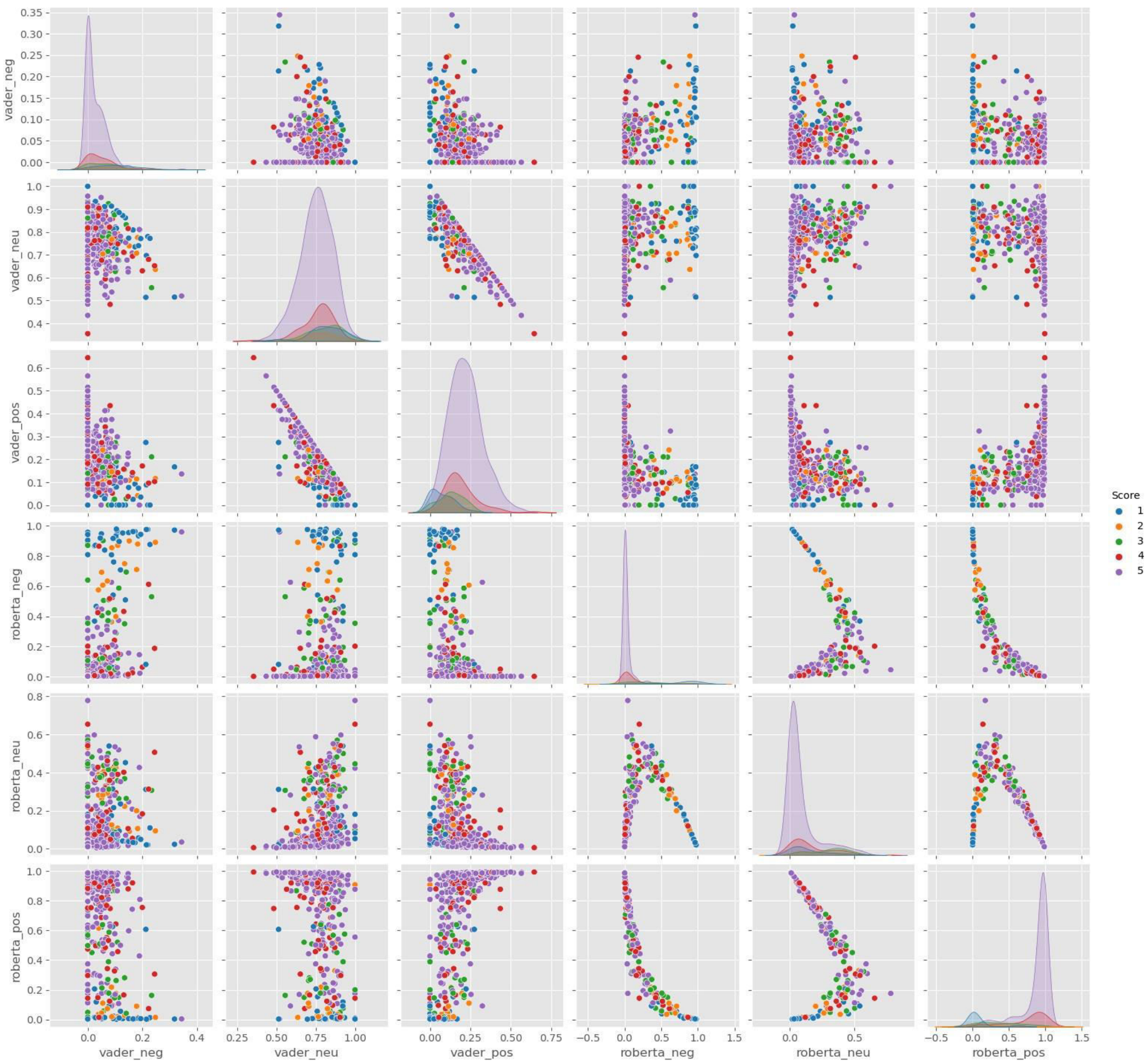


Fig 3 Comparison pairplot between VADER and RoBERTa scores

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

VI. CONCLUSION AND FUTURE WORK

To reveal the shortcomings of both the models, some examples that has lowest rating but were given maximum positive score and has highest rating but were given minimum negative score are shown below. These statements show that either models are getting confused or people has by mistake given wrong rating. In any case these are the showstoppers.

	VADER	RoBERTa
One star rated statements with highest positive score	So we cancelled the order. It was cancelled without any problem. That is a positive note...	I felt energized within five minutes, but it lasted for about 45 minutes. I paid $3.99 for this drink. I could have just drunk a cup of coffee and saved my money.
Five star rated statements with highest negative score	this was sooooo deliscious but too bad i ate em too fast and gained 2 pds! my fault	this was sooooo deliscious but too bad i ate em too fast and gained 2 pds! my fault

The paper has successfully two very extreme techniques of NLP based sentiment analysis. VADER which is a very basic, inaccurate and obsolete technique of sentiment/opinion analysis and RoBERTa which is a state-of-the-art and modern analysis technique. Using the Hugging Face Transformer library tokenization and model training has been done. Machine was pretrained with Twitter data but worked successfully in predicting Amazon reviews. A dataset of around 5 lacs was used to train the model. RoBERTa model was able to accurately predict the sentiment behind most of the comments.

The readers are recommended to train a BERT model that is older than RoBERTa and check its accuracy and performance. While both the models use the same underlying architecture, some people still favour the traditional BERT technique.

REFERENCES

[1] Tetsuya Nasukawa and Jeonghee Yi, “Sentiment analysis: capturing favorability using natural language processing”, In Proceedings of the 2nd international conference on Knowledge capture (K-CAP '03), Association for Computing Machinery, 70–77, 2003.

[2] Dey, Lopamudra, et al. "Sentiment analysis of review datasets using naive bayes and k-nn classifier." arXiv preprint arXiv:1610.09982 (2016).

[3] Aliman, George B., et al. "Sentiment Analysis using Logistic Regression." Journal of Computational Innovations and Engineering Applications JULY 35 (2022): 40.

[4] X. Ouyang, P. Zhou, C. H. Li and L. Liu, "Sentiment Analysis Using Convolutional Neural Network," 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, 2015, pp. 2359-2364

[5] Murthy, G. S. N., Shanmukha Rao Allu, Bhargavi Andhavarapu, Mounika Bagadi, and Mounika Belusonti. "Text based sentiment analysis using LSTM." Int. J. Eng. Res. Tech. Res 9, no. 05 (2020).

[6] Kurniasari, Lilis, and Arif Setyanto. "Sentiment analysis using recurrent neural network." In Journal of Physics: Conference Series, vol. 1471, no. 1, p. 012018. IOP Publishing, 2020.

[7] Santur, Yunus. "Sentiment analysis based on gated recurrent unit." 2019 International Artificial Intelligence and Data Processing Symposium (IDAP). IEEE, 2019.

[8] Hoang, Mickel, Oskar Alija Bihorac, and Jacobo Rouces. "Aspect-based sentiment analysis using BERT" Proceedings of the 22nd nordic conference on computational linguistics. 2019.

[9] Ahmad, Munir, Shabib Aftab, and Iftikhar Ali. "Sentiment analysis of tweets using svm." Int. J. Comput. Appl 177.5 (2017): 25-29.

[10] Elbagir, Shihab, and Jing Yang. "Twitter sentiment analysis using natural language toolkit and VADER sentiment." Proceedings of the international multiconference of engineers and computer scientists. Vol. 122. 2019.