Leveraging NLP to efficiently accommodate Sentiment Analysis on Textual Data

Spandan Sanket Maharana

Abstract- Text Sentiment Analysis (SA) has become an increasingly popular method for organizations to understand the perceptions and opinions of their employees. It is the study of opinions, sentiments, attitudes, emotions, mood of a conveyor expressed in written language. SA involves using natural language processing (NLP) techniques to identify the emotional tone expressed in text. This can be used to determine the sentiment of employee feedback, which can be used to identify areas that require improvement or to highlight successful practices. The aim of this project is to present a detailed analysis of the sentiments of a review for the employee provided by their peers and managers using the Bidirectional Encoder Representations from Transformers (BERT) model, a language model that has proven to be highly effective in natural language processing tasks such as sentiment analysis, text summarization, machine translation, etc. With the help of BERT, we aim to identify the sentiment expressed in employee reviews and understand the factors contributing to the positive or negative sentiments. The dataset used in this study consists of employee reviews from various industries and organizations which comprises both negative and positive reviews. The first step in this project is to perform data pre-processing on the dataset and to remove any redundant data points. This entails cleaning the data, removing stop words and punctuation, and converting the text to a numerical format that the BERT model can process. Following that, the dataset will be divided into three parts: training, validation, and test. The goal is to fine-tune the pretrained BERT model on the training set using a binary classification task to predict whether a review is positive or negative, and then to evaluate the model on the validation and test datasets to determine performance. The study's conclusions are expected to have a wide range of effects on organizations. The analysis can primarily be used to pinpoint the aspects of the working environment that need to be improved. It can then be utilized to pinpoint productive procedures in the workplace. Finally, the model can offer information that might be useful in identifying the elements that contribute to employee retention and satisfaction across all organizational departments. In summary, this study shows how well the BERT model does sentiment analysis on employee reviews. The programme was able to correctly categorize the tone used in employee reviews and pinpoint crucial factors that can influence whether a review is positive or negative.

Keywords: Text Sentiment Analysis, Employee Review, NLP, BERT model, Dataset

1. INTRODUCTION

Natural Language Processing (NLP) is a branch of computer science, more specifically the branch of Artificial Intelligence (AI) which is concerned with giving computing ability to the computers which can clarify the context behind the spoken words and written texts by using various machine learning and artificial intelligence algorithms. It enables the computers to understand the natural language as the humans do, whether the language is in a written or spoken format and process it to make sense in a way the computer can understand. It has applicability in real-world scenarios, such as analysis of chats, customer reviews, employee review, sentiment analysis of the text. The data collected in the above format can be processed in using the machine learning algorithms like linear regression, random forest classification, naïve bayes, SVM and many other algorithms which can be helpful in deducing the efficiency of various state-of-the-art models like BERT(Bidirectional Encoder Representations from Transformers), VADER(Valence Aware Dictionary for sEntiment Reasoning), Hugging Face Model, etc. and make a comparative analysis on which model is the most efficient in reading and understanding the sentiments from a series of data segments with an optimal space and time complexity.

The benefits of NLP includes but is not limited to, checking the accuracy and efficiency of a document, ability to summarize a complex data, helpful in the framework of AI such as Google Assistant, Alexa, Siri, intent classification in the chatbots used by organization for customer support and reviews, provides data filtering and clearing which can be helpful in getting an advanced insight into the data from the analytics, and many more. The challenges to NLP are Lack of precision, Unabridged intent of the voice and tone, Evolving use of the language. In order to adapt to the benefits and eliminate the challenges in NLP we are more inclined towards using the modern machine learning and deep learning models which can be helpful in facilitating efficient yet effective results in real-time.

In this project, we have focused on Text Sentiment Analysis, which is a popular application of natural language processing (NLP) which involves running an analysis on a piece of text and categorizing the detected sentiments in a text and the goal of sentiment analysis is to approximate the emotion in the text, that is whether the text is likely to express a positive, negative or neutral sentiment towards a given subject. The results obtained can be helpful in a variety of applications, such as predicting the customers behavior, subjective analysis of the members in a team, monitoring the brand value, employee review analysis, and in various other trending areas.

Bidirectional Encoder Representation from Transformers (BERT) is a pre-trained state-of-the-art deep learning model which provides high efficiency and performance on a wide range of NLP tasks, including but not limited to Text Sentiment Analysis. It is designed on the transformer architecture, that allows it to seize a long array of dependent relations within the words in the sentences and moreover learn a context-based representation of the written and spoken languages, making it more suited for the areas of task where understanding the context of natural language is a primary concern. Combining BERT with NLTK (Natural Language Toolkit) we tend to get a higher accuracy from the data, by providing various tools for data filtering using tokenization, stemming, POS, Stop words, text categorization, parsing, semantic reasoning, wrappers for industrial strength NLP libraries and an active discussion forum. At first, the model is fine-tuned on the labeled dataset, where we use various algorithms to efficiently train the model for predicting the sentiment in a text based on a label provided in the libraries. Once the model is trained, we have to test the efficiency by providing more data into the program, if the result provided is as effective and accurate as the trained model, we can classify the new obtained result from the data as a predefined sentiment.

In this paper, we explore the integration of BERT and NLTK for Sentiment Analysis in various domains of the industry. We present a comprehensive and comparative analysis between the various models of NLP using the BERT, VADER and Hugging Face model.

2. LITERATURE SURVEY

Text sentiment analysis is a popular field in natural language processing (NLP), which involves determining the polarity of a given text, i.e., whether it is positive, negative, or neutral. BERT (Bidirectional Encoder Representations from Transformers) is a state-of-the-art language model developed by Google, which has achieved remarkable performance on various NLP tasks, like text sentiment analysis.

In the literature survey of this paper, we will discuss a few of the studies which have used BERT for text sentiment analysis.

[1] "BERT for Sentiment Analysis on Large-Scale Datasets" by Y. Yang et al. (2020)

This paper evaluated a BERT-based approach for sentiment analysis on large-scale datasets. The authors have fine-tuned the BERT model on two benchmark datasets, Amazon, Yelp and achieved state-of-the-art results on both datasets.

[2] "Fine-tuning BERT for Sentiment Analysis of Short Texts" by D. Wang et al. (2020)

This paper has put forth a BERT-based approach for sentiment analysis of short texts, such as tweets and reviews. The authors have fine-tuned the BERT model on a dataset of Twitter tweets and achieved better performance than the competent models.

[3] "BERTweet: A pre-trained language model for English Tweets" by V. Sanh et al. (2020)

This paper has given a brief introduction to ‘BERTweet’, a pre-trained BERT model for sentiment analysis of English tweets. The authors have fine-tuned the model over varied pinnacle datasets and achieved state-of-the-art results on most of them.

[4] "Comparative Analysis of BERT and Traditional Machine Learning Techniques for Sentiment Analysis" by S. Venkatraman et al. (2021)

This paper has provided a comparative analysis of the performance of BERT with traditional machine learning techniques for sentiment analysis. The authors have fine-tuned the BERT model on two datasets and compared the performance of the model with various other models. The results obtained justified that BERT outperformed the other models on not just one but both datasets.

[5] "Enhancing the Performance of Sentiment Analysis with BERT Embeddings and Data Augmentation" by H. Zhang et al. (2021)

This paper has proposed a method for tweaking the performance of sentiment analysis using BERT embeddings and various data augmentation techniques. The authors have fine-tuned the BERT model on several benchmark datasets and achieved better results as compared to the other models.

In conclusion, BERT has shown great potential for text sentiment analysis, and recent studies have demonstrated its effectiveness on various datasets and tasks. Fine-tuning the BERT model on specific datasets and applying data augmentation techniques can further improve its performance.

3. RESEARCH METHODOLOGY

There are a wide variety of applications, including social media monitoring, brand reputation management, customer experience analysis, market research, employee review analysis and many more. All the data obtained can be used for extracting the subjective information from the text including attitude, opinions, emotions and feelings. In this paper, we discuss the mostly used methods and techniques to obtain the results.

3.1 Rule based methods

It uses a predefined set of rules to identify the emotions/sentiments in a text. These are apparently based on a more linguistic feature of the text, likely the positive or negative words, use of intensifiers, syntactic structure of sentences. For instance, a sentence containing the words ‘happy’, ‘love’, ‘high’, ‘thanks’, ‘elated’, might be tagged as a positive sentiment whereas the sentences containing the words ‘hate’, ‘sorrow’, ‘unhappy’, ‘disappointed’, will be tagged as negative sentiment. This method is easy to interpret and implement in real-time, but when the complexity of the language increases, the results may tend to be more obnoxious and limited.

3.2 Lexicon based methods

This method uses existing dictionaries of sentiment words and phrases to determine the sentiment in a text. Here, each sentiment is assigned a score, i.e. the positive emotions are assigned a value of +1 and negative emotions are assigned a value of -1, and then the aggregate of all the emotions is done to get a sentiment score. This method is helpful for the domains where the labeled data is scarce, as they do not require training data, making it more limited to coverage and quality of the nuances in the language.

3.3 Hybrid Methods

This is a combination of multiple approaches which can help in improving the accuracy of the data. For example, we can use a rule-based approach to get an overview of the overall sentiment of the text, and can use a machine learning model which can be helpful in classifying the sentiment of specific aspects of the entities present in the text. It can be helpful in mitigating the weakness of using a single method or algorithm, but it is more difficult to implement and requires more resources for computing the data.

3.4 Machine Learning methods

The statistical use of algorithms which are helpful in learning from the data and making predictions can be classified into the machine learning methods, which requires a labeled dataset to perform analysis, with each data annotated with a corresponding sentiment label of positive, negative, neutral. This method is mainly focused on learning to identify the patterns in the data and use them to predict the patterns in the new data by creating models.

3.5 Data Preprocessing

It is a critical step which involves cleaning, normalizing, transforming the raw text into a machine-readable format which is suitable for analysis. The techniques included are tokenization, stemming, stop words, POS. Tokenization is the method of splitting the text into individual words or tokens and then checking for the emotion of each of the words and then estimating an aggregate emotion of the text data, whereas stemming reduces the words to their root form to handle the variations in the endings of the sentences. While using the stop words, we filter out all the words which are not stopping phrases or characters in the sentence and what remains is analyzed to obtain a result from the data. On the other hand, POS involves labeling the words into its grammatical category which can help in identifying the sentiment of the words before and after them.

3.6 Feature Extraction

It is the process of converting the preprocessed text into a set of numerical features which can be used as input to the machine learning algorithms. The techniques like Bag-of-words which uses a list of words containing few specific words in the list, which when found in a sentence, can be splitted from the sentence and be stored in a new list and picked up from that list to give a specific output of the emotion represented by the word.

3.7 BERT Model

Bidirectional Encoder Representation from Transformers (BERT) is a state-of-the-art language model developed by Google in 2018, which uses a deep neural network architecture called Transformers. It is trained on a large amount of textual data which enables it to generate high quality language representations which capture the semantic meaning and intent of the words and sentences. The working of the model involves classifying text into a large category of sentiments, by running algorithms and performing preprocessing of the data over the dataset, which helps in filtering the data by removing the redundant data from the dataset and only focusing on the cleaned and processed data for training and testing purposes and then fine-tuning the pre-trained model on a specific text, involving adjusting the parameters to classify the sentiments in the text, followed by classifying the data obtained after fine tuning into proper function vectors which can be further exploited and changed to get a varied result on a larger dataset, by evaluating the test scores, precision, accuracy on the test dataset.

BERT is a powerful language model which can be tweaked for sentiment analysis by adjusting the parameters to classify the sentiments by adjusting its parameters to classify the sentiments in a textual data, this ability to capture the semantic meaning of the words and sentences makes it an efficient tool for NLP.

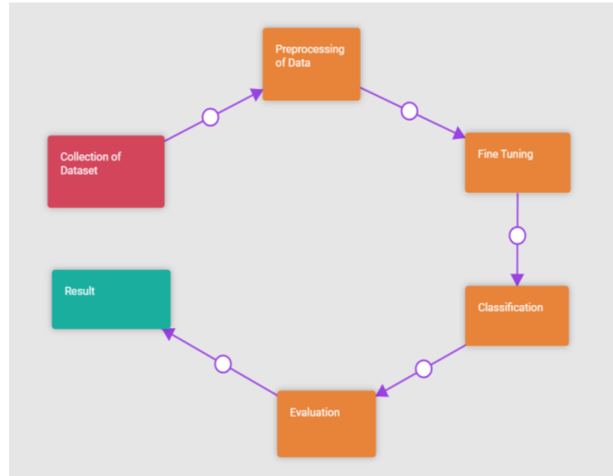

Fig 1- Steps of working in the Bidirectional Encoder Representation from Transformers (BERT) model

3.8 VADER Model

Valence Aware Dictionary and sEntiment Reasoner (VADER) is a rule-based sentiment analysis model which makes use of the lexical analysis method to facilitate text sentiment analysis. It is designed to analyze the sentiments of the various social media texts, which tend to contain coarse language which gets classified as invalid data by other machine learning models and the other algorithms. But, when using VADER, we can train the values and get the result into a more accurate version. This is helpful in mitigating the potential risks for the industries by giving out a detailed analysis at the earliest. The working of the model in performing text sentiment analysis involves preprocessing the data, by cleaning and removing the stop words, punctuations, special characters, by tokenizing the text into individual words and storing them into a list for later use, then scoring each word of the text using a lexical tool consisting of various words and their corresponding sentiment scores(positive, negative, neutral), the scores can range from -1 to +1 where former being the most negative sentiment and latter being the most positive sentiment. Subsequently, we can score the sentences by combining the sentiment scores of each of the words, and then use various rules to adjust the sentiment scores based on the factors like intensifiers, negations, capitalizations, etc. Then, the proposed final output will range from -1 to +1 along with the intensity of the sentiments as well as the scores for the different components of the sentiments and finally, evaluating the performance of VADER model on the test dataset by computing the metrics like accuracy, precision, recall and F1 score.

VADER is a rule-based sentiment analysis method which uses lexicon analysis to score every individual word and compute the sentiment score of the entire textual data. The rules to adjust the sentiment scores are based on various factors, making it effective for analyzing the sentiment of social media texts.

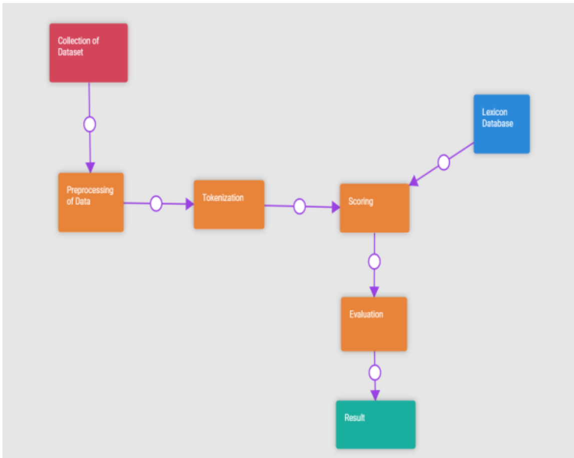

Fig 2- Steps of working in the Valance Aware Dictionary and sEntiment Reasoner (VADER) model

In conclusion, text sentiment analysis is a valuable tool which can be used to extract insights from textual data, the methods mostly depend on the availability of the labeled, unlabeled data, complexity of language, robustness, accuracy and frequency of the characters in the text. The rule-based methods like VADER, are easily interpreted and implemented but can be limited by the specific requirements and parameters of the rules. On the other hand, machine learning methods like BERT provide higher accuracy as the data undergoes a series of various filtering and clearing algorithms which can help in increasing the accuracy of the models. Also, since the data trained in a machine learning method is segregated into train, test and validation data, it provides a specific data value to the data through labeling the data nodes in the dataset. Therefore, we made a comparative analysis between the rule-based methods and machine learning methods, which helped us to ballpark and pinpoint on the efficacies of the entire dataset and facilitate the conclusion on which method is the most efficient to perform text sentiment analysis.

4. RESULT ANALYSIS

Text Sentiment Analysis is an important technique used to analyze and understand the sentiment, tone and emotions expressed in textual data. In this paper we have used the [Amazon Fine Food Reviews dataset]() to perform the analysis on the textual data to categorize the sentiments by applying various models and obtaining the most efficient model to proceed in designing an approach to perform sentiment analysis on the datasets for a large array of data and moreover contributing to the future scope in the domain of NLP.

The dataset consists of reviews of fine foods from amazon. The data span a period of more than 10 years, including all ~500,000 reviews up to October 2012. Reviews include product and user information, ratings, and a plain text review. It also includes reviews from all other Amazon categories. The dataset contains all the reviews from Oct 1999 to Oct 2012 which adds up to approximately 568,454 reviews in total, and the reviews were given by 256,059 users who brought 74,258 products in total and amongst those users around 260 users have given more than 50 reviews. For our analysis we have used around 10,000 reviews.

Here we have implemented the testing of the data using Valence Aware Dictionary and sEntiment Reasoner (VADER) model and Bidirectional Encoder Representation from Transformers (BERT) model.

VADER uses a rule-based method which uses a sentiment lexicon to assign sentiment scores to each word in the text. The scores which are assigned to the tokenized texts are segmented as polarity scores and the summation of all the polarity scores in the tokenized list, gives the polarity score for the entire text. The polarity score is a float value ranging from -1 to +1 where the former being the most negative sentiment in a text and the latter being the most positive sentiment in the text. The results obtained using VADER were highly accurate with the polarity scores ranging from 0.7 to 0.8, it was particularly effective in identifying the intensity of sentiment and the context embedded behind the text.

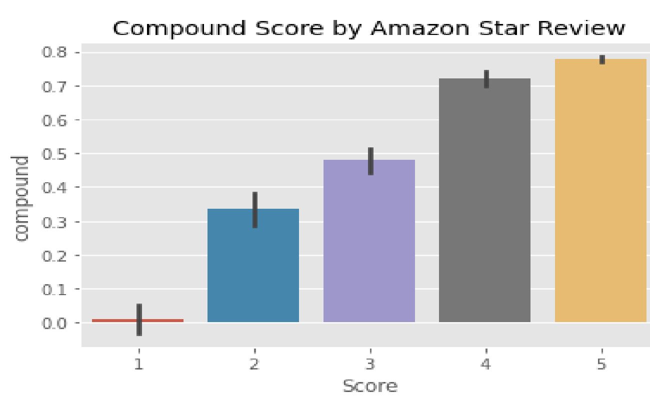

Fig 3- Compound Scores generated using VADER on the dataset.

The compound scoring technique is used to generate the leveraging score for the sentiment in a text, it is faster, efficient and did not require large amounts of data for training and tuning using multiple parameters, but the coverage of the sentiment was limited to a lexical analysis which at certain data points made it difficult to get the sentiment behind the texts, also in the cases when the complexity of the language increased the model produced ambiguous results which led to decrease in the efficiency of the model. On testing the model on the dataset, we have graphically presented the sentiments from the data.

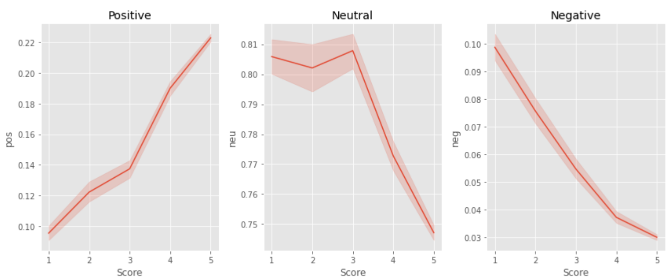

Fig 4- Polarity Scores generated by the VADER model.

After using both the models on the same dataset, we found both were highly accurate with polarity scores ranging from 0.7 to 0.9. However, Bidirectional Encoder Representation from Transformers (BERT) was highly effective in identifying the sentiments in the text, even the subtle sentences and the sentences including irony, sarcasm, pun, etc. while VADER identified the intensity of the sentiment and the context of the text more accurately. However, BERT required a large amount of data and hence was computationally expensive and was also more versatile and can support multiple parameters while performing NLP tasks as compared to VADER which was faster and did not require large amounts of data, but had limited coverage of sentiment lexicons.

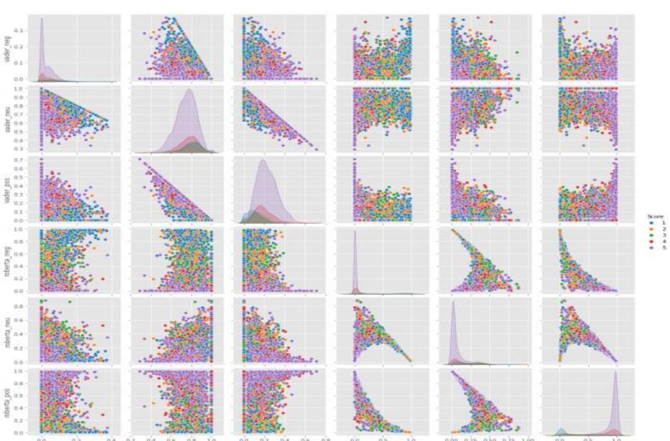

Fig 5- Comparative Analysis between BERT and VADER

In conclusion, Bidirectional Encoder Representation from Transformers (BERT) and Valence Aware Dictionary and sEntiment Reasoner (VADER) both are highly effective tools for sentiment analysis, but BERT is more effective in identifying the subtle overtones in the sentiments and is more versatile but requires a large amount of data while VADER is faster and efficiently identifies the intensity of sentiment and the context of the text, but has a limited coverage and at certain data points overgeneralizes the sentiment of the text. The choice of the model depends on specific application and requirements for accuracy, efficiency and coverage.

5. ADVANTAGES AND LIMITATIONS

Sentiment Analysis has a wide range of advantages for business including valuable insights into customer opinions and feedback which helps businesses grow and prosper by understanding how their products are being perceived in the market and this information can be used to make improvements and strategies according to the needs of the customer to enhance their experience, proactive reputation management which is helpful to ballpark the online reputation of any company by identifying and addressing negative sentiment early which can mitigate the potential damage of their brand value, real-time social media monitoring which allows the businesses to respond quickly to the customer feedbacks and enquiries, which helps improve customer engagement and brand perception on social media , informed market research which provides an array of varied opinions and a detailed analysis on the data, risk management can be done on the potential risks such as negative customer review or reputational damage which can be helpful, and competitive analysis which is used to analyze sentiment towards competitors by understanding how the product is actually being perceived by the consumers, automation, objectivity, scalability, and many more.

Even though there are many advantages of sentiment analysis, it also has potential disadvantages such as subjectivity and ambiguity because it relies on interpretation of human emotions which can be subjective as different people may have a different way of interpreting emotions and thus sentiment analysis can lead to potential minister[pretation, there can be inaccuracies in prediction because while dealing with complex or subtle emotions, figurative language or sarcasm false positives and false negatives can impact the reliability of the final result. Some models may struggle with understanding the contextual information because the sentiment will vary according to the context and lack of context can lead to inaccurate results for example, “It’s not that I don't like it, I hate it" can be a challenging sentence for some models to interpret the true emotions correctly because of the sarcasm involved. sentiment expression, tone and cultural norms normally vary across different languages and places and sentiment analysis models may not be equally effective in all cultural contexts and languages. Some models can be biased, and this can result in inaccurate predictions. Sentiment analysis on social media can pose challenges due to informal language used, acronyms, and several other evolving language trends. It is important for us to be aware of these limitations and use the results obtained judiciously and consider all the biases of the model that is being used.

The advantages of using BERT for text sentiment analysis is that it is susceptible to capturing the varied intent of the natural language, likely irony, anger, sadness, sarcasm, happiness, excitement, etc. Additionally, there is a scope to fine-tune the model on specific domains of data, which can effectively improve the performance and can be applied to specific domains and industry, facilitating an ease of access and effective conglomeration of data trends with the day-to-day needs of the industry. BERT has been pre trained in many datasets in various languages. which makes it more effective when it comes to analyzing sentiments in different languages, making BERT very useful for businesses that operate in multiple countries as they need to monitor emotions in different languages. As BERT is a pre-trained model, it is an effective tool for transfer learning in sentiment analysis.

The limitations of using BERT for text sentiment analysis is that there is a contextual ambiguity in identifying the correct sentiment of words or phrases with multiple meanings or connotations. Additionally, the domains where the vocabulary and language usage is highly specific, there may not be specific trained data to facilitate proper working of the model, and the requirement of large amount of labeled data for training to perform well and the datasets with less data points results in overfitting or underfitting of the data. Moreover, it requires higher GPU and CPU usage for computation, making it challenging for many users with limited resources.

The advantages of using VADER for text sentiment analysis is that it is an open source tool accessible to everyone and is fast and efficient to implement and does not require large dataset for training the model, along with a predefined lexicon of words which have been annotated with the sentiment score which can handle sarcastic remarks and identify the emotions in the emoji's used in text, even can handle the negations which can change the sentiment of the text whilst providing a high accuracy in text analysis, especially in short texts like the social media comments and posts.

The limitations of using VADER for text sentiment analysis is that there is a limited vocabulary scope i.e, the performance is limited by the dictionary in the model which may not include all the words in the text, leading to a lower accuracy while performing data analysis. Moreover, the model follows a rule-based model which mainly relies on the predefined rules and cannot be customized to the specific needs of an application and cannot capture the complexity and the overtone of the language always, resulting in errors and due to its language specificity and limited scope which can subsequently add to loss of insights into the final analysis.

6. CONCLUSION

The set of conducted experiments showed that BERT outperformed the state-of-the-art machine learning model VADER with a basic configuration tweak for the optimal training speed in the dataset. It showed a promising result and is now considered as a popular approach in NLP tasks, and its ability to capture the context and meanings of the words have demonstrated the improving accuracy in the longer texts. Regardless of the limitations, it is a powerful tool which can be used in a variety of applications like social media monitoring, customer feedback analysis, market research, mitigation of negative reviews in a website, and many more.

ACKNOWLEDGEMENT

It is a matter of great pleasure for us to get this opportunity expressing our sincere sense of gratitude to Siksha 'O' Anusandhan Deemed to be University. Firstly, we would like to express our heartfelt thanks to the Institute of Technical Education and Research for providing an opportunity and other relevant facilities. Our guide Daisy Monika Lal was the main force behind all these efforts because of her valuable suggestions and proper guidance for this project. We express our sincere thanks to the HOD, Department of Computer Science and Engineering, Prof. (Dr.) Debahuti Mishra who has allowed us to use the facilities of the institute. We are also thankful to all those who have helped us in this endeavor either directly or indirectly, especially all our teachers. Finally, we would like to express a big thank you to all friends and all known & unknown people who had helped us directly or indirectly.

REFERENCES

[1] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In Proceedings of the Workshop on Language in Social Media (LSM 2011), pages 30–38, Portland, Oregon. Association for Computational Linguistics.

[2] Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models.

[3] Esteban Castillo, Ofelia Cervantes, Darnes Vilari˜no,David B´aez, and Alfredo S´anchez. 2015. UDLAP: Sentiment analysis using a graph-based representation. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pages 556–560, Denver, Colorado. Association for Computational Linguistics.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. CoRR, abs/1810.04805

[5] Mickel Hoang, Oskar Alija Bihorac, and Jacobo Rouces. 2019. Aspect-based sentiment analysis using BERT. In Proceedings of the 22nd Nordic Conference on Computational Linguistics, pages 187–196, Turku, Finland. Link¨oping University Electronic Press

[6] Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 3730–3740, Hong Kong, China. Association for Computational Linguistics.

[7] J. McAuley and J. Leskovec. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. WWW, 2013.