INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER
INDIA

Impact Factor: 8.379

Real-Time Sentiment Analysis System Using the BERT Model

Ayush Kumar, Dr. Kamalraj R

PG Student, Department of CS and IT, Jain (Deemed-to-be University), Bengaluru, India

Professor, Department of CS and IT, Jain (Deemed-to-be University), Bengaluru, India

ABSTRACT: Opinion mining or we can say the Sentimental analysis is a critical issue where a large amount of information related to public opinion is widely spread in the world. So, in this paper, we will try to understand the public opinion and decision-making process in real time for the given text data. This proposed system aims to develop a real-time sentiment analysis of text data from various sources, such as social media, news websites, etc. The system will involve several processes like data collection, data preprocessing, sentiment analysis using the BERT (Bidirectional Encoder Representation from Transformers) model, and real-time processing to find the sentiment trends and patterns. based on [4] we can say that the performance of the BERT model is highly accurate as compared to the other traditional models like Random Forest, KNN, and Decision Tree, etc. The result which is obtained using the BERT model shows the accuracy of the system.

KEYWORDS: Sentiment analysis; Social media; BERT; Deep learning; Subgroup discovery.

I. INTRODUCTION

In the present time the one of the most important research topics is sentiment analysis or we can say opinion mining. This topic mainly refers to the study of people's emotions, Opinions, or sentiments that can be expressed in a written text. We will use the pre-trained language model like BERT for sentiment analysis in various domains such as media analysis, market research, and decision-making processes.

Companies and organizations are trying to find public opinion about their products and services based on real-time data analysis and can make better decisions based on the system output. We can use different platforms like Twitter, Facebook, Threads, etc. to extract and analyze the user's opinions and reviews that they have posted. The information amount that is extracted from the above platforms is too large to be analyzed by normal users. To avoid this, we will use the sentiment analysis technique we will use the pre-trained language model like BERT which was developed by Google in 2018.

In the paper, I have discussed real-time data analysis by following processes like data extraction from the online platforms and then we did the data preprocessing to remove all the unwanted text like #, URL, @, etc. Then we will do the tokenization and create tokens from the following text data.

The paper introduces the methodology for real-time sentiment analysis using deep learning models like BRET to analyze large data and find the sentiment or opinion of the users that are present on platforms like Twitter, Facebook, Threads, etc.

II. LITERATURE STUDY

The sentiment analysis also known as opinion mining comes under the domain of natural language processing (NLP), which focuses on the extraction of information from the given textual data. A lot of research done on this topic with different aspects and the deep learning method has transformed sentiment analysis by increasing the performance across a range of different tasks. Based on the study we can say that the utilization of BERT (Bidirectional Encoder Representations from Transformers) and other deep learning architectures is performing very well in this field.

In [1], the author presents a methodology where he used BERT for analyzing sentiment in social media content. To achieve high performance in sentiment analysis he combines the BERT and Bidirectional LSTM (Long Short-Term Memory) and addresses challenges such as bias reduction and efficient representation learning.

In [2], the author merges CNN (Convolutional Neural Network) and Bi-LSTM (Bidirectional Long Short-Term Memory) architectures for sentiment analysis. This model captures sentiment expressions and demonstrates promising results in sentiment classification tasks.

In [3], the author used Twitter data to analyze climate action and the Sustainable Development Goals (SDGs). The study covers preprocessing techniques, statistical methods, and deep learning models like BERT to analyze sentiments and provide valuable insight into public opinion dynamics and policy development in the field of climate action.

In [4], a methodology where BERT with conventional machine learning techniques for sentiment analysis of airline reviews is used. The study compares the performance of BERT-based models to traditional methods and highlights their efficiency and effectiveness in sentiment analysis tasks.

In [5], the author focused on improving transparency in the decision-making process by offering easy to understandable explanations for crowd decisions. The method leveraging deep learning based on sentiment analysis offers insights into the reasoning behind crowd decision-making in various scenarios.

In [6], researchers introduced a method where he used the Natural Language Processing (NLP) technique for extracting tweet polarity. By using AdaBoost classifier to enhance the performance and process reviews from various E-commerce sites.

In [7], the author proposed a model that can perform real-time sentiment analysis on product reviews. Providing real-time analysis to the users in the E-commerce application. This innovative approach allows users to access the sentiment analysis results and provide valuable insight into consumer sentiment.

III. PROPOSED TECHNIQUE

The proposed technique aims to create a model by training the present BERT model in such a way so that it can perform real-time sentiment analysis on the given text data which is obtained by social media like Twitter, Facebook, Thread, etc. The process that we are going to follow to training the model will be divided into components as shown in Fig 1.

The data extraction from the online platform then it will do the preprocessing to eliminate the unwanted text elements like hashtags, URLs, and user mentions and then it will perform the Tokenization to create tokens from the text data. The model will undergo training using deep learning techniques. By using pretrained language model like BERT we can efficiently analyze the sentiment by capturing the semantic associated with the text data.

Deploy the trained BERT model to perform the real-time sentiment analysis on incoming text data. Sentiment classification has three categories like Positive, Negative, and Neutral, and based on this we will dynamically update the sentiment analysis.
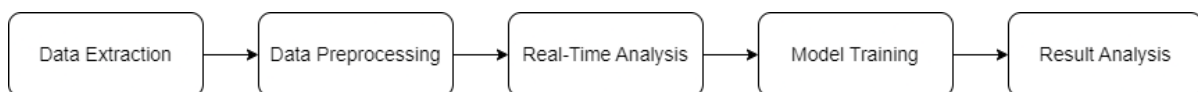

Fig 1. Process to train the model

IV. METHODOLOGY

The main objective of this paper is to leverage deep learning model like BERT to enable real-time sentiment analysis and categorize the given data into the sentimental classifications like Positive, Negative, and Neutral.

A. *Data Extraction:*
We will gather text data from various sources such as social media platforms (e.g., Twitter, Facebook, Threads), news websites, forums, blogs and any other relevant platforms where public operation are expressed. To do this we can use APIs or web scraping techniques to collect textual data in real-time as show in Fig 2 .

B. *Data Preprocessing:*

Clean the collected data by eliminating unwanted text element including hashtags, URLs, user mentions, special characters, emojis, and any other irrelevant information that may interfere with sentiment analysis. Now applying the normalization technique, we will standardize the text data by converting it to lowercase, removing all punctuation, and handling contractions, abbreviations, and slang terms to ensure consistency. Now we will do the tokenization by segmenting the preprocessed text data into individual tokens (words or characters or sub-words) to facilitate further analysis as show in Fig 3.

C. *Model Training:*

The bidirectional Encoder Representation from Transformers (BERT) model is selected because it is effective in capturing semantic within textual data based on [4]. Now we will pretrain the BERT model on large corpora so that it can learn the representation of words and phrases. Then we will see the output and try to fine tune the pre-trained BERT model using the collected and preprocessed text data from the domain of interest (e.g., social media, news) to adapt it for the real-time sentiment analysis tasks as show in Fig 4.

D. *Real-Time Sentiment Analysis:*

After training the BERT model we will deploy it to the production environment which is capable of handling real-time data streams. Now the deployed BERT model can classify the sentiments of the incoming text data into predefined categories such as Positive, Negative, and Neutral. The continuously update the sentiment analysis results in real-time as new data arrives, we can see the change in trends and patterns live as show in Fig 5.

E. *Result Analysis:*

At the end we will analyze the sentiment result to get the better understanding of public opinion dynamics, market trends, customer sentiment, or any other relevant area of interest. We will represent the result by some visualizations technique such as charts, graphs, and dashboards to facilitate easy interpretation and decision-making.
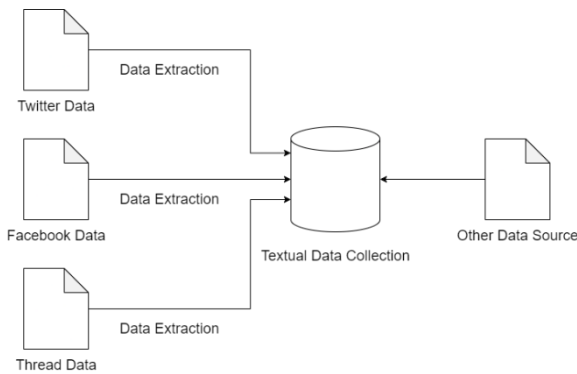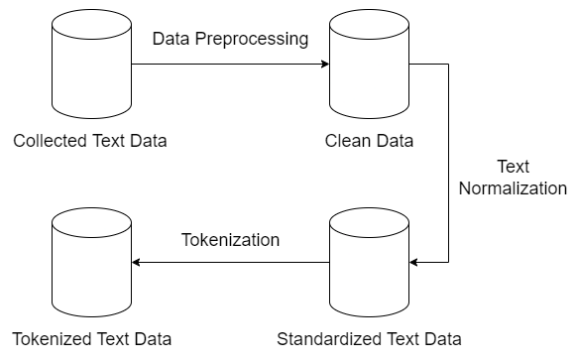
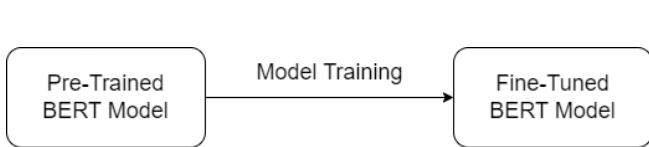Fig 2. Data Extraction.
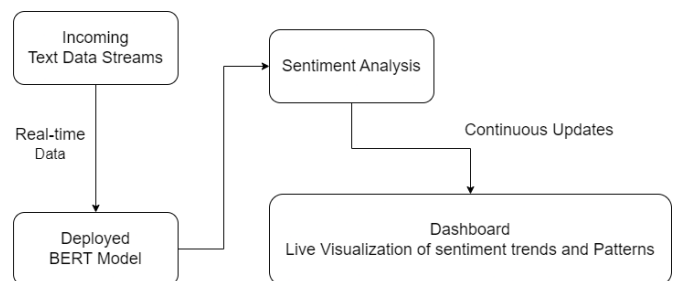
Fig 3. Data Preprocessing.

Fig 4. Model Training.

Fig 5. Real-Time Sentiment Analysis.

V. RESULT AND DISCUSSIONS

F. *Sentiment Analysis Performance:*
The real-time sentiment analysis system which uses the BERT model work very well in analyzing sentiments from textual data. We can check how accurate and fast it was by looking at different measures, and it did a great job at handling lots of text data from online platforms like social media, news, and online forums.

G. *Comparison with Traditional Models:*
When we compared our system with traditional model like Random Forest, KNN, and Decision Tree, we found that the BERT model works better and accurately [4] as shown in the Fig 6. The reason why BERT works better than the traditional model is because it looks at words in a way that help it understand the meaning and context better, making its predictions about feeling more exact.

H. *Real-Time Processing:*
The big strength of our system is to analyze the feelings right away as they happen which gives us quick insights into how people's opinions were changing. It will keep looking at the new data which is coming in all the time, so we always have the latest update on how feelings were shifting. This will help to make the decision fast and make smart choices based on what people were feeling.

I. *Adoptability and Scalability:*
One of the most important advantages of using BERT model is that it can work very well in different area and handle lots of data without any issues. It is like having a tool that fits into many different situations smoothly. BERT performs well and shows how flexible and useful in different scenarios.

J. *Insights and Patterns:*
When we start analyzing feelings, we come to know about how people feel about things, what are the trends happening in the market, and how customers think about products or services. By seeing the feelings in the form of charts and graphs makes it easy to make decisions, and understand how people behave and what they prefer. This makes it easier for them to make smart decisions.

K. *Limitations and Future Directions:*
Even though our real-time sentiment analysis system using BERT works very well, we found some challenges like the way different people talk online, biases in the data, and language differences made it tricky sometime. If we need to make our system better, we need to keep refining it and looking for ways to improve it. We can try different techniques, adapting them to specific areas, and considering different type of data to make sure our system stays accurate and reliable in the future.
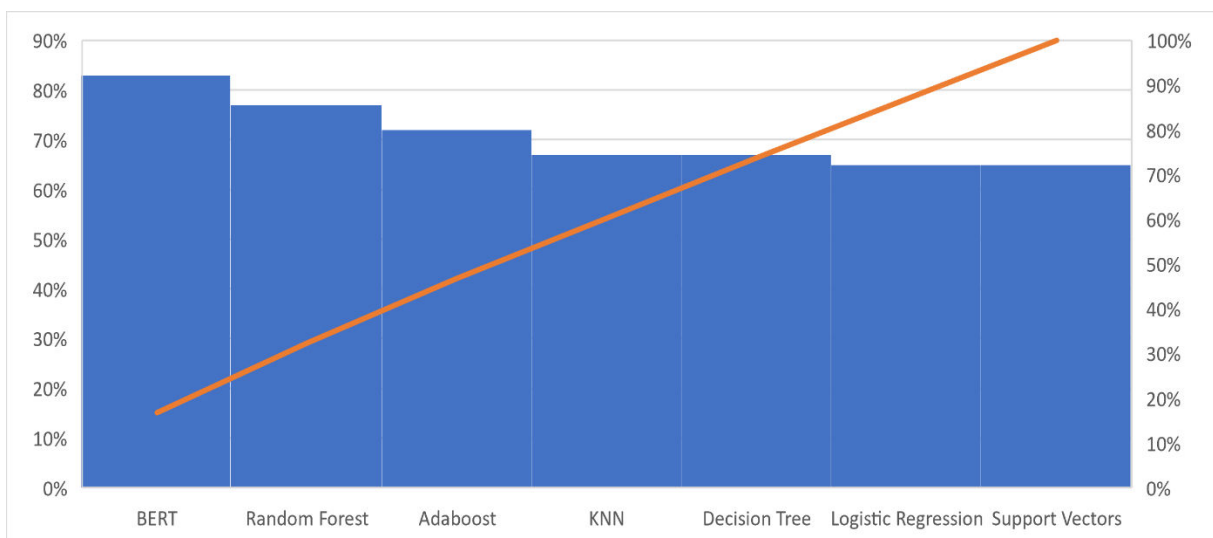

Fig 6. BERT model compared with different models.

VI. CONCLUSION

This research paper introduced a thorough method for analyzing feelings in real-time using the BERT model. This model works very well at understanding the meaning of text. Our system works very accurately, fast, and can handle a large amount of information from different places, which helps decision-making. We used advanced technology like deep learning and analyzing data as it come in. Our system gives a great way to keep track of how people feel, what is happening in the markets, and what customers think, which is very important in today's fast paced digital world.

REFERENCES

1. Sayyida Tabinda Kokab, Sohail Asghar, and Shehneela Naz. "Transformer-based deep learning models for the sentiment analysis of social media data". Array, Volume 14, July 2022.
2. Debabrata Maity, Suvarna Kanakaraddi, and Shantala Giraddi. (2023). "Text Sentiment Analysis based on Multichannel Convolutional Neural Networks and Syntactic Structure". Procedia Computer Science, Volume 218, 2023.
3. Emelie Rosenberg, Carlota Tarazona, Fermín Mallor, Hamidreza Eivazi, David Pastor-Escuredo, Francesco Fuso-Nerini , and Ricardo Vinuesa. "Sentiment analysis on Twitter data towards climate action". Results in Engineering, Volume 19, September 2023.
4. AKSH PATEL, PARITA OZA, and SMITA AGRAWAL. "Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model". Procedia Computer Science, Volume 218, 2023.
5. Cristina Zuheros, Eugenio Martínez-Cámara, Enrique Herrera-Viedma, Iyad A. Katib, and Francisco Herrera. "Explainable Crowd Decision Making methodology guided by expert natural language opinions based on Sentiment Analysis with Attention-based Deep Learning and Subgroup Discovery". Information Fusion, Volume 97, September 2023.
6. Alaa S. Al Shammari. "Real-time Twitter Sentiment Analysis using 3-way classifier". IEEE Xplore 21st Saudi Computer Society National Computer Conference (NCC), December 2018.
7. Jahanzeb Jabbar, Iqra Urooj, Wu JunSheng, Naqash Azeem. "Real-time Sentiment Analysis On E-Commerce Application". IEEE Xplore International Conference on Networking, Sensing and Control (ICNSC), June 2019.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING