International Journal of Technology, Science and Engineering ISSN: 2457-1016

Vol. 4, Issue I, Jan-Mar, 2021 http://www.bharatpublication.com/journal-detail.php?jID=25/IJTSE

EMPLOYABILITY OF THE SENTIMENT ANALYSIS IN THE DEVELOPMENT OF CUSTOMER REVIEW CLASSIFICATION TECHNIQUES

Ariz Abbas Naqvi

Department of Liberal Arts, Aligarh Muslim University, Aligarh, India

ABSTRACT

There is no doubt that the popularity of online shopping has increased in recent years as a result of the rapid expansion of e-commerce websites. The sellers are attracted to it as well. However, sellers can differentiate based on both quantity and quality. Some sellers offer high-end products to cater to the requirements of their customers, while others concentrate on meeting those requirements at reasonable prices. Customer reviews on e-commerce websites are a smart way for customers to get what they want, making it easier for them to choose a product that meets their requirements. In today's world, customer reviews can significantly affect product sales.

INTRODUCTION

A natural language processing (NLP) method known as sentiment analysis or opinion mining is used to determine whether data is positive, negative, or neutral. Text data is frequently subjected to sentiment analysis, which enables businesses to monitor brand and product sentiment in customer feedback and comprehend customer requirements.

There are three methods for sentiment analysis:

Rule-based: Based on a set of rules that have been manually created, these systems carry out sentiment analysis automatically.

Typically, rule-based systems make use of a set of artificial rules that assist in determining subjects of subjectivity, polarity, or opinion. Because it doesn't take into account how words are arranged, the rule-based system is very simple. Obviously, you can support new expressions and vocabularies by employing more advanced processing methods or adding new rules. However, introducing new rules can have a significant impact on the system as a whole and alter previous outcomes. Rule-based systems frequently necessitate fine- tuning and upkeep, necessitating ongoing investment Automatic Systems: To learn from data, these systems use machine learning techniques. Automated methods, in contrast to rule-based systems, do not rely on rules that are created by hand but rather on machine learning techniques. Typically, tasks involving sentiment analysis are modelled after classification problems in which text is entered into a classifier and categories are returned.

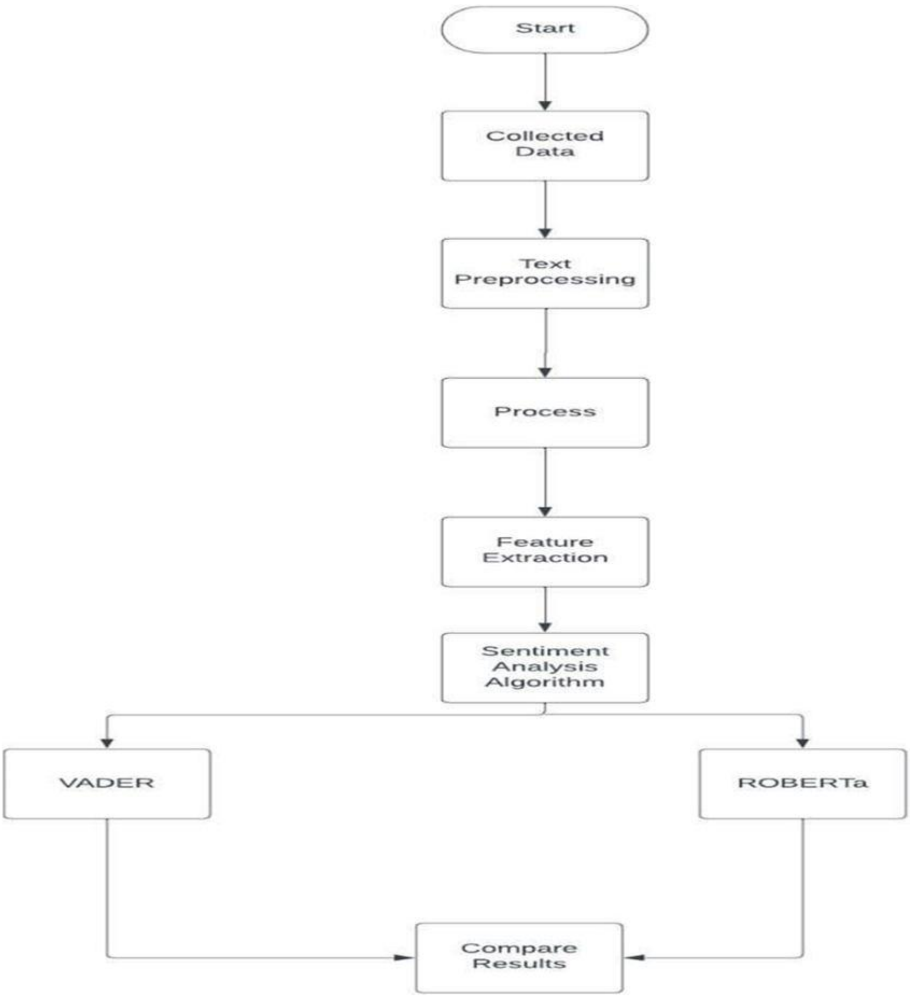
B. Either positive or negative

Based on the test patterns used for training, the model learns to associate particular inputs (text) with corresponding outputs (tags) in the training process (a). A text input is transformed into a feature vector by a feature extractor. pairs of positive, negative, or neutral feature vectors and tags are fed into an algorithm for machine learning in order to create a model. A feature extractor is used in prediction process (b) to turn the unseen text input into feature vectors. The model is then fed these feature vectors to generate predicted tags (once more positive, negative, or neutral). Algorithms for Classification: A statistical model like Naive Bayes, Logistic Regression, Support Vector Machines, or Neural Networks are typically used in the classification step:

1) Naïve Bayes: A group of probabilistic algorithms that predict a text's category using Bayes's Theorem.

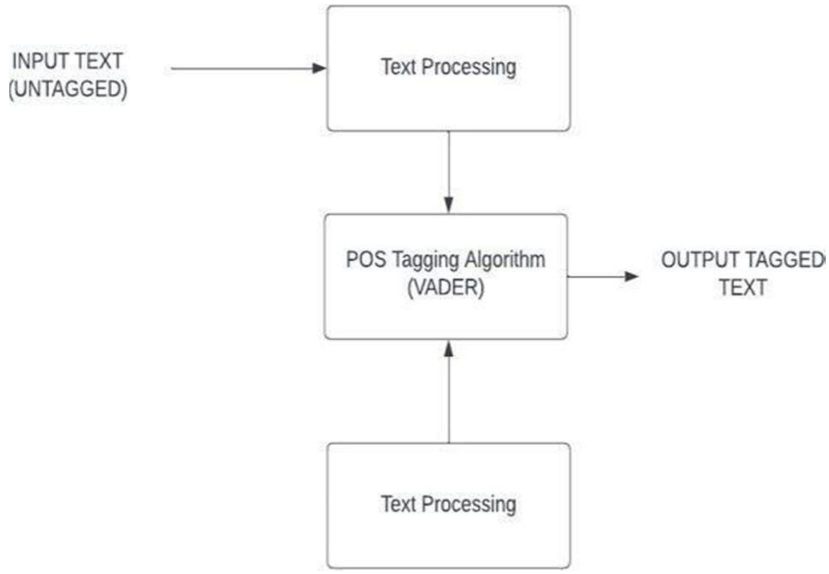
2) Linear Regression: A well-known algorithm in statistics that uses a set of features (X) to predict a value (Y).

PROPOSED METHODOLOGY

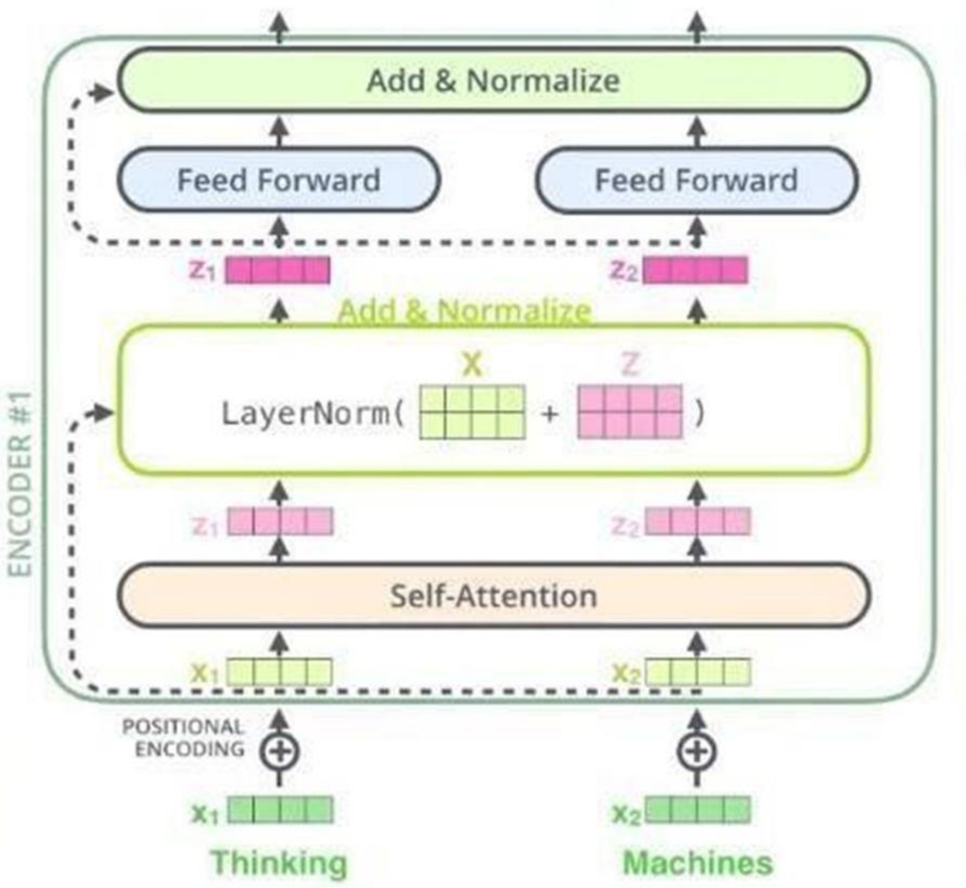


The workflow of the proposed system and how it will operate are depicted in the figure above. We will collect the data and implement two distinct models:

1) VADER Model



Roberta.



These models are distinct from one another.

IMPLEMENTATION

A. Collection of Data

The dataset we used, in this case, consists of all of Amazon's reviews of food products.

B. Libraries and Modules used

- NLTK
- Seaborn
- Tokenizers

C. Implemented Model

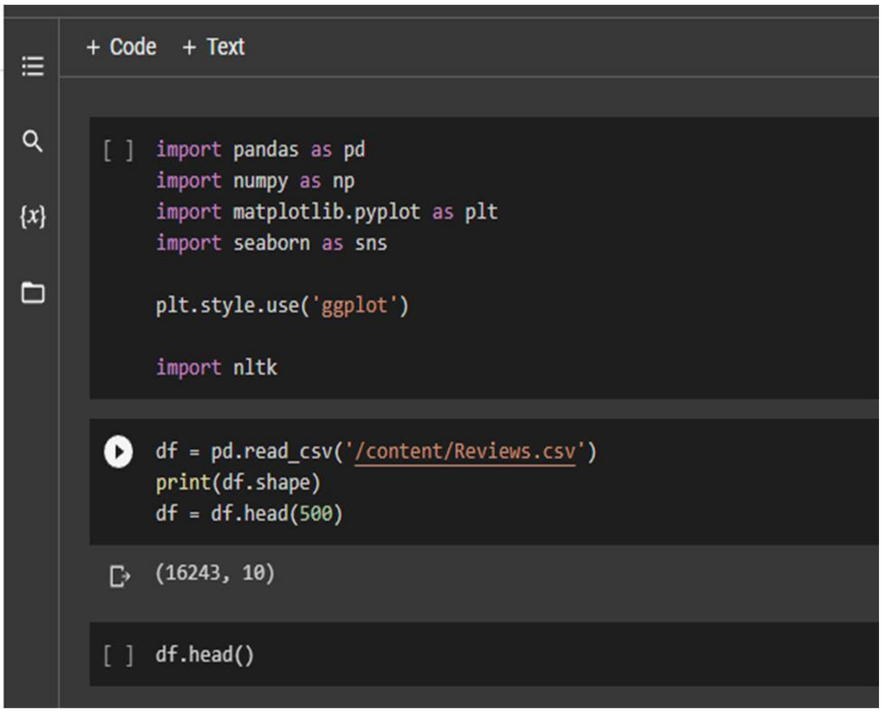
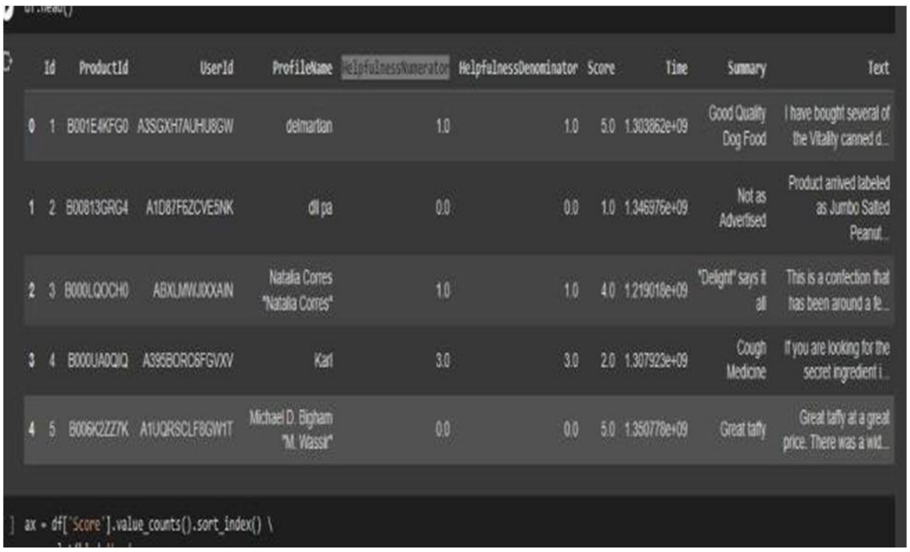


Figure: Loading the Data Set



df.head()

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1.0	1.0	5.0	1.303862e+09	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0.0	0.0	1.0	1.346976e+09	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1.0	1.0	4.0	1.219018e+09	"Delight" says it all	This is a confection that has been around a fe...
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3.0	3.0	2.0	1.307923e+09	Cough Medicine	If you are looking for the secret ingredient i...
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0.0	0.0	5.0	1.350778e+09	Great taffy	Great taffy at a great price. There was a wid...

] ax = df['Score'].value_counts().sort_index() \

Figure: Overview of Dataset

D. Programs Implemented in the Model

1) To get Polarity Score of the Whole Dataset

```
res = {} for i, row in tqdm(df.iterrows(), total=len(df)):

text = row['Text'] myid = row['Id']

res[myid] = sia.polarity_scores(text)
```

2) To Plot the Results from VADER model

```
ax = sns.barplot(data=vaders, x='Score', y='compound') ax.set_title('Compund Score by Amazon Star Review')

plt.show()
```

3) Running Roberta Model

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

output = model(**encoded_text) scores = output[0][0].detach().numpy()

scores = softmax(scores)

scores_dict = { 'roberta_neg' : scores[0], 'roberta_neu' : scores[1], 'roberta_pos' : scores[2]

}

print(scores_dict)
```

4) Combining and Comparing both the results

```
sns.pairplot(data=results_df, vars=['vader_neg', 'vader_neu', 'vader_pos', 'roberta_neg', 'roberta_neu', 'roberta_pos'],

hue='Score', palette='tab10')

plt.show()
```

The score, review time, profile name, product, User ID, text, review summary, helpfulness numerator, and helpfulness denominator are just a few parameters in this dataset.

The unique identifier for the item is the Id Product ID. The user's unique identification number is their ProfileName. Helpfulness Numerator: The number of people who thought the review was helpful (helpfulnessDenominator). Ratings range from one to five. Time, or the review's date and time. A summary of the evaluation Text, also known as the review's text.

TESTING

A. Positive Reviews and Polarity Scores: "I am so happy I ordered it."
As we can see, the review is about a satisfied customer who is pleased with this product, so this is a positive review. The comment here is, "I am so happy I ordered it."
Now, if we look at the polarity scores, Negative (neg) – 0.0; in this case, there are no negative words or feelings; As a result, the negative score is 0. The positive score is good because the words suggest that the review is positive.
Compound Score – 0.646 This is the one that ultimately matters, so it typically ranges from -1 to 1, with -1 representing the most negative and 1 representing the most positive.
The score, as shown here, is 0.646; Therefore, the review is favourable.
B. Polarity Scores for negative reviews:
Commentary: "This is bad. I wouldn't say I liked it.

The comment is here. The viewpoint concerns a dissatisfied customer with this product, so I don't say I liked it. As a result, this review is negative. This could be better.
Now, if we look at the negative polarity score (neg): 0.658, this indicates that a customer is dissatisfied with the product, which indicates that the emotion is negative. As a result, this value will be negative.
Positive(pos) – 0.0 There are no words or feelings of positivity; As a result, the neg score is 0
Compound Score – 0.8271 This is the most important factor overall, so it typically ranges from -1 to 1, with -1 representing the lowest and 1 representing the highest.
This indicates that the score is -0.8271; The review is, therefore, negative.

CONCLUSIONS

The value of sentiment analysis cannot be overstated, despite the numerous difficulties and potential issues facing the sector. Sentiment analysis is destined to become one of the most important factors in many future business decisions because it bases its findings on fundamentally compassionate factors. Text mining methods that are more accurate and consistent can address a number of issues with sentiment analysis that currently exist. Sentiment analysis will be used to build a genuine social democracy that relies on the collective wisdom of the people rather than a select group of "experts." a democracy where all viewpoints and feelings are taken into account when making decisions. As a preventative measure, sentiment analysis's findings are useful. It cannot be used to predict a company's success or measure anything else. Sentiment analysis can sometimes be unnecessary and only useful as a reporting metric after the damage has already been done. Uneven results based on the sources: In sentiment analysis, data separation can be a significant obstacle. Situational analysis that relies on insufficient data can lead to biased results. Sources like Twitter and Facebook can be mined for complete data.

REFERENCES

[1] Abdul-Mageed, M., M.T. Diab, and M. Korayem, (2011), Subjectivity and sentiment analysis of modern standard Arabic. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: shortpapers*

[2] Akkaya, C., J. Wiebe, R. Mihalcea, (2011), Subjectivity word sense disambiguation. *In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP- 2009).*

[3] Alm, C.O. (2011), Subjective natural language problems: motivations, applications, characterizations, and implications. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: shortpapers (ACL-2011).*

[4] Andreevskaia, A. and S. Bergler. (2006), Mining WordNet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. *In Proceedings of Conference of the European Chapter of the Association for Computational Linguistics (EACL-06)*, 2006.

[5] Abdul-Mageed, M., M.T. Diab, and M. Korayem. (2011), Subjectivity and sentiment analysis of modern standard Arabic. *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: shortpapers.*

[6] Akkaya, C., J. Wiebe, and R. Mihalcea. (2009), Subjectivity word sense disambiguation. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP-2009).*

[7] Li, F., C. Han, M. Huang, X. Zhu, Y.J. Xia, S. Zhang, and H. Yu. (2010), Structure-aware review mining and summarization. *In Proceedings of the 23rd International Conference on Computational Linguistics (Coling-2010).*

[8] Li, S., C.-R. Huang, G. Zhou, and S.Y.M. Lee. (2010), Employing Personal/Impersonal Views in Supervised and Semi-Supervised Sentiment Classification. *In Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010).*

[9] Alexander Pak and Patrick Paroubek. 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA)

[10] Dave, Kushal & Lawrence, Steve & Pennock, David. (2003). Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. *Proceedings of the 12th international conference on World Wide Web* 775152. 10.1145/775152.775226.

[11] B. Aisha, (2010), A Letter Tagging Approach to Uyghur Tokenization, *International Conference on Asian Language Processing*, pp. 11-14, doi:10.1109/IALP.2010.72.