iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** I **Month of publication:** January 2021

DOI: https://doi.org/10.22214/ijraset.2021.32874

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue I Jan 2021- Available at www.ijraset.com

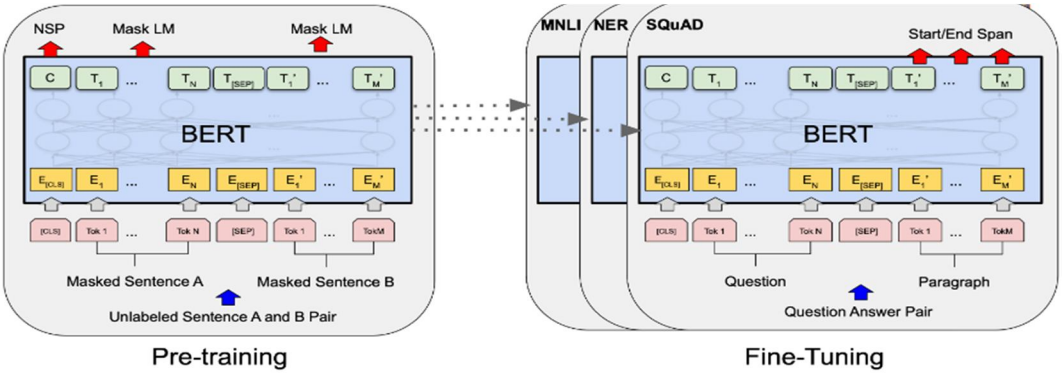
Interpretation of Sentiment Analysis of Tweets through Artificial Intelligence

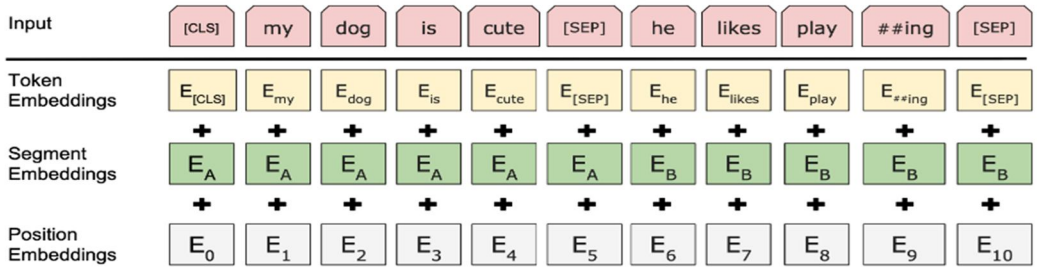
Onkar Apte[1], Mohit Bhandari[2]
[1, 2,] *KC College of Engineering*

Abstract: ***Sentiment analysis, also known as opinion mining is a sub-field of Natural Language Processing (NLP) that tries to identify and extract opinions from a given text. This project presents a software tool for sentiment analysis of tweets which combines an advanced technique of machine learning BERT (Bidirectional Encoder Representation from transformer) with sentiment analysis for analysing a given tweet. The project by using its state-of-the-art software model also briefly studies about the different sentimental patterns outputted by a given set of data while analysing the features of given text, of different text adding to various research aspects.***

I. INTRODUCTION

Sentiment analysis or opinion mining based on Natural language processing and machine learning to interpret and classify emotions in subjective data. It is a kind of text classification technique which classify technique based on direction of the object they contain. It is also known as process of detecting the contextual polarity of text. Sentiment analysis is often used in various organization like business, social media etc, whereas to detect sentiment in social data, gauge brand reputation, and understand one's .and its application predominantly have propagate to every feasible domain, from services, financial activities, specific company product ,basically in the given text it can identify whether it is positive ,negative or neutral. The idea of instinctive sentiment analysis is significant for publicizing research, to find out the people thinks for their artefact.one of the most important as in recent trend social media one's own comment. Fast and automatic detection of flickering is necessary, for scrutiny of people's feedback or as instructive augmentation for search engines. Social media act as a medium where the user can post many opinions a day, and these can be used for classification, where it can derive the appraisal, emotion or attitude of the reciter. Due to trend in souk and the changing need of people a lot of research work held in the field of sentiment analysis. Sentiment analysis requires the usage of a training dataset for evaluating the process.





ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue I Jan 2021- Available at www.ijraset.com

II. RELATED WORK

The assortment and sentiment analysis of data has already been addressed by multiple papers. In this section, we will use the following tools that were used for this approach as well as previous findings regarding this specific research in this paper

A. Problem Statement

Sentiment analysis is essentially a process of computationally determining whether a given text expresses a positive, neutral, or negative emotion. Sentiment analysis for social media content can be used in various ways. It is a field of exploration to systematically identify, extract, quantify, and study affective states and subjective information. It plays an important role for identifying the emotion. Therefore, to get proper response then give it as input to the most powerful State-of-the-Art Natural Language Processing algorithm to decode the given sentences and finally predict the output.

III. NLP AND ITS APPLICATION FOR SENTIMENT ANALYSIS OF TWEET:

Natural Language Processing (NLP) refers to AI method of communicating with an intelligent systems using a natural language such as English. Processing of Natural Language is required when you want an intelligent system like robot to perform as per your instructions, when you want to hear decision from a dialogue based clinical expert system, etc. Natural language empathetic is particularly difficult for machines when it comes to opinions, given that humans often use cynicism and irony. Sentiment analysis, however, is able to recognize subtle nuances in emotions and opinions – and determine how positive or negative they are.

Analyzing sentiment in real-time, can monitor mentions on social media (and handle negative comments before they escalate), gauge customer reactions to your latest marketing campaign or product launch, and get an overall sense of how customers feel about your company whereas performing sentiment analysis periodically, and understand what customers like and dislike about specific aspects of your business. Those insights can help you make smarter decisions, as they show you exactly what things to improve.

By the late 90s and with dawn of the 21st century, there were significant development in technologies with better hardware providing much more improved software optimizations, which added to a significant shit focus of AI applications in precision medicine, particularly cancer diagnosis and treatment With increasing applications in classification, AI-based systems have become part of nearly considerably decreasing human errors and increasing efficiency, thus proving reliable in required applications.

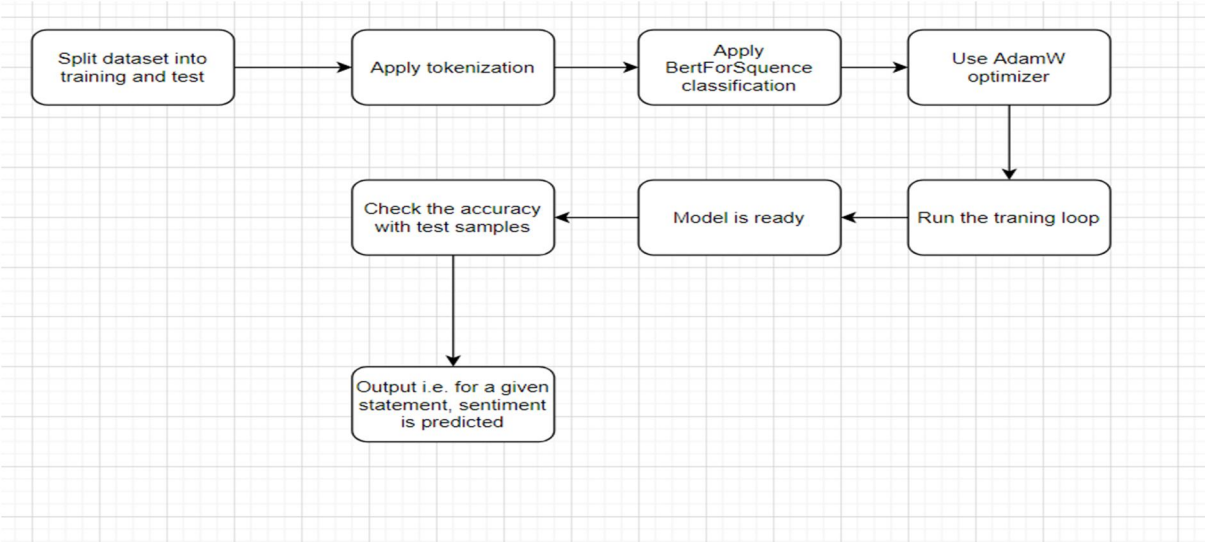
IV. TECHNIQUE

This project uses the method of NLP (Natural language processing), extensively focused on a software platform based on neural networks. The project uses BERT (Bi Directional Encoder Representation from transformers) based on machine learning technique developed by Google, for the purpose of making analysis of sentiment analysis of tweets. This model has 12-layer, 768-hideen, 12-heads, 110M parameter neural network architecture. It also uses Hugging Face, which is a library that provides transformers across different domains within NLP. Transformers essentially are just advancement recurrent neural nets which are able to parallelize processing and training. It also has optimizers such as Adam W, which is Adam optimizer with weight decay. Dataset we used for our sentiment analysis is SMILE Twitter Emotion dataset and then the dataset is retrieved for the training purpose of model This collection of tweets mentioning 13 Twitter handles associated with British museums was gathered between May 2013 and June 2015.It was created for the purpose of classifying emotions, expressed on Twitter towards arts and cultural experiences in museums. It contains 3,085 tweets, with 5 emotions namely anger, disgust, happiness, surprise and sadness. Whereas in achieving our end is goal is tokenization. Tokenization is conversion of text into tokens before applying transformations. Basically it assigns a number to each word from the dataset which makes easier for further processing.

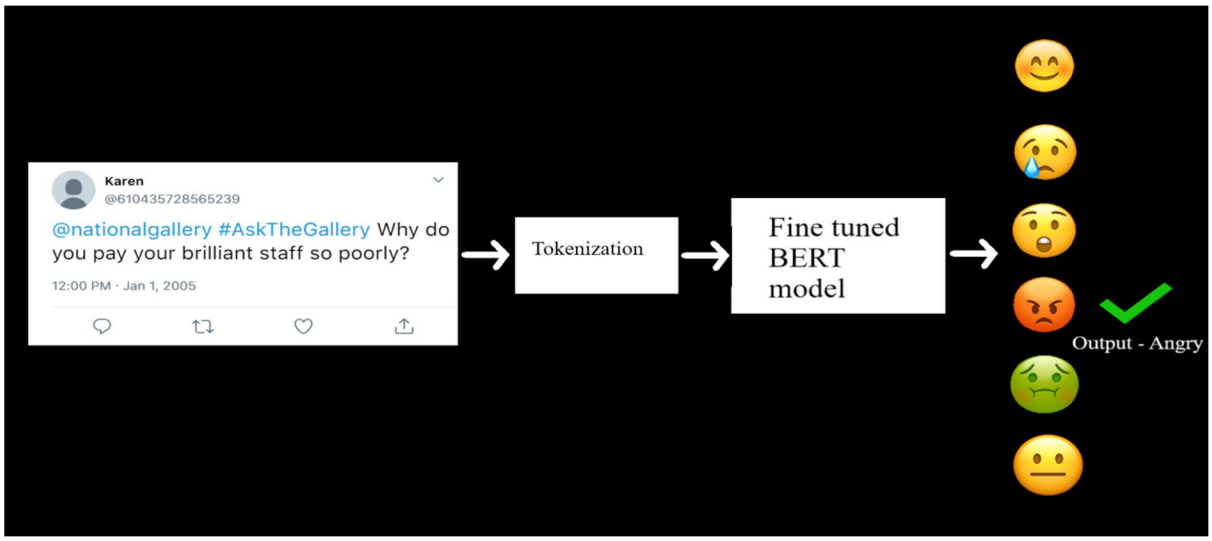
In stage 1 we encoded our data using BERT Tokenizer which converted our text data into numbers so that PyTorch can infer the data, infer results from the data as well as train from it. In the step 2 and 3 we use “BertForSequeceClassification” from Hugging Face library. Here, we are treating each tweet as its own unique sequence. So one sequence will be classified into one of six classes namely five emotions anger, disgust, happiness, surprise and sadness and not-relevant class which basically is a class for all the tweets which have no clear emotions expressed in them. Next is learning optimization. We are using AdamW for optimization and “get_linear_schedule_with_warmup” from transformers library from Hugging Face for scheduling. Optimizer is just defining our learning rate and Scheduler controls how it changes through each epoch. Learning rate which we decided after testing exhaustively is 1e-5. We found that this hyper parameter works best for our use case and after this we defined our performance metrics.

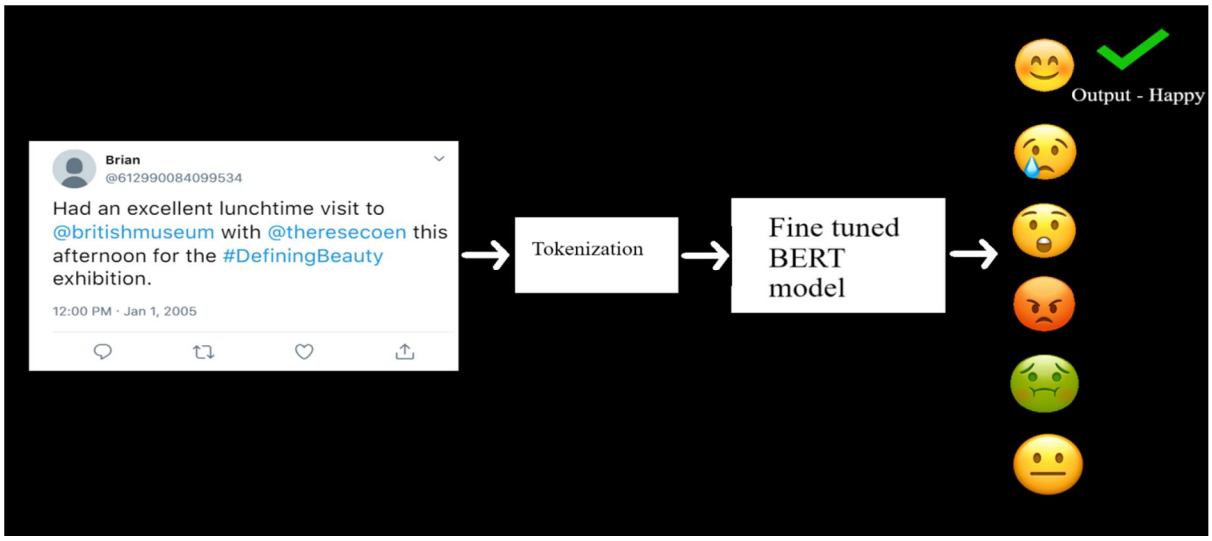
After this we are ready to train our model which is the most important and most intricate task. Our training loop is based on “run_glue.py” script from HuggingFace which is recommended by HuggingFace when you train BERT model on your own dataset. By checking the output we conclude that the program is able to recognize the desired result correctly.

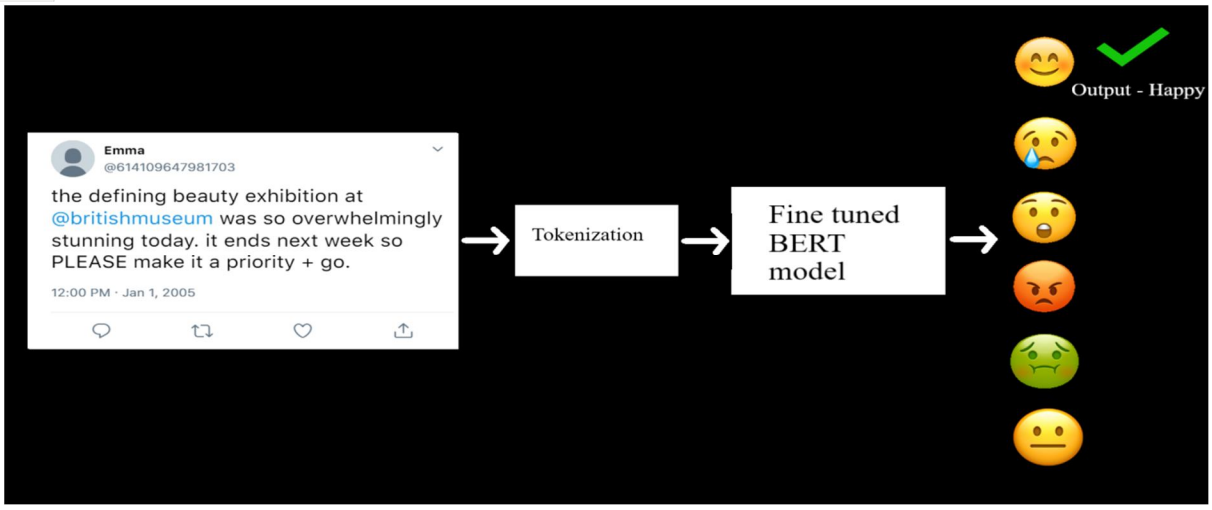
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue I Jan 2021- Available at www.ijraset.com



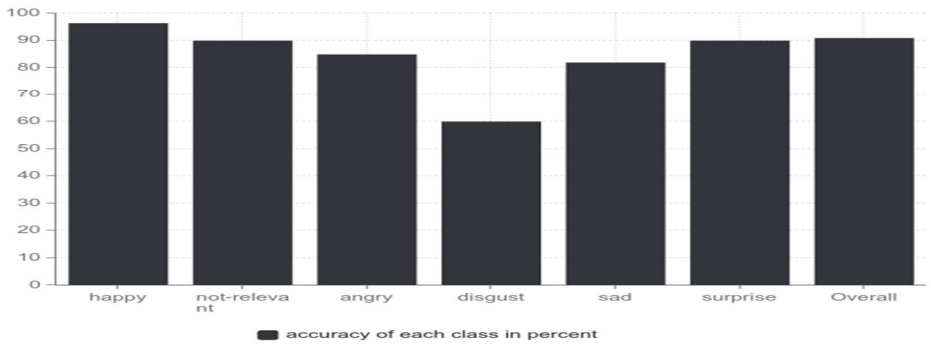
V. TEST SAMPLES



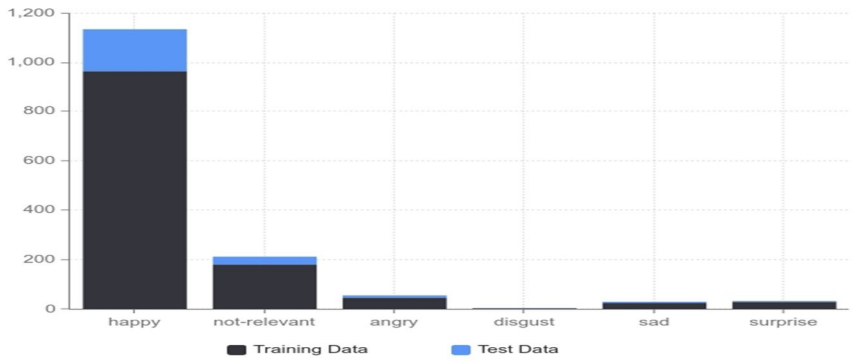




Accuracy



Dataset Distribution

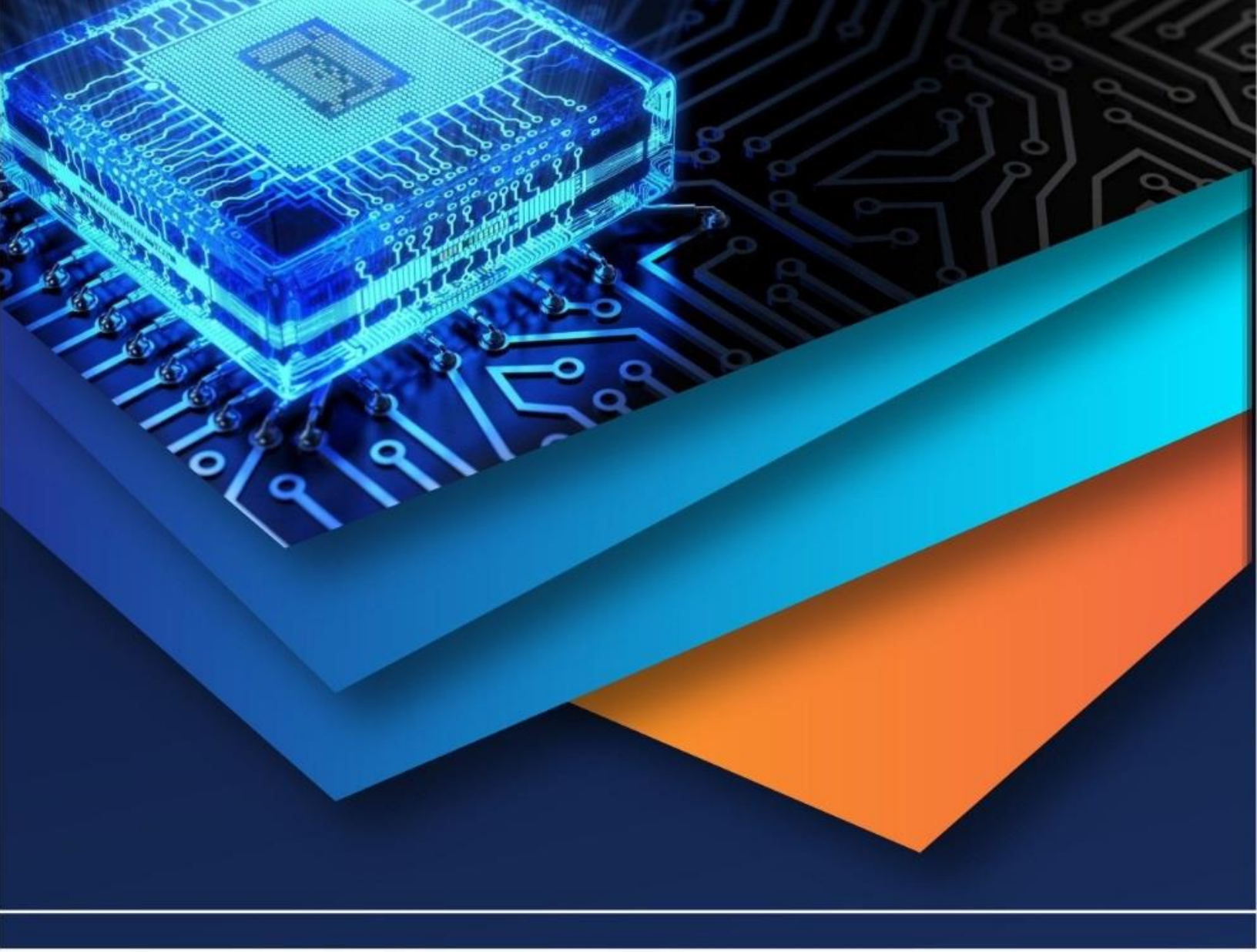


VI. DISCUSSION AND SCOPE OF STUDY

In this project, we propose a possible software based solution for the analysing of tweets through contextual trepidations based on the BERT Masked Language Model. This can be useful in many fields. Amazon like e-commerce can use it to know emotions of product review, Sites like You tube can use it to recognize and remove hate speech comments. Companies can scrap data from twitter which mentions their company name and analyse what people's emotions are towards their company and products. We propose inserting and/or replacing tokens from a sentence, in their order of importance for the text classification task. Automatic and human evaluation on several datasets demonstrates the strength and effectiveness of our attack. We have succeeded to identify positive, negative and neutral feelings to the entity under consideration 91.08 % accuracy based on various algorithm. In addition, our experiments give us insights into the portability of the learned models across various texts.

REFERNCES

[1] https://huggingface.co/transformers/model_doc/bert.html
[2] https://figshare.com/articles/dataset/smile_annotations_final_csv/3187909/2
[3] https://arxiv.org/abs/1810.04805
[4] Shashi Narayan, Shay B Cohen, and Mirella Lapata. 2018. Ranking sentences for extractive summarization with reinforcement learning. In Proceedings of the NAACL Conference
[5] Payal Bajaj, Daniel Campos, Nick Craswell, Li Deng, Jianfeng Gao, Xiaodong Liu, Rangan Majumder, Andrew McNamara, Bhaskar Mitra, Tri Nguyen, Mir Rosenberg, Xia Song, Alina Stoica, Saurabh Tiwary, and Tong Wang. 2016. MS MARCO: A human generated MAchine Reading COmprehension dataset. arXiv:1611.09268.
[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805.
[7] Jianqiang, Z., Xiaolin, G., Xuejun, Z.: Deep convolution neural networks for Twitter sentiment analysis. IEEE Access **6**, 23253–23260 (2018)
[8] Abid, F., Alam, M., Yasir, M., Li, C.: Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter. Future Gener. Comput. Syst. **95**, 292–308 (2019)
[9] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis
[10] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications", Expert Syst. Appl., vol. 77, pp. 236-246, Jul. 2017
[11] R. Baly, G. El-Khoury, R. Moukalled, R. Aoun, H. Hajj, K. B. Shaban, et al., "Comparative evaluation of sentiment analysis methods across arabic dialects", Procedia Comput. Sci., vol. 117, pp. 266-273, May 2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY