

Depression Detection by Analyzing Social Media Post of User: A Review

Prof. A. P. Kshirsagar, Aditya Ghodke, Sahil Kadam, Shubhum Jadhav, Parikshit Jadhav, Aditi Bhosale
Department of Computer Science And Engineering
Karmveer Bhaurao Patil College of Engineering, Satara, India
aditya.ghodke@kbpcoes.edu.in, sahil.kadam@kbpcoes.edu.in, shubhum.jadhav@kbpcoes.edu.in
parikshit.jadhav@kbpcoes.edu.in, Aditi.bhosale@kbpcoes.edu.in

Abstract: *Today, the problem of early diagnosis of depression is one of the most important problems in psychology. Mental health problems are often one of the most important health problems in the world, with depression alone currently affecting more than 300 million people. As social media platforms create more male or female user accounts, researchers are increasingly using evidence-based data to determine whether content can be used to spread mental health issues in users. Scientists around the world say that depression is a disease that causes serious problems in our lives and is still a cause for concern. With the advent of all-inclusive devices such as smartphones, predicting depression remains an open question. Social testing is often used to solve this problem. This paper aims to use a depression assessment and suicidal ideation detection system to predict the level of depression supporting suicidal behavior. The purpose of this system is to provide the best tools to understand the process by which a man or woman is depicted through words as being in a relationship that may be disturbing. For this purpose, experts and productive people are used to determine whether the person is depressed or not by using their ability to do physical work at work. They were trained to use similar equipment and divided into different levels of depression on a scale of 0-100%. It also collects the logs in the report and sorts them by whether it's the best tweet you've ever sent. Whether you suffer from depression or not, using cutting-edge technology is a method for early detection of depression or mental illness. The main purpose of this evaluation is to investigate relevant resources and their impact on determining depression levels. The aim of this review is to understand the criteria used to classify people with depression by looking at some of the cases discovered by postgraduate studies examining the education of men and women. With the ability to combine all tag groups, you can create periodic reports that can be used to identify customers suffering from depression. This article shows that there is variability in reporting patterns between depressed and non-depressed individuals, as evidenced by the association of reporting groups. This study used cognitive tools to create data collected from consumer social media posts. Natural Language Processing (NLP) has been classified using the BERT code set to represent depression in a simpler and more understandable way.*

Keywords: machine learning, NLP, BERT algorithm, depression, classification, social media

I. INTRODUCTION

Today, the problem of early diagnosis of depression is one of the most important problems in psychology. Depression is also a matter of vile grace. The stress of life in today's world has to be stressful. Depression affects more than 350 million people worldwide, approximately 5% of the population. Nearly 800,000 people die from suicide each year, and distance is the second leading cause of decreased life expectancy for ages 15-29. Meanwhile, the main type of suicide is related to depression. Recent research shows that depression also leads to disability and serious physical illness. The proliferation of the Internet and communication technology, especially online dating, has improved the way people interact and communicate with different people electronically. Facebook, Twitter, Instagram etc. programs now offer not only written and multimedia content, but also customers' thoughts, feelings, and opinions about a topic, issue, or problem online. On the other hand, this is surprising for the discussion site's customers, who can help open and answer statrelated questions online for free;

On the other hand, it creates an opportunity for healthcare professionals to understand what can happen if one person reacts to an issue in forward-thinking America. To provide such an insight, the tool for cognitive skills has a special tool that will help to identify certain patterns hidden in online communication and use them to see the complexity of Americans (such as "happiness", relationship "Sadness", "anger", "anger"). must have something. "Anxiety and depression among online consumers." There is also a growing body of research on the characteristics of social anxiety, including the relationship between social anxiety and aesthetic disorders ("melancholia"), "depression," "bipolar disorder," etc.), smoking, relapse, sexual violence and suicidal thoughts. Young people of both genders, minorities, key workers and unpaid carers report impaired cognitive abilities, increased drug use and mental instability. It is characterized by a daily process of physical, emotional and social adaptation at the age of twenty-four. Young people want the easiest way to achieve happiness, love, mobility and freedom to grow and improve health, and to have a job in life. The development of many cultural attitudes during this period of the herbal approach will lead to many types of normal or cognitive disorders. Depression can cause affected individuals to experience severe pain and dysfunction at work, university, and relationships. No matter what you are currently doing on your phone or computer, social media is clearly visible to people. Have you checked in with your friends on Facebook, posted a snapshot of your conversation or a video of your daughter's first walk on Instagram? This will be the link Twitter sent you here. Nowadays, parents love to share their thoughts, feelings and daily lives with the spread of social media such as Twitter, Facebook and In. Instagram. These articles are often included in snaps, videos, and posts. In this study, we aim to analyze social media posts to reveal all factors related to social media users' depression. For this purpose, various tools are used to obtain technical information. Considering the main purpose of this study, the following are the factors relevant to the next research mentioned in this article. We love using insight and algorithmic tools to identify clients' social media challenges.

II. NLP

The artwork described in this article is in the field of natural language processing (NLP), especially beautiful text. The origin of the text beautification function can be identified in early research, particularly in the classification of documents based on a 1961 analysis of specific content. and finally within a year the business started to shift a bit towards machine analysis algorithms. 2000. It will first be combined with machine analysis to find important or dangerous aspects in video analysis, then expanded to specialized analysis such as social media quality and authenticity. Recently, an in-depth analysis has been made about the beauty of the text, as it has many uses besides the beauty of the image. For example, some of the state-of-the-art operations of text processing can be applied to the representation of language through a transformation process, such as Universal Language Model Fine-Tuning (ULM Fit) and Google Research Project Transformers' Bidirectional Encoder. Representation (BERT). BERT's rules and various planning standards.

III. LITERATURE REVIEW

Useful opportunities are often created to study user behavior in networks. Techniques from discourse analysis have been used effectively to examine social media in particular. The data was collected from customer posts on leading social media sites Twitter and Facebook. In this study, a learning tool was used to process data collected from SNS (social networking site) users. Natural Language Processing (NLP) recorded using Support Vector Machines (SVMs) and Naive Bayes Rulesets can eliminate stress in a more accessible way. [1]

To conduct research using positive language processing techniques (NLP) to target for the purpose of smuggling order order to order, to brush, to brush) to be done. [2]

Investigate depression and anxiety in exercise tweets by combining tweets using Multinomial Naive Bayes and Support Vector Regression (SVR) algorithms as the distribution on the form sheet [3]. 4) In the paper, the researchers proposed a method to analyze and extract thoughts from text using thinking theory, learning strategies, and instinctive language processing strategies for seeing someone in distress. [4]

The article aims to use the materia medica study of the Twitter site for depression-specific emotional assessment. The tweets themselves are labeled as unethical or negative and are often based on a series of statements that appeal to

negative emotions. Bootstrap vector and negative Bayes classifier were used for complex prediction. Results from the first category evaluation include the F1 score, accuracy, and confusion matrix. [5]

This article describes the depression and suicidal ideation test, which often predicts suicidal behavior depending on the level of depression. Notes are then written in the form of tweets and questions. It is then divided into 5 collapse levels based on weight using the feature algorithm. [6]

Yates et al. used a neural network version to demonstrate the dangers of self-harm and depression, often relying on Reddit and Twitter posts, and confirmed the accuracy of this diagnosis. In addition to clinical research, the authors suggest that the proposed strategy can also be used in large-scale health studies. [8]

OâDea et al. Research shows that Twitter is increasingly used as a way to monitor mental health conditions, including depression and suicide. Their study found that the fear level of suicide-related tweets could be easily inferred using both a human coder and a programmed device classifier. Determining depression levels in social media posts about suicide. In our study, we provide a description of strategies used for people with anxiety using the process of writing words to describe depression using the BERT rule. The framework consists of a data preprocessing step, a feature extraction step, followed by a machine learning classifier, implicit feature evaluation, and testing.

PROBLEM STATEMENT

Depression has been shown to affect self-expression. Create software that uses machine learning tools to detect and resolve problems on customers' social media. This project uses language processing, machine learning, and neural network architectures to develop, improve, and learn models to classify users' social media posts as "depressed" or "not having a hard time."

OBJECTIVE

The objectives are as follows:

The system will continue to track users' posts and chat history. If the device detects bad behavior, it will automatically post positive content to your wall based on your depression level.

Help people escape depression

IV. METHODOLOGY

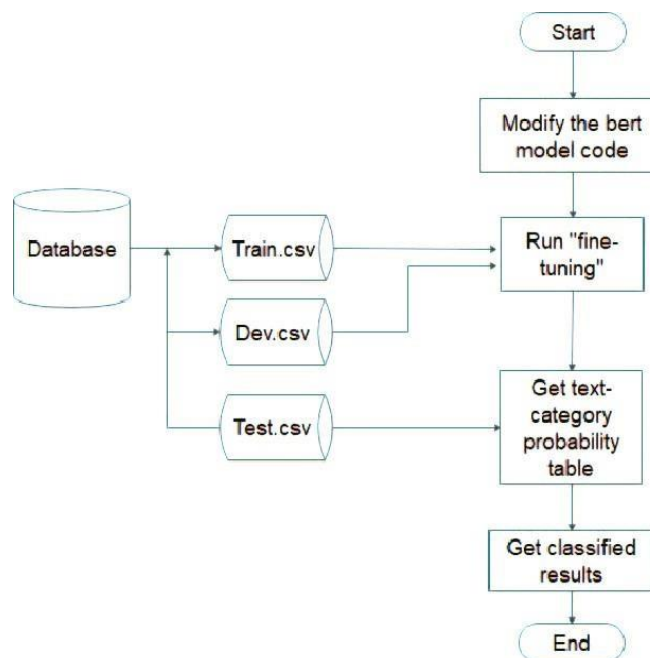


Fig1. Bert algorithm

Machine learning classification techniques are used by such

BERT Algorithm:

BERT stands for Bidirectional Encoder Representation of Transformers. It aims to provide a deep representation of untitled text using a valuable resource that combines all elements from left and right. Therefore, BERT foundations can be improved with larger algorithms to create routine models for many NLP tasks. The sample language is as described in the section. For this project this output set (aggregate objects) was modified for binary analysis. Of the many pre-professional models, we decided to use the non-English (all lowercase before tokenization) BERT model because factual information is not particularly important for broadcast request type work.

MATHAMATICAL MODEL

Counting numbers related to the project: System Description:

S=I, O, F, DD, NDD, Inactive, Successful

Where, S= system I= Login O=Output F=Failure S=Success

I is the input of the system

Input I = input set

Where

I = (user social media post)

F is the function of the system) F = (function set)

Among these,

F1= (input dataset)

F2= (Json to CSV conversion) F3=(preprocessing) F4=(cleaning)

F5= (training test section) F6= (Emotion Dictionary)

F7= (Classifier (BERT algorithm) F8=(Word Segmentation) Output for the system

Output O1= (Time Melancholy Detection)

Success status: Product is good it works. Complete the stress test.

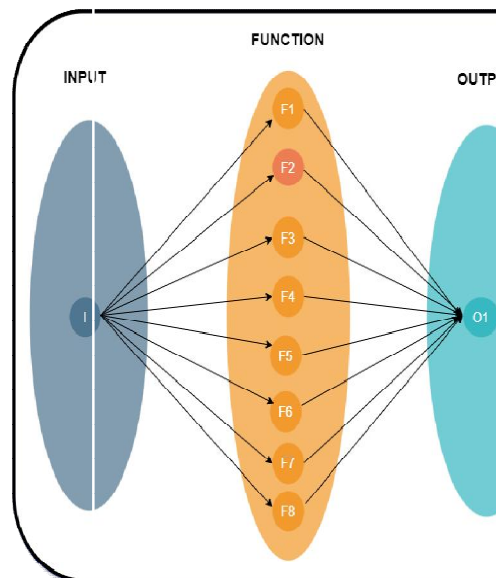


Fig2 . Venn Diagram

At,

I = (post of social media) F1= (input dataset)

F2= (Json to CSV conversion) F3=(preprocessing)

F4 = (Clean)

Copyright to IJAR SCT

www.ijarsct.co.in

DOI: 10.48175/IJAR SCT-18671



F5= (Training Test Split) F6= (Emotional Dictionary)
 F7= (Classifier (BERT algorithm)) F8=(Tokenization)
 Output O1 = (Detection of Depression)

SYSTEM ARCHITECTURE

Depression leads to poor performance in private and public health. One of the best answers to this question is to pay attention to the person's personality. These behaviors are common on social networking sites such as Facebook, Twitter and Instagram. He belongs to the young generation. People share their unique thoughts, daily activities, thoughts on various topics, etc. they share on social media. These social networks provide people with information, ideas, relationships and identity. Advanced measurements of individual impact are not always available, but these of customer-generated content on social media can provide an early indicator of an individual's level of engagement wisdom and desperation. The motivation for our study is to extract information from social media and predict customer frustration by clearly understanding individual behaviors and questions. Conduct extensive research to inform and evaluate the tool for many separate studies to determine how the customer's content is published on the social network, the process that begins from the customer's exposure, or the physical activity he engages in on social media. The following judgments show a surprising use of the likes and contents of the potential data model. First, all tweets about shortages and shortages are returned, as well as information on consumer and sports activities, as well as the number of followers, number of bites Followers, duration of posts, number of comments and number of retweets. . Then, all of the account's posts are summarized in a report. All documents were completed as text first. First, the corpus is created and the posts in each report are tagged. You can use the BERT classification algorithm.

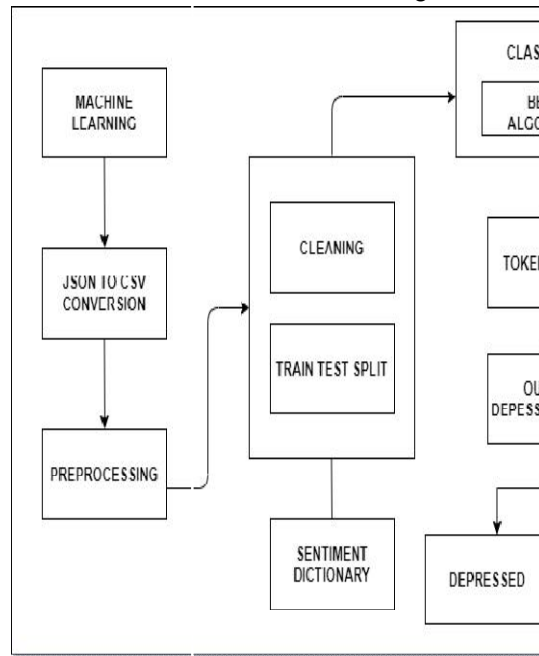


Fig3. System Architecture

V. EXISTING SYSTEM

This free tool provides an easy-to-use method to determine a customer's trust level using the Naive Bayes algorithm. Data extraction is done by Facebook's Extraction Beauty with the help of Facebook Graph API. Once extracted, it undergoes actual pre-treatment. Missing or duplicate features are processed first. Techniques such as tokenization, conversion to small numbers, stemming, and time removal are used ahead of the truth. In the proposed tool, the pattern of an average user's Facebook posts can determine whether he or she is now depressed. But analyzing the text alone does not give accurate results, so we also look at the messages of the user and his friends, and we also analyze the

conversation history, because the user can definitely express his frustration towards his friends. Based on these measurements, customers can be labeled as stressed or non-stressed.

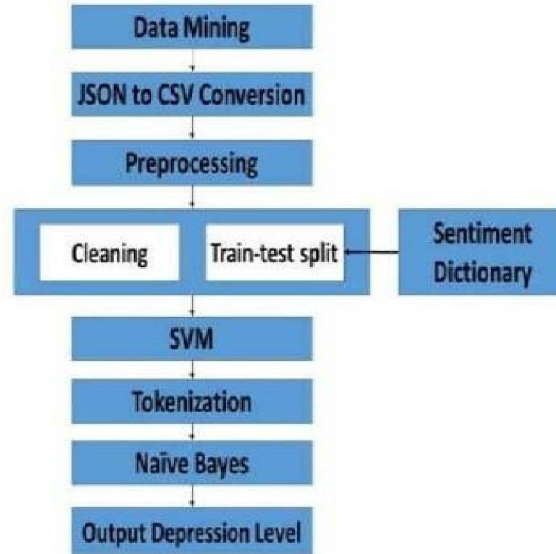
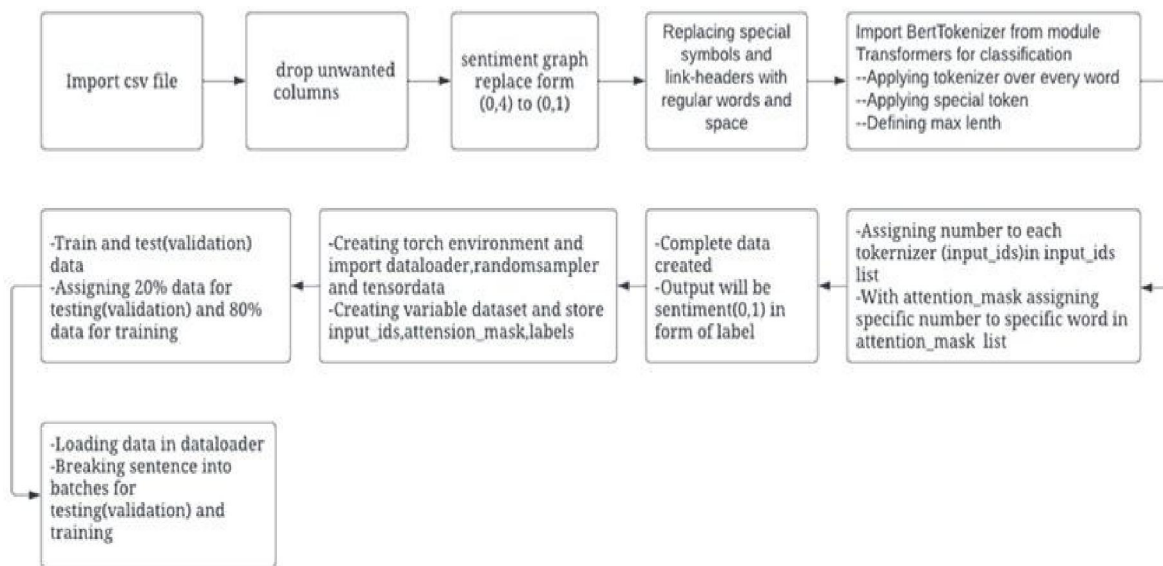


Fig4.Exisiting System

Implementation

- Processing Of The Data
- Testing After Training Data
- Creating and Connecting Frontend

Data Processing:



Screen view Result 1:

```
df = pd.read_csv('C:/Users/karan/OneDrive/Desktop/Depression/Dataset/depression.csv', encoding='latin-1', header = None)
df.columns=['Sentiment', 'id', 'Date', 'Query', 'User', 'Tweet']
df = df.drop(columns=['id', 'Date', 'Query', 'User'], axis=1)
✓ 13.2s
```

```
df.head()
✓ 0.3s
```

	Sentiment	Tweet
0	0	@switchfoot http://twitpic.com/2y1zl - Awww, t...
1	0	is upset that he can't update his Facebook by ...
2	0	@Kenichan I dived many times for the ball. Man...
3	0	my whole body feels itchy and like its on fire
4	0	@nationwideclass no, it's not behaving at all...

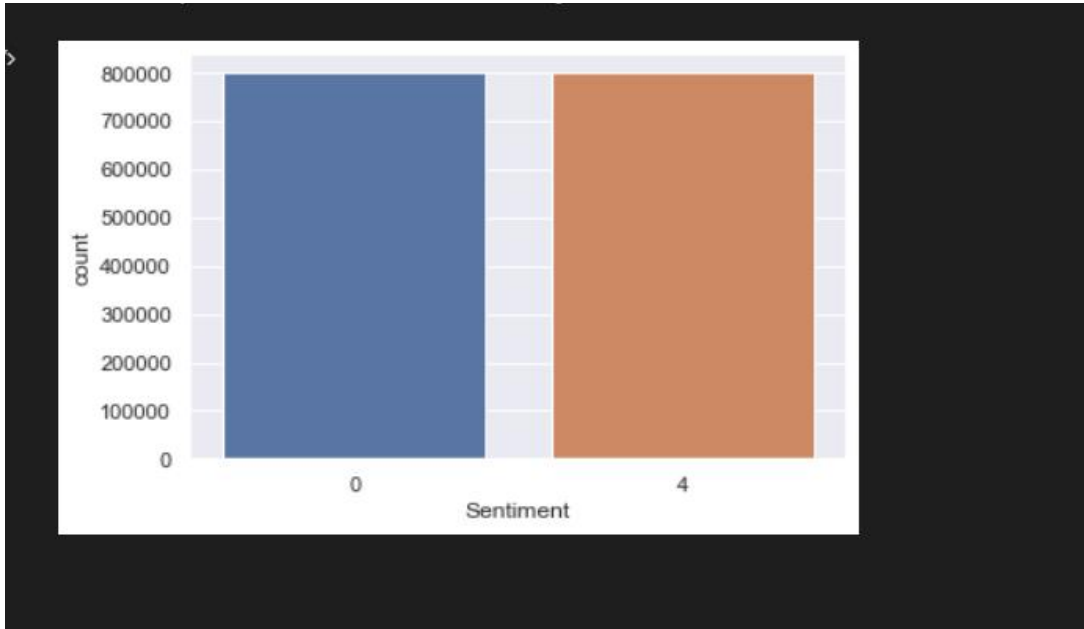
```
hashtags = re.compile(r"^\#\S+|\s#\S+")
mentions = re.compile(r"^\@\S+|\s@\S+")
urls = re.compile(r"https?:\/\/\S+")

def process_text(text):
    text = re.sub(r'http\S+', '', text)
    text = hashtags.sub(' hashtag', text)
    text = mentions.sub(' entity', text)
    return text.strip().lower()
✓ 0.1s
```

```
df['Tweet'] = df.Tweet.apply(process_text)
✓ 41.2s
```

```
df.head()
✓ 0.3s
```

	Sentiment	Tweet
0	0	entity - awww, that's a bummer. you shoulda ...
1	0	is upset that he can't update his facebook by ...
2	0	entity i dived many times for the ball. manage...
3	0	my whole body feels itchy and like its on fire
4	0	entity no, it's not behaving at all. i'm mad. ...



```
df['Sentiment'] = df.Sentiment.replace(4,1)
✓ 0.1s

sns.countplot(df.Sentiment)
✓ 0.8s

C:\Users\karan\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Passing the following argument will be `data`, and passing other arguments without an explicit keyword will result in an error in the future.
  warnings.warn(

<AxesSubplot:xlabel='Sentiment', ylabel='count'>
```



A bar chart with a white background and a black border. The y-axis is labeled 'count' and ranges from 0 to 800,000 in increments of 100,000. The x-axis is labeled 'Sentiment' and has two categories: '0' and '1'. The bar for '0' is blue and reaches approximately 780,000. The bar for '1' is orange and reaches approximately 780,000.

Sentiment	Count
0	~780,000
1	~780,000


```
train_size = int(0.8*len(dataset))
val_size = len(dataset) - train_size

train_dataset, val_dataset = random_split(dataset, [train_size, val_size])

print('Training Size - ', train_size)
print('Validation Size - ', val_size)
```

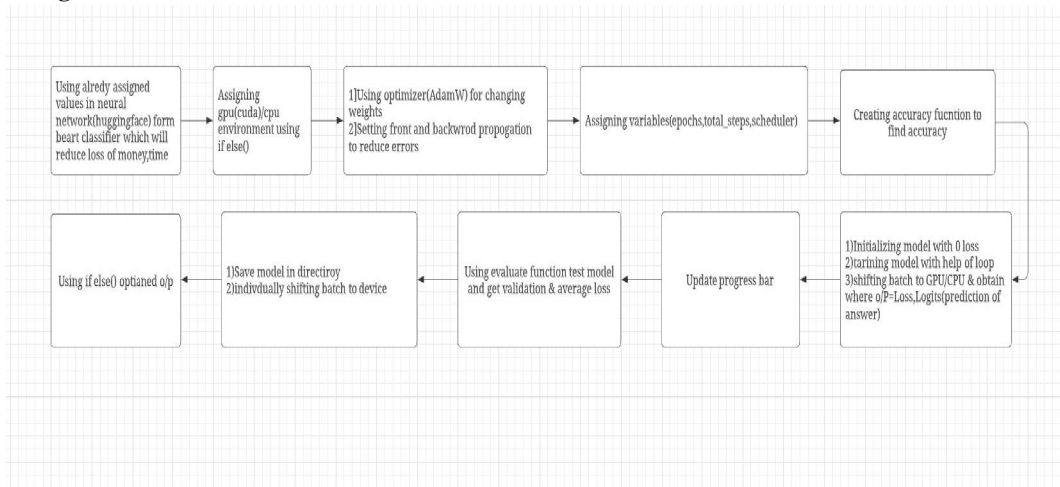
```
Training Size - 1280000
Validation Size - 320000
```

```
train_dl = DataLoader(train_dataset, sampler = RandomSampler(train_dataset),
                    batch_size = 32)
val_dl = DataLoader(val_dataset, sampler = SequentialSampler(val_dataset),
                  batch_size = 32)
```

```
len(train_dl), len(val_dl)
```

```
(40000, 10000)
```

Data Testing



Screen View Result 2:

```

Module II

model = BertForSequenceClassification.from_pretrained(
    'bert-base-uncased',
    num_labels = 2,
    output_attentions = False,
    output_hidden_states = False)
[44] Python
...
</>

Some weights of the model checkpoint at bert-base-uncased were not used when initializing BertForSequenceClassification: ['cls.predictions.bias',
'cls.predictions.transform.dense.weight', 'cls.predictions.transform.dense.bias', 'cls.predictions.decoder.weight', 'cls.seq_relationship.weight', 'cls.seq_relationship.bias',
'cls.predictions.transform.LayerNorm.weight', 'cls.predictions.transform.LayerNorm.bias']
- This IS expected if you are initializing BertForSequenceClassification from the checkpoint of a model trained on another task or with another architecture (e.g. initializing a BertForSequenceClassification model from a BertForPreTraining model).
- This IS NOT expected if you are initializing BertForSequenceClassification from the checkpoint of a model that you expect to be exactly identical (initializing a BertForSequenceClassification model from a BertForSequenceClassification model).
Some weights of BertForSequenceClassification were not initialized from the model checkpoint at bert-base-uncased and are newly initialized: ['classifier.weight', 'classifier.bias']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
  
```

Defining Accuracy and Evaluate Function

```
def accuracy(preds, labels):  
    pred_flat = np.argmax(preds, axis=1).flatten()  
    label_flat = labels.flatten()  
    return np.sum(pred_flat==label_flat)/len(label_flat)
```

```
def evaluate(data_loader_test):  
    model.eval() = eval mode  
    loss_val_total = 0  
    predictions, true_vals = [], []  
    for batch in data_loader_test:  
        batch = tuple(b.to(device) for b in batch)  
        inputs = {  
            'input_ids': batch[0],  
            'attention_mask': batch[1],  
            'labels': batch[2]  
        }  
        with torch.no_grad(): # No gradient descent  
            outputs = model(**inputs)  
            loss = outputs[0] #1. Loss  
            logits = outputs[1] #2. Logits  
            loss_val_total += loss.item() # Validation Loss
```

```
output_dir = 'Model/'  
model_to_save = model.module if hasattr(model, 'module') else model  
model_to_save.save_pretrained(output_dir)  
tokenizer.save_pretrained(output_dir)
```

```
( './tokenizer_config.json',  
  './special_tokens_map.json',  
  './vocab.txt',  
  './added_tokens.json')
```

Loading save model and predicting

```
from transformers import BertTokenizer, BertForSequenceClassification  
import torch  
# Load the BERT tokenizer.  
print('Loading BERT tokenizer...')  
output_dir = 'Model/'  
tokenizer = BertTokenizer.from_pretrained(output_dir)  
model_loaded = BertForSequenceClassification.from_pretrained(output_dir)
```

Loading BERT tokenizer...

```

model_loaded = model_loaded.to(device)
#previously i was shifting whole batch into device
input_id = input_id.to(device) #individually shifting to device
attention_mask = attention_mask.to(device)#individually shifting to device

with torch.no_grad():
    outputs = model_loaded(input_id, token_type_ids=None, attention_mask=attention_mask)

logits = outputs[0]
answer = logits.argmax()
return answer
    
```

Python

```

ans = Sentiment('i am happy')
    
```

Python

Truncation was not explicitly activated but 'max_length' is provided a specific value, please use 'truncation=True' to explicitly truncate examples to max length. Defaulting to 'longest_first' truncation strategy. If you encode pairs of sequences (GLUE-style) with the tokenizer you can select this strategy more precisely by providing a specific strategy to 'truncation'.

C:\Users\karan\anaconda3\lib\site-packages\transformers\tokenization_utils_base.py:2271: FutureWarning: The 'pad_to_max_length' argument is deprecated and will be removed in a future version, use 'padding=True' or 'padding='longest'' to pad to the longest sequence in the batch, or use 'padding='max_length'' to pad to a max length. In this case, you can give a specific length with 'max_length' (e.g. 'max_length=45') or leave max_length to None to pad to the maximal input size of the model (e.g. 512 for Bert).

```

warnings.warn(
    
```

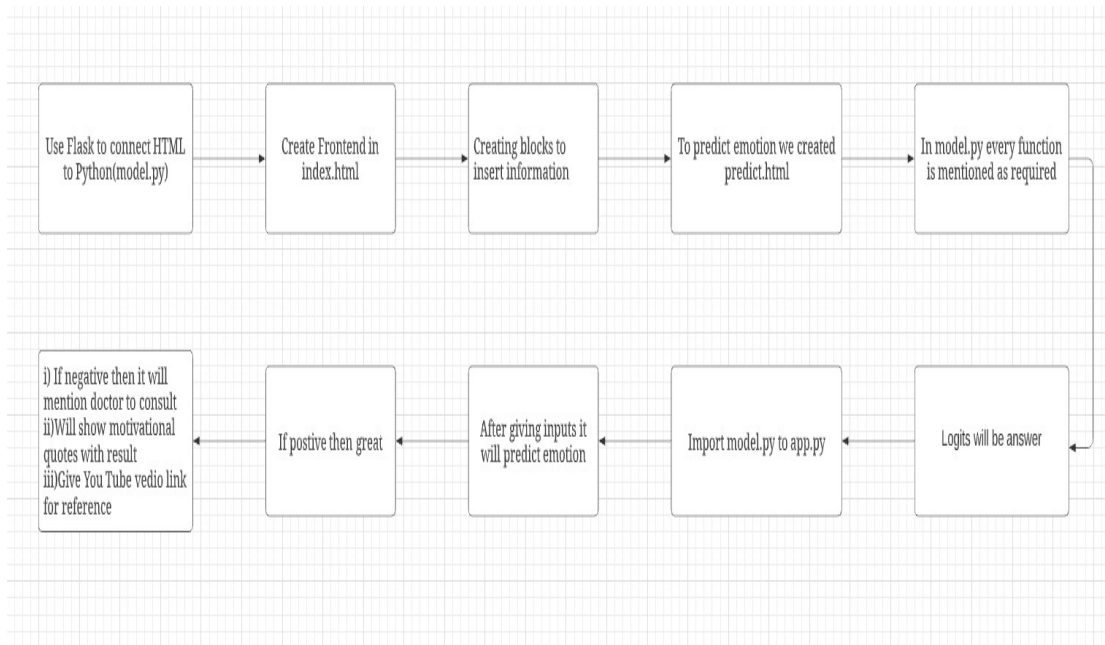
```

if ans == 1:
    print("Happy")
else:
    print("Depressed")
    
```

Python

Depressed

Front End:

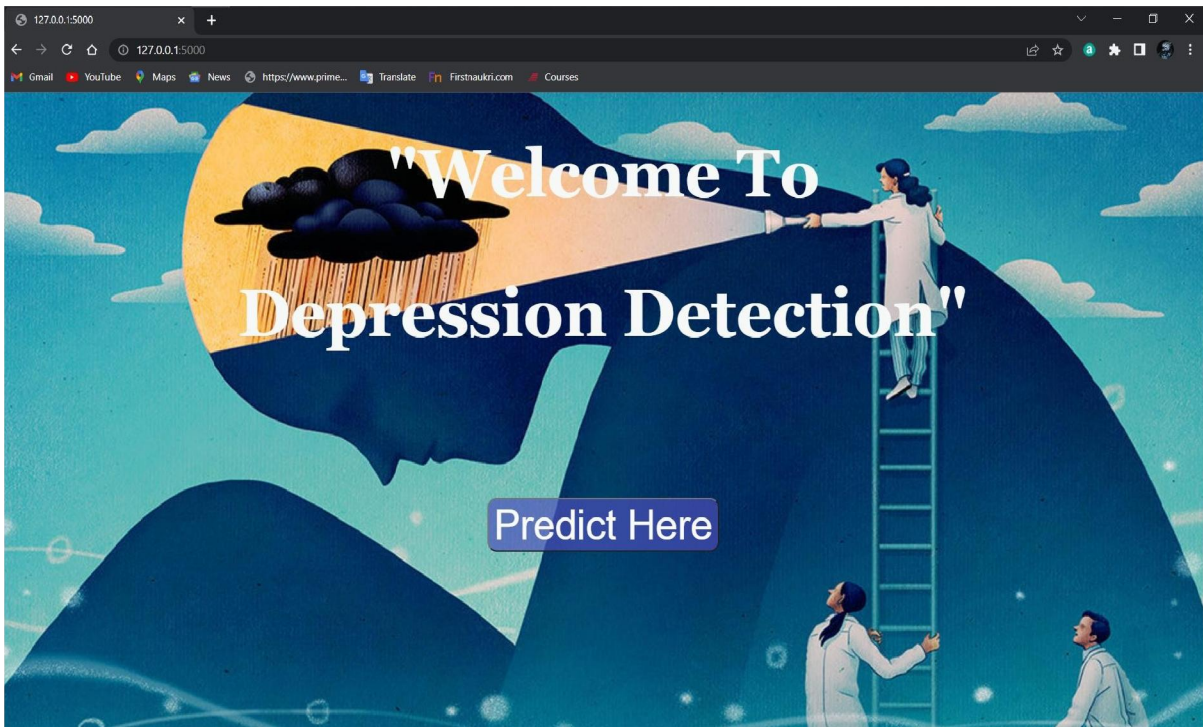


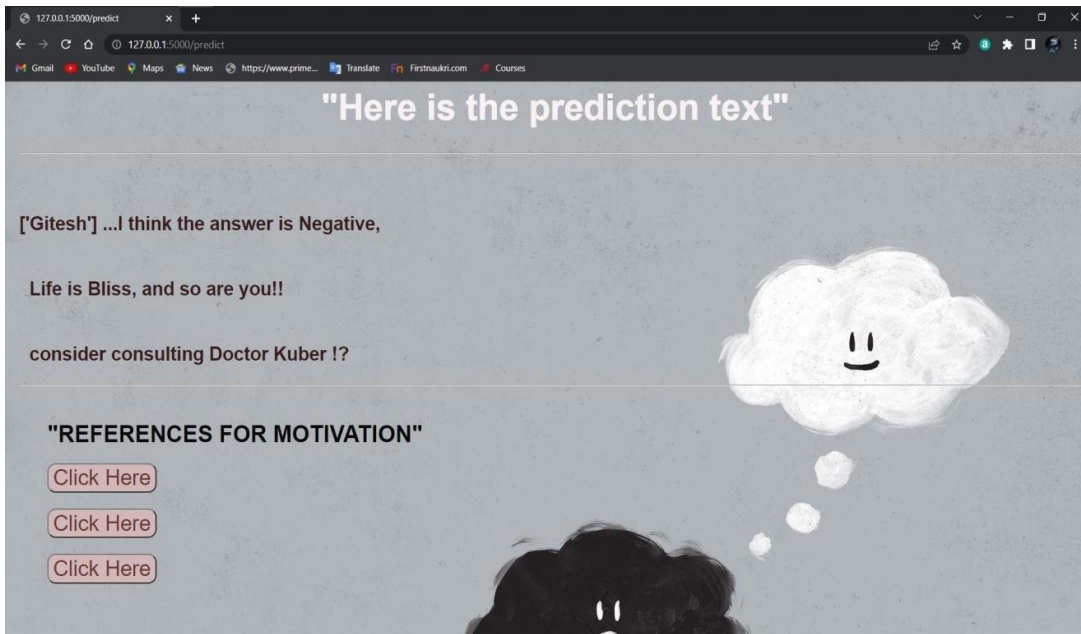
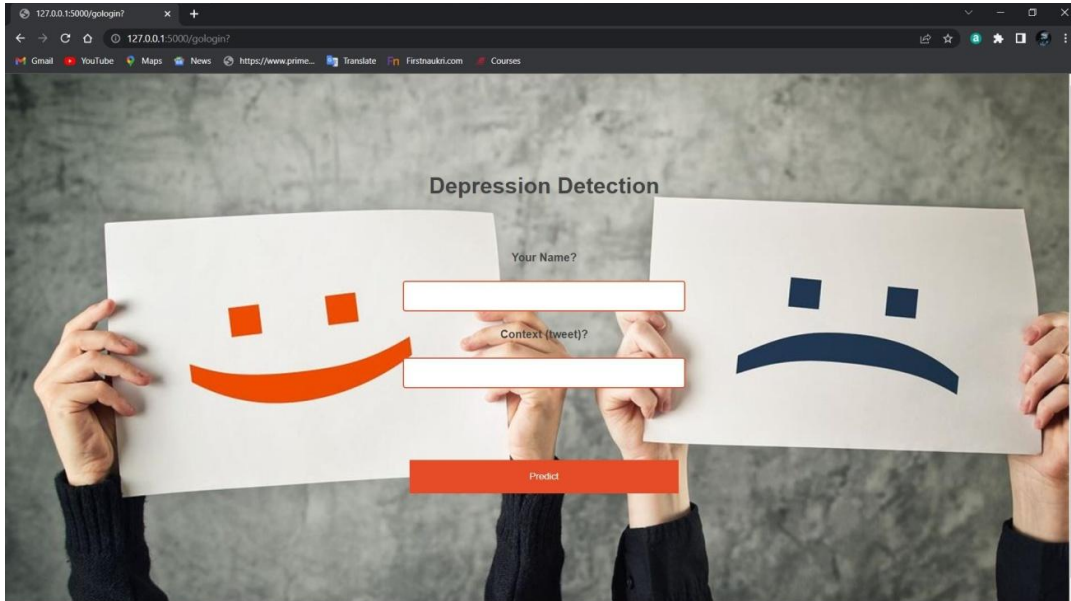
Screen View Result 3

```
Anaconda Prompt (anaconda3) - python app.py

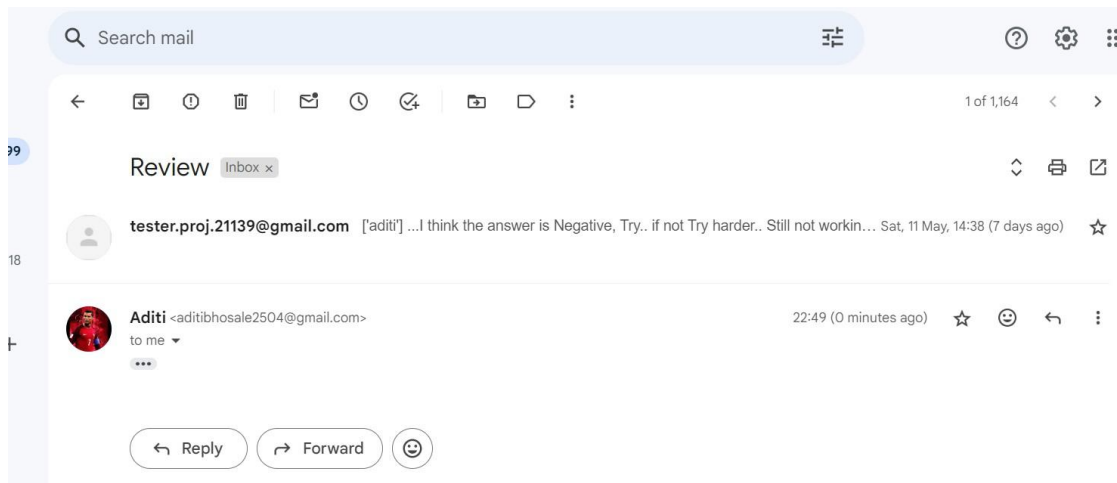
(base) C:\Users\karan>cd C:\Users\karan\OneDrive\Desktop\Depression_final\Bert-Sentiment

(base) C:\Users\karan\OneDrive\Desktop\Depression_final\Bert-Sentiment>python app.py
* Serving Flask app "app" (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: on
* Restarting with watchdog (windowsapi)
* Debugger is active!
* Debugger PIN: 534-379-669
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```





Email Notification:



VI. CONCLUSION

The planning process can help the unsuspecting buyer save his life by knowing in advance, even if the customer is not currently stressed, or the system will send some incentive pressure to customers based on the current situation. Depression level. In today's world, our body is very useful because most of us, due to our busy schedule, do not have time to meet our friends and understand their thoughts and feelings as we used to. Therefore, our body plays an important role here to prevent the decline of human life. The system will inform family members or spouse and children about the man or woman suffering from depression. Therefore, someone from his family or circle of friends can help a man or woman get out of depression.

REFERENCES

- [1] Asad, Mahmoud Pranto, S.K. Afreen and M. M. Islam, "Detecting depression by analyzing users' social media posts," 2019 IEEE International Conference on Signal Processing, Information, Communications and Systems (SPICSCON), Dhaka, Bangladesh, 2019, p. 13- 17, doi : 10.1109/SPICSCON48.2019.9065101.
- [2] K. Katchapakirin, K. Wongačikaseree, P. Yomaboot, thiab Y. Kaewpitakkun, "Facebook social media for detection of community depression in Thailand", 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhon Pathom Province, 2018 Page 1 - 6, doi:10.1109/JCSSE.2018.8457362.
- [3] P. Arora and P. Arora, "Mining Twitter data for depression detection," 2019 International Conference on Signal Processing and Communications (ICSC), Noida, India, 2019, p. 186-189, doi: 10.1109/ICSC45622.2019.8938353.
- [4] A.Ü. Hasan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee, "Emotional analysis of social network (SNS) data using machine learning to measure depression," 2017 International Conference on Information and Communication Technologies Convergence (ICTC), Jeju, 2017, P. 138. - 140, doi: 10.1109/ICTC.2017.8190959.
- [5] Deshpande and V. Rao, "Exploring stress using emotional intelligence," 2017 International Conference on Intelligent Sustainable Systems (ICISS), Pal ladam, 2017, p. 858- 862, doi: 10.1109/ISS1.2017.8389299. >
- [6] S. Jain, S.P. Narayan, R.K. Dewang, UA Bhartiya, N. Meena and V. Kumar, "Machine learning based depression analysis and suicidal ideation detection system using surveys and Twitter," 2019 IEEE Conference on Student Engineering and Systems (SCES), Allahabad, India, 2019, p. 1-6, doi: 10.1109 / SCES46477.
- [7] S. Abbaraju, V.S.S. Nadella, S. V. Alluri, "Acoustic- based depression detection using machine learning techniques," 2020 International Conference on Emerging Trends in Information Technology and Engineering (icETITE), Vellore, India, 2020, p. 1-6, doi: 10.1109/icETITE47903 .2020.394.
- [8] Yates, A., Cohan, A., and Goharian, N.: In-person risk assessment of depression and violence in online forums. arXiv preprint arXiv:1709. 01848 (2017). . Journal of Medical Internet Research 20 (5), e168 (2018). Search for suicidal thoughts on Twitter. Online Discussion 2015; 2(2): 183-188. doi: 10.1016/j.invent.2015.03.005.

- [9] S. Grover, A. Dutt and A. Avasthi, "Context of depression research in India," Indian Journal of Psychiatry, vol. 52, no. Appendix 1, p. We 178, 2010. Reddy, "Melancholy: Pain and Burden," Indian Journal of Psychological Medicine, vol. 32. No. 1 second. January 1, 2010. Available: <https://www.who.int/news-room/fact-sheets/detail/depression>
- [10] M. M. Aldarwish and H. F. Ahmad, "Investigating stress levels using broadcast communication," 2017 IEEE 13th International Symposium on Autonomous Decentralized Systems (ISADS). IB, 2017, page 14. 277-280: p.
- [11] P. Tzirakis, G. Trigeorgis, M.A. Nicolau, B.S. W. Schuller thiab S. Zafeiriou, "End- to-end multimodal emotion recognition using deep neural networks," IEEE Journal of Selected Topics in Signal Processing, vol. 11. Tsis muaj. Nplooj 8. 1301-1309, 2017 O. Available: <https://emedicine.medscape.com/article/805459-overview>