IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DEPRESSION DETECTION BY ANALYZING SOCIAL MEDIA POST OF USER: Research Paper

[1]Karan Pagare, [2]Gitesh Bhole, [3]Pooja Mankar, [4]Rucha Mahajan, [5]Prof.Abhay Gaidhani, [6]Dr.Mangesh Ghonge, [7]Dr.Amol Potgantwar

[1,2,3,4]Student, [5,6,7]Professor, [1,2,3,4,5,6,7]Computer Engineering,
[1,2,3,4,5,6,7]Sandip Institute of Technology and Research Centre, Nashik, Maharashtra, India.

Abstract: *Today, the problem of early detection of depression is one of the most important in psychology. Mental health problems are often among the most important health stressors in the world, with over 300 million people currently affected by depression alone. As large amounts of male or female signups are generated on social media platforms, researchers are increasing the use of evidence-gathering devices to determine if this content material can be used to uncover mental health issues in users. Depression is a disease that poses a major problem in our society and is a continuing concern, according to researchers around the world. With ubiquitous computing devices such as smartphones, predicting depressive moods remains an open question. Social media testing is often implemented to address this issue. In this article, a depression rating and suicidal ideation detection system were proposed to predict suicidal acts supporting the magnitude of depression. To do this, expert and well-established classifiers were used to discriminate whether someone is depressed or not, using competencies from their sporting activity within positions. Similar tool algorithms are used to train it and classify it into different levels of depression on a scale of 0-100%. In depression or not, the use of Art Machine Learning algorithms is a predictive method for the early detection of depression or unusual mental illnesses. The main contribution of this exam is the exploration of a competency network and its implications for the recognition of the degree of depression. This system aims to gain an in-depth understanding of the model used to categorize users with depression by understanding some cases where male or female undergraduate labels are examined to uncover postgraduate labels. By combining all of the post tag category possibilities, you can create temporary post profiles that are then used to categorize customers with depression. This paper shows that there are fluid versions in the posting patterns between depressed and non-depressed clients, as represented by the combined odds of the posting tag category. Natural Language Processing (NLP), categorized the use of the BERT set of regulations to find out depression probably in a greater on hand and inexperienced way.*

Index Terms **- Machine Learning, NLP, BERT Algorithm, Depression, Classification, Social Media Post.**

I. INTRODUCTION

Nowadays the hassle of early depression detection is one of the maximum critical withinside the area of psychology. Depression is moreover a mean intellectual problem. In today's world, the stresses of lifestyle sports in one's lifestyle ought to develop the possibility of depression. Over 350 million people globally are bothered by depression, which is prepared 5% of the whole population. Close to 8,00,000 people die because of suicide every 12 months and its miles are statistically the second one important purpose for loss of lifestyle amongst people 15–29 years old. At the equal time, the number one form of suicide is related to depression. Recent researches display that depression is likewise the primary purpose of incapacity and intense somatic diseases. The proliferation of the net and communique technologies, especially online social networks have rejuvenated how people have interacted and talked with every unique electronically. The packages which encompass Facebook, Twitter, Instagram, and alike now, now no longer nice host the written and multimedia content material however furthermore provide their customers precise feelings, emotions, and sentiments approximately a subject, hassle, or a problem on line. On one hand, that is terrific for customers of social networking websites to overtly and freely contribute and reply to any hassle matter online; on the alternative hand, it creates possibilities for people of America of a person who reacted to a subject in a particular manner. To offer such a notion, device studying strategies ought to probably provide a few precise skills that could help in inspecting the precise styles hidden in online communique and way them to reveal the intellectual us Moreover, there can be a developing frame of literature addressing the characteristic of social networks withinside the shape of social relationships which incorporates breaking up relationships, intellectual illness ('depression', 'anxiety', 'bipolar' etc.), smoking, and eating relapse, sexual harassment and for suicide ideation. Younger adults, racial/ethnic minorities, crucial staff, and unpaid character caregivers advocated having intimate disproportionately worse reputation outcomes, raised substance use, and improved unstable cerebration. Youth is outlined as age fifteen to twenty-4 years, and it consists of center and past due to adolescence. It's characterized through the way of the approach of present-day modifications in physical, psychological, and social dimensions. For healthful increase and development, young adults need to have a nice manner of happiness, love, action, and independence and to have a reason in lifestyle. Throughout this natural way stage, many forms of conduct rectangular degree evolved that my purpose every normalcy or reputation malady. Regardless of what you almost definitely did currently on your mobile phone or PC, it's obvious that social media changed properly into a concern. Did you capture up with buddies on Facebook, put up snapshots of your cat, or video of your totter strolling for 1st time on Instagram? probably a Twitter hyperlink introduced you here. These days oldsters tend to particular their emotions, reviews accomplice levels reveal their ordinary lives thru the growth of social media systems like Twitter, Facebook, and In. Instagram. These expressions are rectangular degrees commonly thru snapshots, videos, and trendy posts. In this study, we intend to investigate Social Media posts to come across any elements that could mirror the depression of applicable Social Media customers. Various gear for studying strategies are hired for such purposes. Considering the crucial element purpose of this study, the subsequent are next studies require situations addressed withinside the paper. We will be predisposed to tend to intend to make use of gear studying strategies and algorithms for depression detection on social media posts of customers.

NLP (Natural Language Processing):

The paintings defined in this paper belong to the vicinity of Natural Language Processing (NLP) and the textual content's splendor is specific. The origins of textual content splendor responsibilities may be determined in early studies to mechanically categorize files primarily based totally on a statistical evaluation of specific clue phrases in 1961. Later, comparable studies desire to result in rule-primarily based surely textual content splendor structures like CONSTRUE in 1990, and finally, the vicinity started to shift increasingly tool studying algorithms in a few unspecified time withinside the destiny of the 3 hundred and sixty-5 days of 2000. In addition to textual content categorization, tool assessment will become furthermore the usage of strain in outstanding textual content primarily based totally on responsibilities like sentiment evaluation, that's focused on extracting critiques and sentiment from textual content files. It will become first be finished in aggregate with the tool studying to find out high-quality or horrible critiques in film critiques and end up then prolonged to outstanding evaluation domains, in addition to specific regions like social media tracking and preferred evaluation of purchaser attitudes. More recently, a deep assessment has been finished for textual content splendor in addition to its more not unusual location vicinity utilization withinside the photo graph splendor. State-of-the-art work outcomes in numerous textual content-primarily based responsibilities that could, for example, be finished through switch studying strategies like Universal Language Model Fine-tuning (ULM Fit) and the Google studies task Bidirectional Encoder Representations from Transformers (BERT) for the education of language representations, which incorporates ULM Fit and numerous outstanding The code of BERT and numerous pre-expert models.

II. Literature Review

Instrumental opportunities of studying the conduct of customers in social networks are actively developing. In particular, strategies of computational linguistics are efficiently utilized in studying the posts from social media.

1) A records-analytic-primarily based totally version to hit upon melancholy of any individual is proposed withinside the paper. The records are gathered from the customers' posts on famous social media websites: Twitter and Facebook. In this research, device studying is used to manner the scrapped records gathered from SNS (Social Networking Sites) customers. Natural Language Processing (NLP), labeled the usage of Support Vector Machine (SVM) and Naïve Bayes set of rules to hit upon melancholy probably in an extra handy and greenway. [1]

2) The research employs Natural Language Processing (NLP) strategies to increase a melancholy detection set of rules for the Thai language on Facebook in which human beings use it as a device for sharing opinions, feelings, and existence events. [2]

3) The fitness tweets are analyzed for Depression, Anxiety from the blended tweets via way of means of the usage of Multinomial Naive Bayes and Support Vector Regression (SVR) Algorithm as a classifier in paper [3].

4) In the paper, researchers gift a way to discover the melancholy degree of someone via way of means of looking at and extracting feelings from the text, the usage of emotion theories, device studying strategies, and herbal language processing strategies on unique social media platforms. [4]

5) The paper, pursuits to use herbal language processing on Twitter feeds for engaging in emotion evaluation specializing in melancholy. Individual tweets are labeled as impartial or negative, primarily based totally on a curated phrase listing to hit upon melancholy tendencies. In the manner of sophistication prediction, a guide vector device and Naive-Bayes classifier had been used. The consequences had been offered the usage of the number one category metrics inclusive of F1-score, accuracy, and confusion matrix. [5]

6) The paper, proposes a melancholy evaluation and suicidal ideation detection system, for predicting suicidal acts primarily based totally on the extent of melancholy. Real-time records changed into gathered withinside the shape of Tweets and Questionnaires. Then, category device algorithms are used to teach and classify it in 5 degrees of melancholy relying on severity. [6]

7) Yates et al. used a neural community version to show the dangers of self-damage and melancholy primarily based totally on posts from Reddit and Twitter and confirmed the excessive accuracy of this diagnostic approach. The authors suggest that proposed strategies may be used for large-scale research of intellectual fitness in addition to for scientific treatment. [8]

8) O'Dea et al. tested that Twitter is gradually researched as a way for spotting mental well-being status, inclusive of melancholy and suicidality withinside the population. Their research found out that it's miles workable to understand the extent of fear amongst suicide-associated tweets, using each human coder and a programmed device classifier.[10]

There is a severe and developing variety of methodologies and strategies for detection of the melancholy degree from the posts on Social Media networks. In our study, we consolidate a technical description of strategies implemented for melancholy identity the usage of the Natural Language Processing approach labeled the usage of the BERT set of rules to hit upon melancholy. The framework is created from Data pre-processing step, the Feature extraction step following the Machine Learning classifiers, the Feature evaluation of the records, and the Experimental Results.

III. PROBLEM STATEMENT

Depression has been verified to have an impact on the language of individuals. To expand a software to investigate and come across despair of social media posts of clients via tool studying techniques. This challenge aims to use natural language processing, tool studying techniques, and neural network architectures to build, tune, and have a look at models that classify social media Post of Users as "depressed" or "non-depressed

IV. OBJECTIVE

The targets are as follows:

1. System will constantly hold on tracking the posts and chats of users. And if it detects the bad notion sort of conduct then the device will mechanically put up the high-quality put up on his/her wall primarily based totally on the extent of depression.
2. Help the individual to pop out of depression.

V. METHODOLOGY

Machine Learning Classification Techniques used for the mode

1. BERT Algorithm:

"BERT stands for Bidirectional Encoder Representations from Transformers. It is designed to pre-educate deep bidirectional representations from the unlabeled textual content with the beneficial aid of the usage of collectively conditioning on each left and proper context. As a result, the pre-expert BERT version may be fine-tuned with in reality one more output layer to create cutting-edge fashions for a large sort of NLP tasks." We advanced a Bidirectional Encoder Representations from Transformers (BERT)-primarily based virtually version, it is a brand-new language example version as defined in. As the choice suggests, it changed into designed to pre-educate deep bidirectional representations that may be fine-tuned with a further output layer. For this project, this outputs layer - a pooled output - changed into used for the binary shape of the comments. From the various pre-expert fashions available, we determined the English-language uncased (all lowercase earlier than tokenization) version of BERT, as case records are not particularly critical to the venture of social media commentary type.

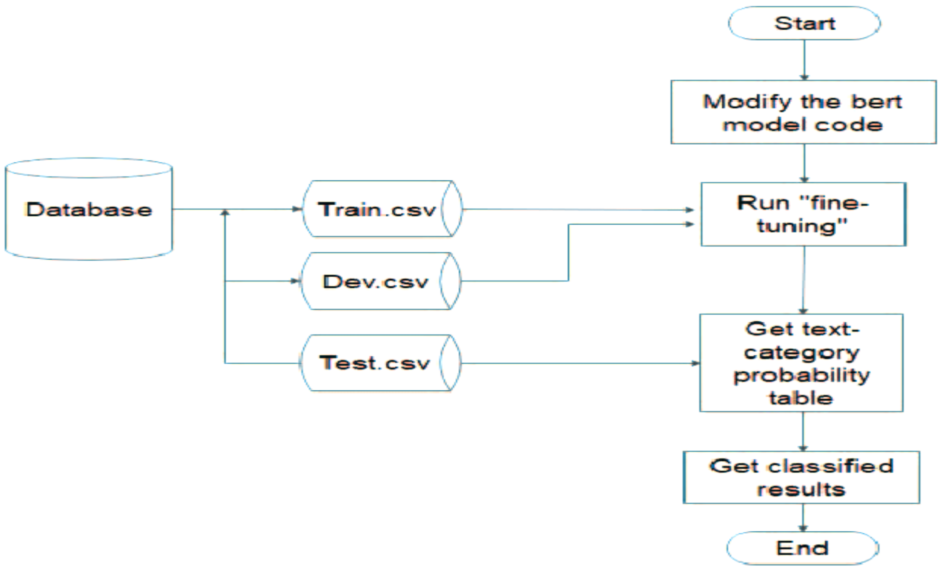

Fig.1 BERT Algorithm

2. Relevant Mathematics Associated with The Project:

System Description:

S= I, O, F, DD, NDD, Failure, Success

Where,

S=System
I= Input
O=Output
F=Failure
S=Success

I is Input of system

Input I = set of Inputs

Where,

I= {Users Social media posts}
F is Function of system
F = set of Function

Where,
F1= {Input Dataset}
F2= {Json to CSV Conversion}
F3={Pre-processing}
F4={Cleaning}
F5= {Train test split}
F6= {Sentiment Dictionary}
F7= {Classifier (BERT Algorithm)}
F8={Tokenization}
O is Output of system
Output O1= {Depression detection}

- **Success Conditions:** Product working Smoothly. depression detection successfully.
- **Failure Conditions:** if internet connection Unavailable.

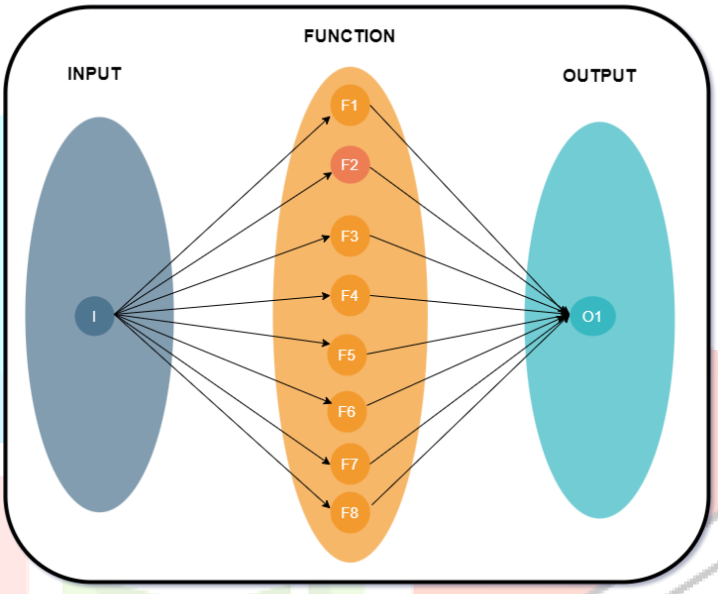

Fig.2 Venn Diagram

Where ,
I = {Users Social media posts}
F1= {Input Dataset}
F2= {Json to CSV Conversion}
F3={Pre-processing}
F4={Cleaning}
F5= {Train test split}
F6= {Sentiment Dictionary}
F7= {Classifier (BERT Algorithm)}
F8={Tokenization}
Output O1 = {Depression detection}

VI. System Architecture

Depression is a severe challenge in private and public health. One of the important solutions to this hassle is an in-intensity have a take examine an individual's behavioural attributes. These attributes are available on numerous social networking websites collectively with Facebook, Twitter, Instagram, etc. Social networking platform is a high-quality way to understand a character's behaviour, questioning style, mood, egoistic networks, evaluations, etc. The use of social networking websites is developing, specifically among most of the several more youthful generations. Human beings on social media specific their feelings, regular sports activities, evaluations about numerous topics, etc. So social networking websites are used as screening gadgets to assume depression ranges. These social networking structures offer a character's experiences, evaluations, socialization, and personality. The superior method of evaluation of the affected character is not so relevant but the use of consumer-generated content material fabric material cloth on social media posts lets one to are looking for the highbrow health ranges and depression of a particular individual. Our challenge purpose is to extract facts from social media posts and thru a manner of the method of getting clean expertise of a character's behavioural attributes and attempted questionnaires, depression ranges of the consumer are predicted. A quantitative appearance is achieved to educate and check numerous gadgets studying classifiers to determine whether or not or now no longer or now no longer or no longer a social media positioned up of the consumer is depressed, from posts initiated thru manner of the method of the consumer or his/her sports activities on social media. The following strength of mind illustrates the depression detection of using the hobby and content material fabric material cloth capabilities kind model. First, all tweets for depressed and non-depressed money owed, similarly to facts of consumer money owed and sports activities collectively with the sort of followers, the amount of following, time of posts, the sort of mentions, and amount of reposts, are retrieved. Next, all posts of an account are assembled in an unmarried document. Text pre-processing is achieved on all documents. First, a corpus is created and posts in each document are tokenized. BERT Classification Algorithm can be used.

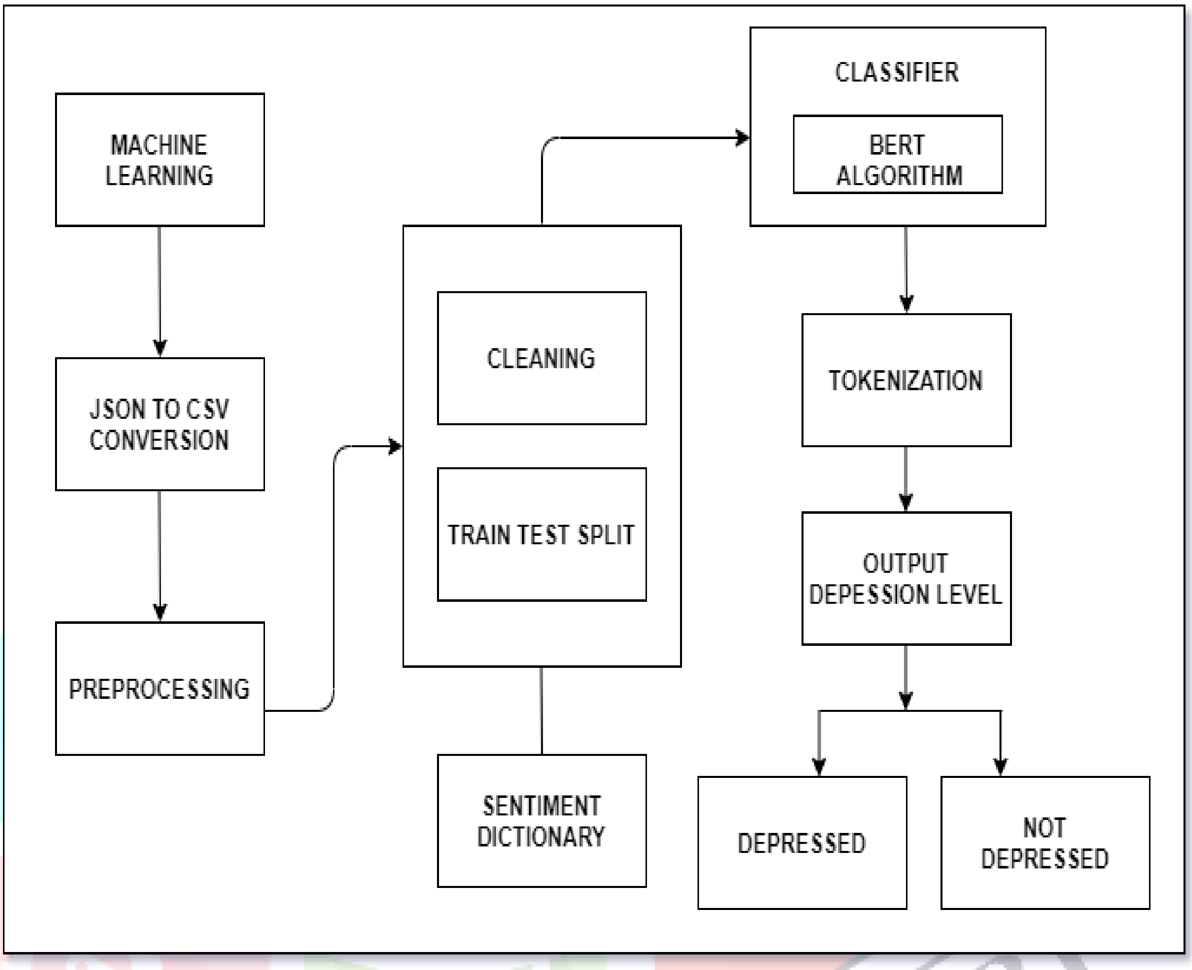

Fig.3 System Architecture

- **Exiting System**

The already-gift tool provides a miles easy flowing way to determine the depression level of clients using the Naïve Bayes algorithm. The extraction of textual statistics is completed via the extraction beauty from Facebook with the help of the Facebook Graph API. After extraction, the statistics are pre-processed. The missing or repetitive attributes are taken care of in pre-processing. Techniques like tokenization, lower case conversion, and word stemming and terms removal are used for Pre-processing of statistics. In the proposed tool, regular clients' Facebook put up model can find out whether or not or now no longer he/she is depressed or now no longer. But most effective studying posts won't supply accurate consequences so we moreover observe the comments via the client and his friends and his chats are also analyzed due to the fact the client will truly percentage his depression collectively alongside together along with his friend. On basis of these analyses, the clients can be categorized as pressured and non-pressured.

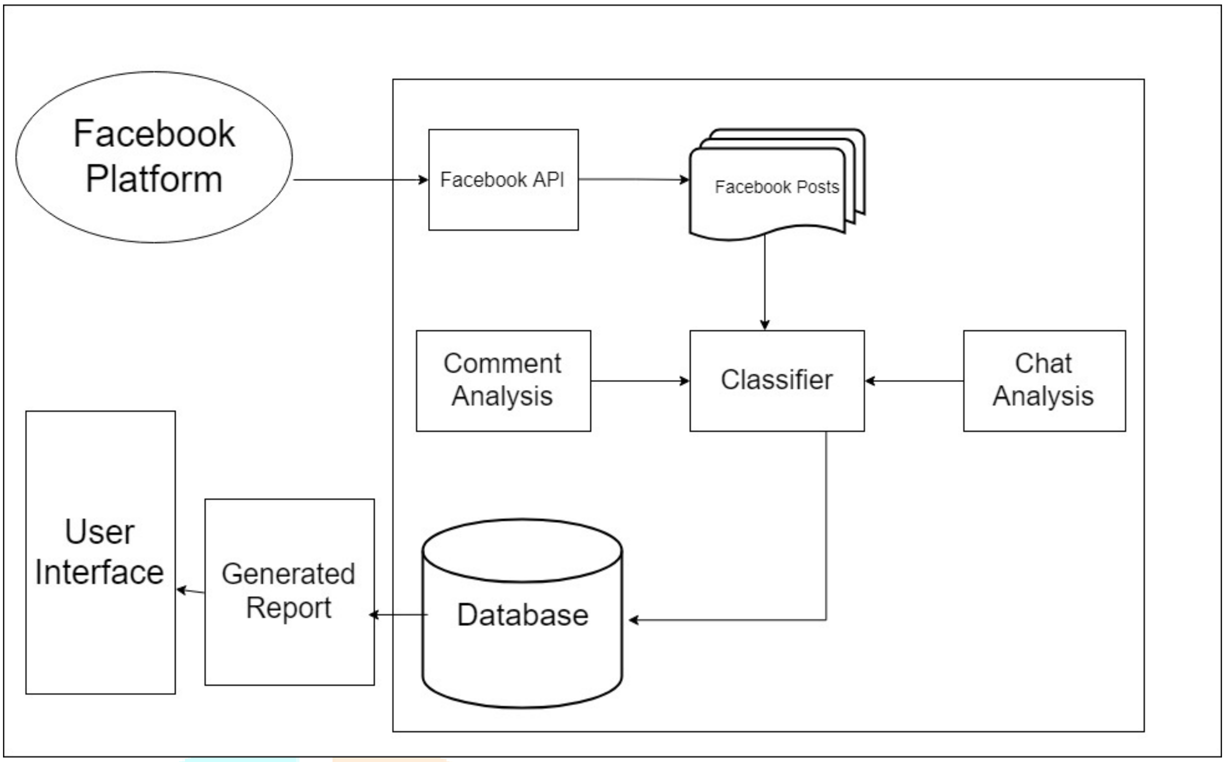

Fig.4 Existing System Architecture

VII. Implementation

- **Module Split-Up:**
  - a) Data Processing (Module-I)
  - b) Data Training, Testing (Module-II)
  - c) Creating Frontend (Module-III)

a) Data Processing (Module-I):

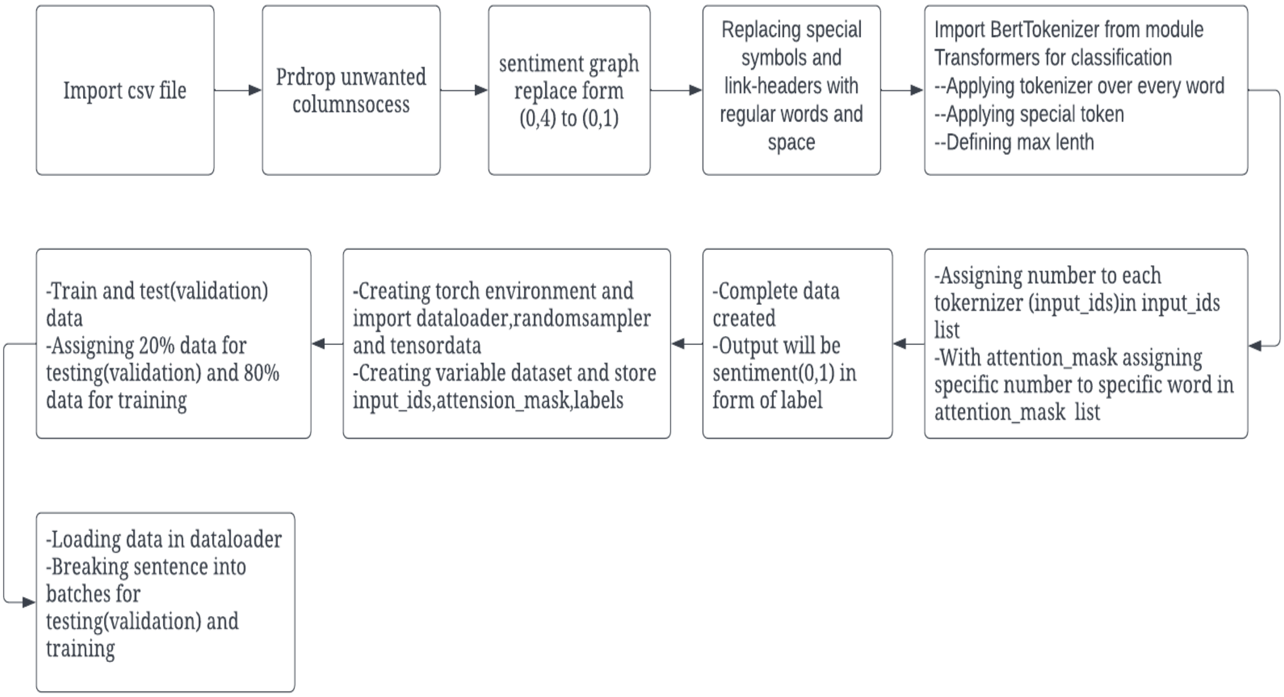

Fig.VII-a-1 Data Processing (Module-I)

- **Screen-View Result:**

```
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....

Fig.VII-a-2 Module-I

```
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. ...

Fig. VII-a-3 Module-I

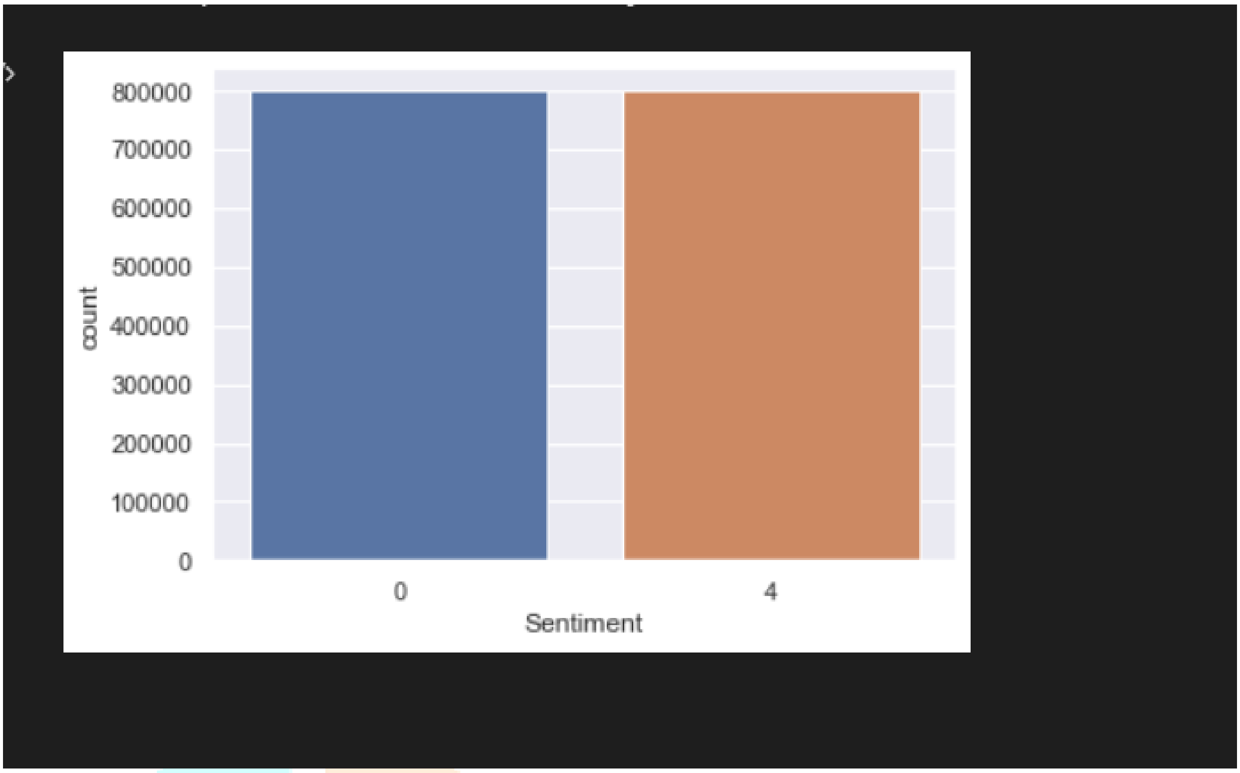

Fig. VII-a-4 Module-I

Fig. VII-a-5 Module-I

```python
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
```

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

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

```
(40000, 10000)
```

Fig. VII-a-6 Module-I

b) Data Training, Testing (Module-II):

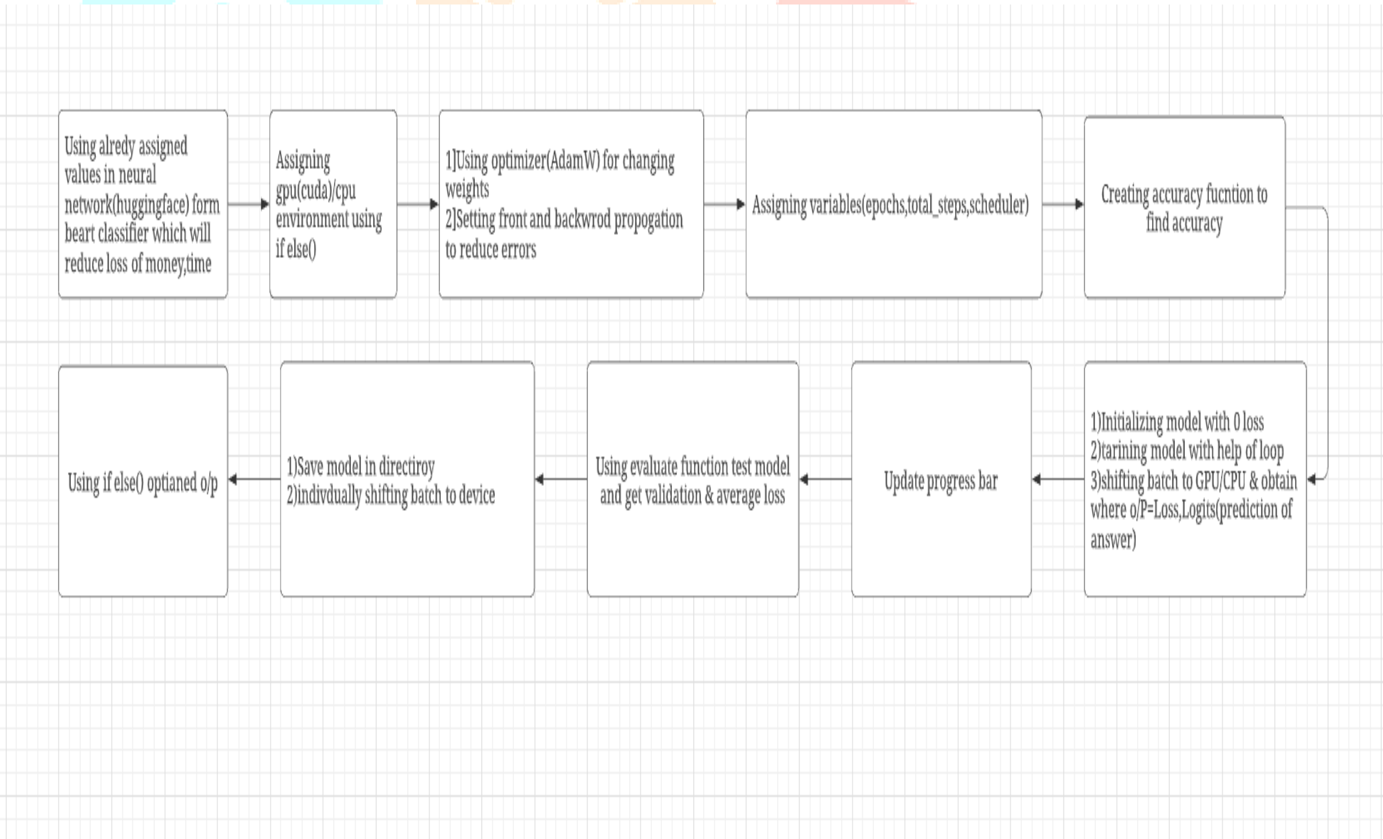

Fig. VII-b-1 Module-II

- **Screen-View Result:**

Module II

```
model = BertForSequenceClassification.from_pretrained(
    'bert-base-uncased',
    num_labels = 2,
    output_attentions = False,
    output_hidden_states = False)
```

```
device = torch.device('cuda' if torch.cuda.is_available() else 'cpu') # GPU/CPU env set for cpu  device = torch.device('cpu' )
model.to(device)


print(device)
```

```
cuda
```

```
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.
```

Fig. VII-b-2- Module-II

Defining Accuracy and Evaluate Function

```python
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)
```

Python

```python
def evaluate(dataloader_test):
    model.eval() # eval mode
    loss_val_total = 0
    predictions,true_vals = [],[]
    for batch in dataloader_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
```

Fig. VII-b-3- Module-II

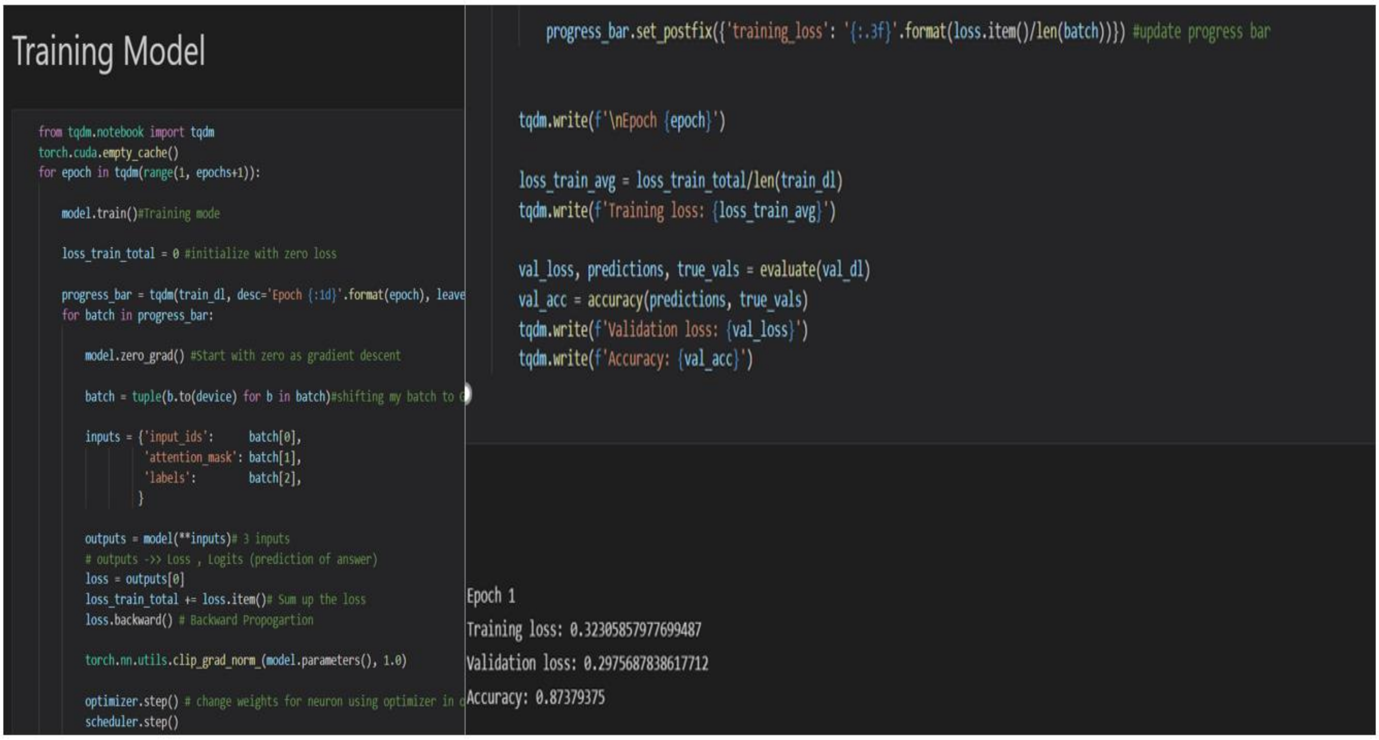

Fig. VII-b-4- Module-II

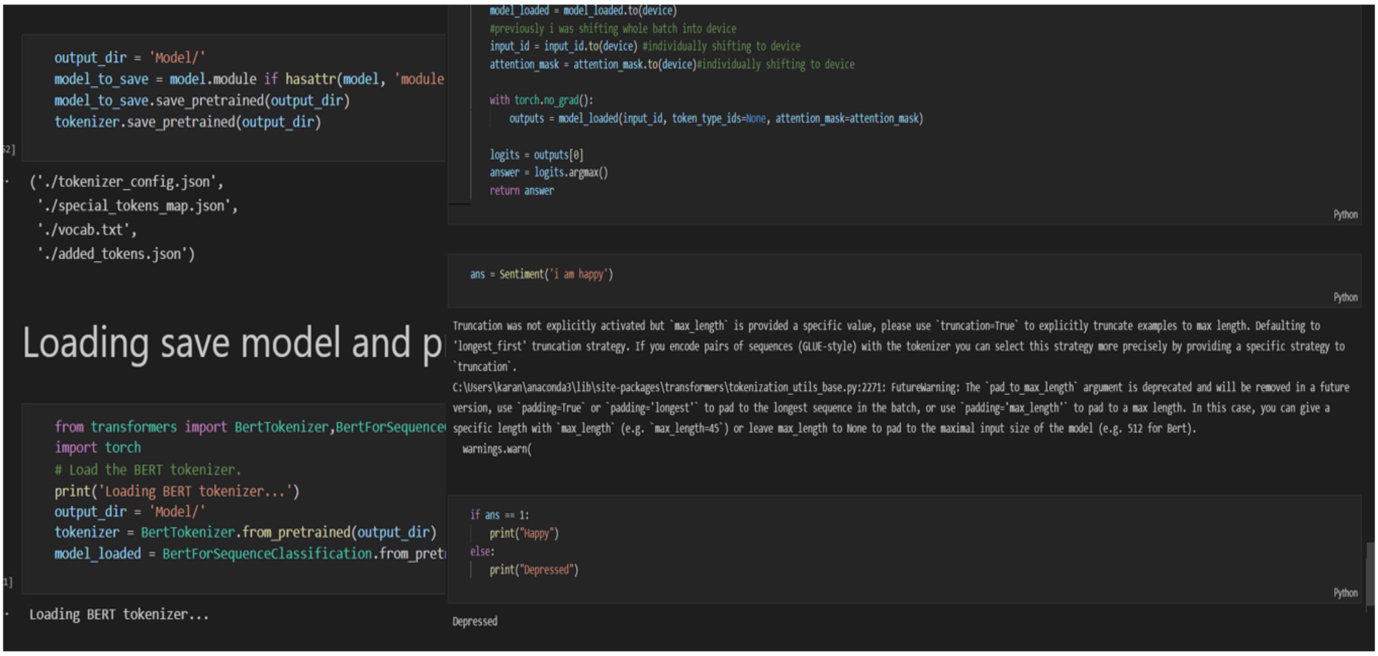

Fig. VII-b-5- Module-II

c) Creating Frontend (Module-III):

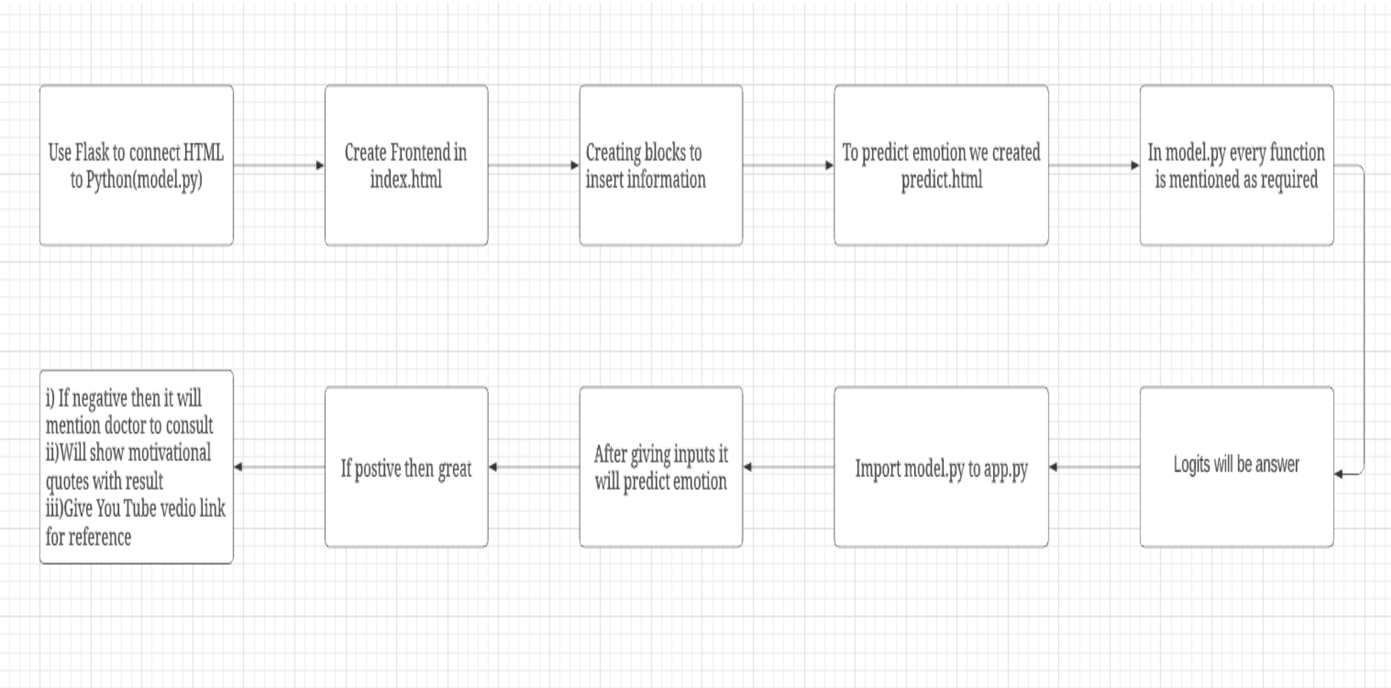

Fig. VII-c-1- Module-III

- **Screen-View Result:**

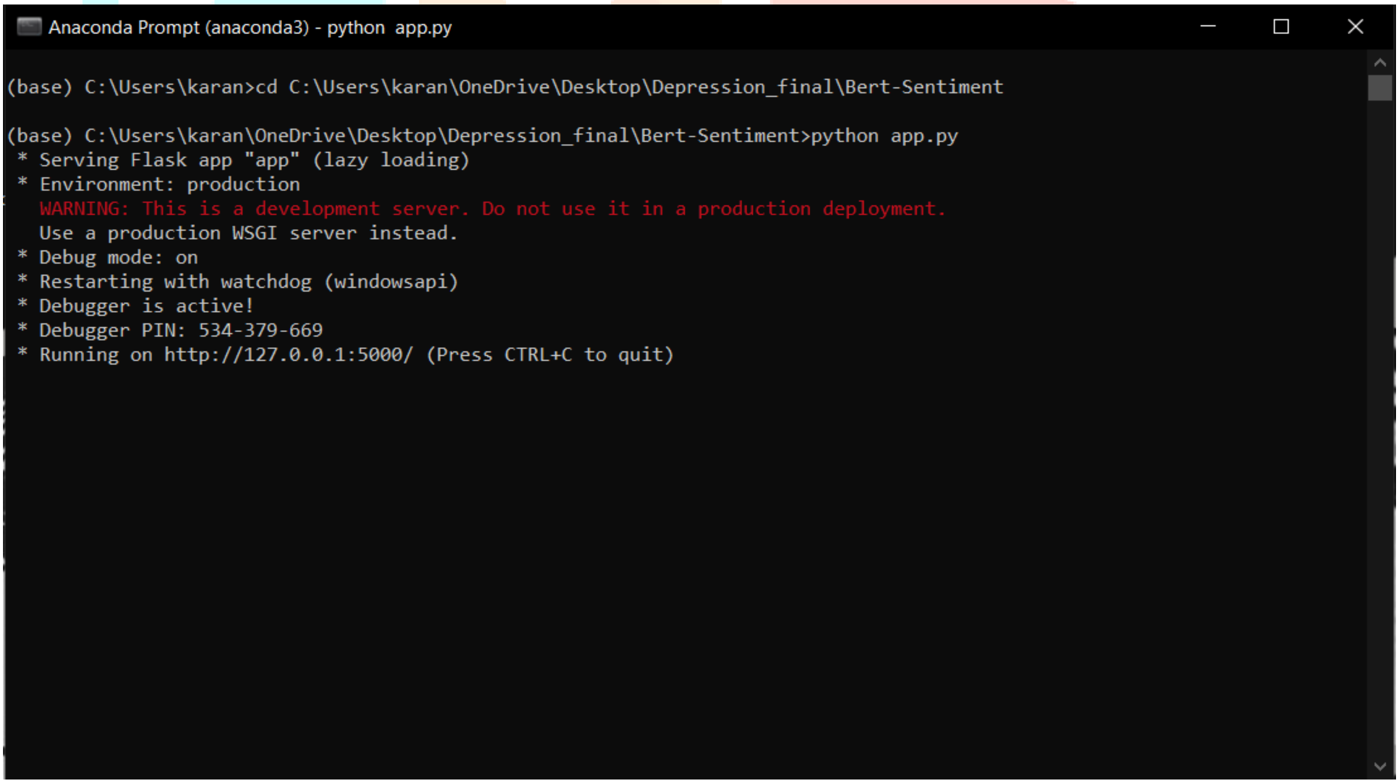

Fig. VII-c-2- Module-III

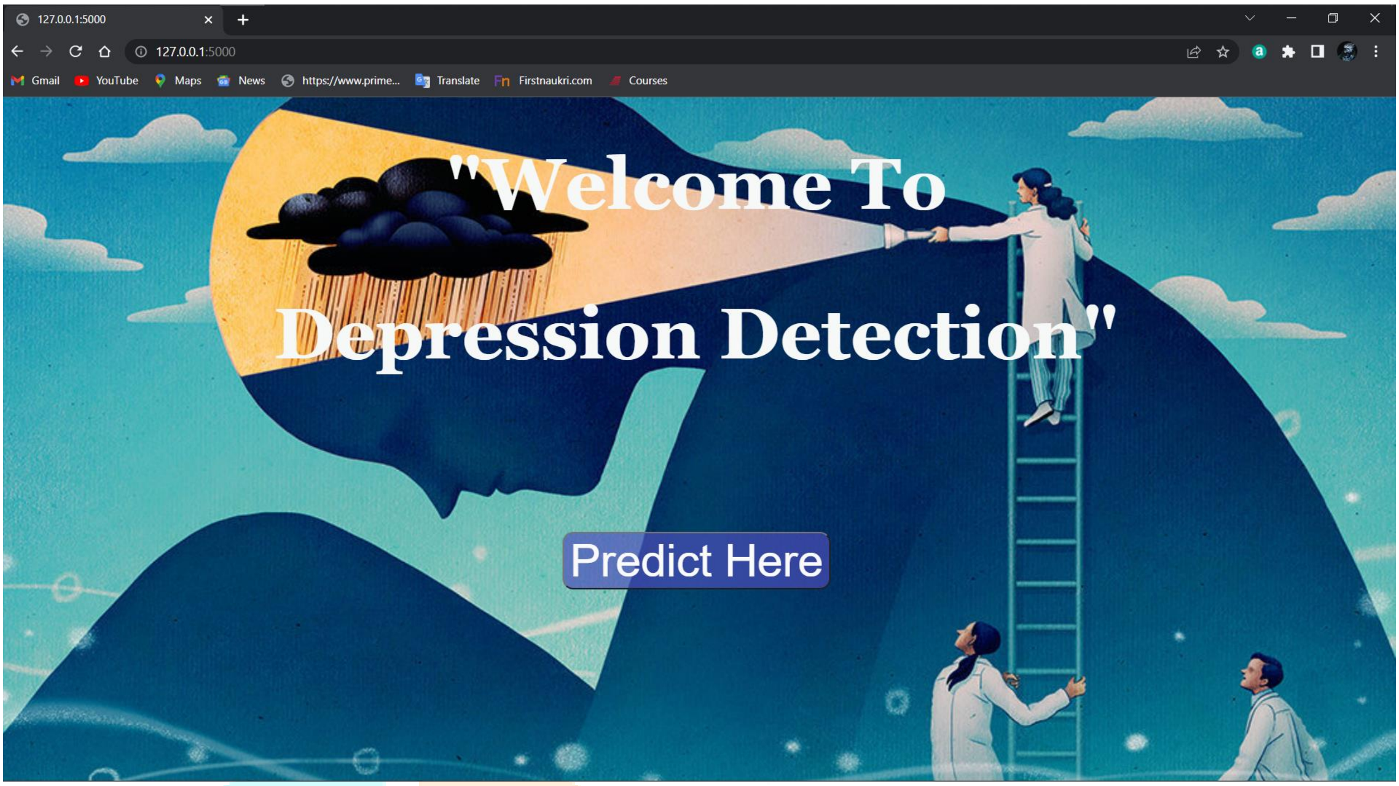

Fig. VII-c-3- Module-III

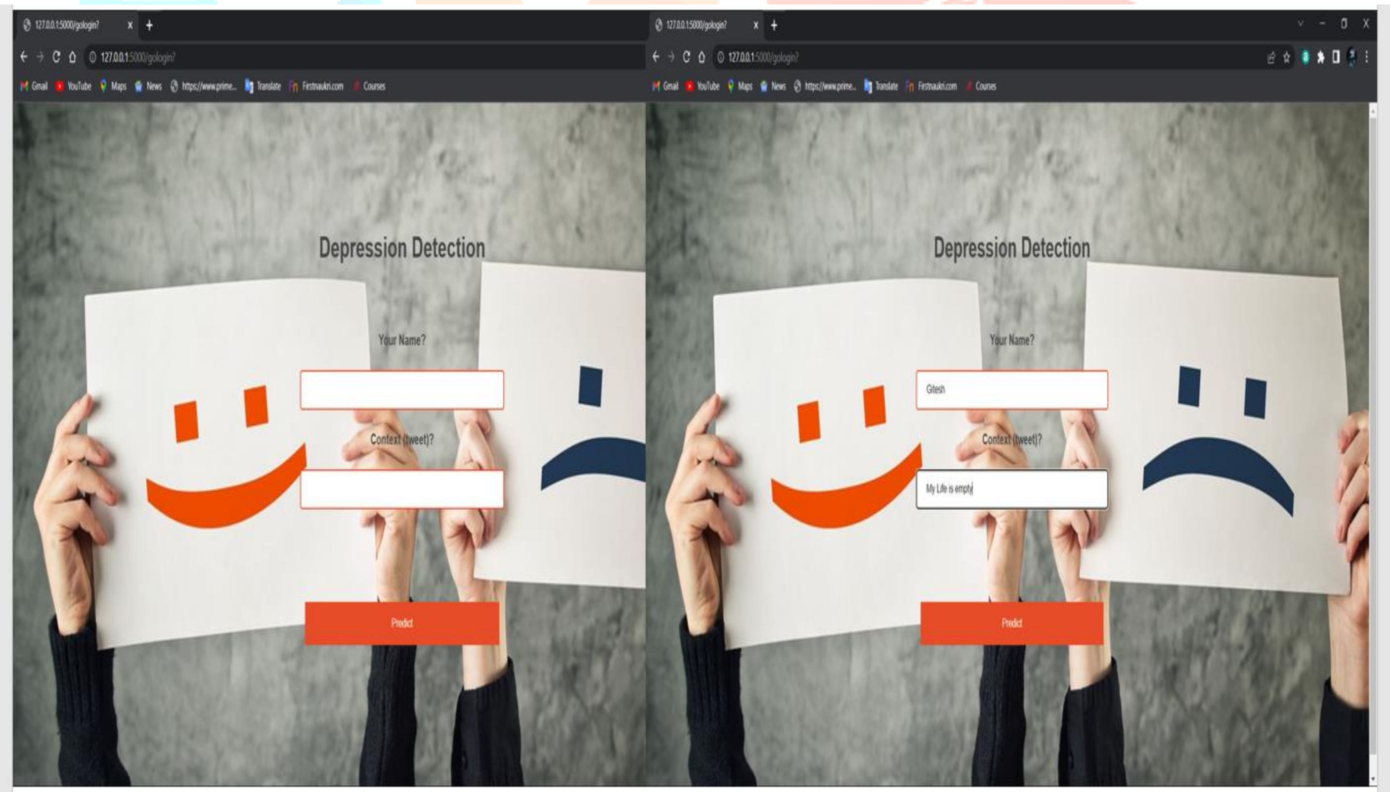

Fig. VII-c-4- Module-III

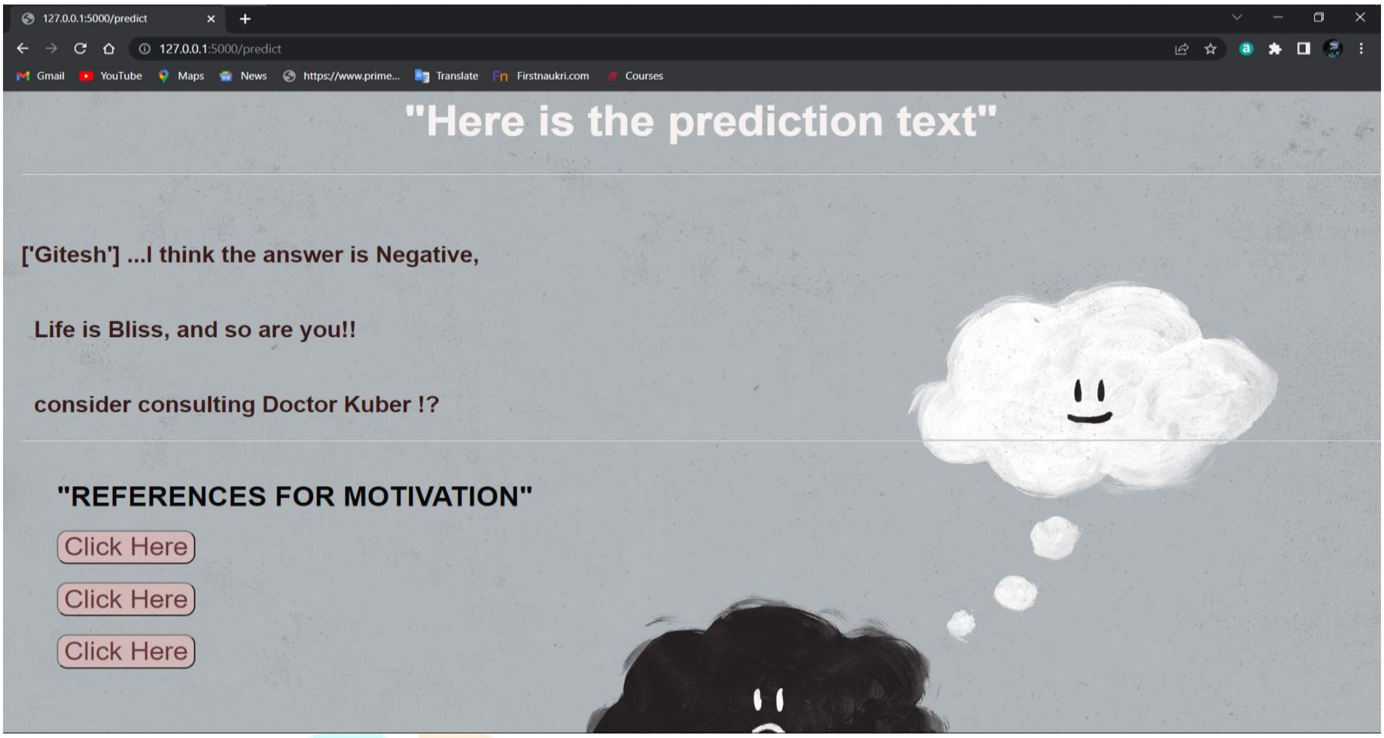

Fig. VII-c-5- Module-III

VIII. Conclusion

The proposed device can also moreover help the suspected patron to keep his/her life, thru manner of the approach of know-how in advance whether or not or now no longer or now not or now not the customer is depressed or possibly the device will deliver some motivational posts to the customer based mostly on the quantity of his depression. We give up the device is probably very useful in today's world wherein most humans don't have time to satisfy our friends, percent their thoughts and feelings as we achieved in older days due to busy schedules. So, our device plays a critical feature proper right here to avoid any unwanted human loss. The device will inform their very non-public circle of relatives' members or spouses and youngsters regarding the state of affairs of a depressed man or woman. So that each own circle of relatives or pal circle will help the man or woman to come out of depression.

IX. Reference

[1] N. A. Asad, M. A. Mahmud Pranto, S. Afreen and M. M. Islam, "Depression Detection by Analyzing Social Media Posts of User," 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), Dhaka, Bangladesh, 2019, pp. 13- 17, doi: 10.1109/SPICSCON48833.2019.9065101.

[2] K. Katchapakirin, K. Wongpatikaseree, P. Yomaboot and Y. Kaewpitakkun, "Facebook Social Media for Depression Detection in the Thai Community," 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhonpathom, 2018, pp. 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 Communication (ICSC), NOIDA, India, 2019, pp. 186-189, doi: 10.1109/ICSC45622.2019.8938353.

[4] A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq and S. Lee, "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression," 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2017, pp. 138- 140, doi: 10.1109/ICTC.2017.8190959.

[5] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, 2017, pp. 858-862, doi: 10.1109/ISS1.2017.8389299.

[6] S. Jain, S. P. Narayan, R. K. Dewang, U. Bhartiya, N. Meena and V. Kumar, "A Machine Learning based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter," 2019 IEEE Students Conference on Engineering and Systems (SCES), Allahabad, India, 2019, pp. 1-6, doi: 10.1109/SCES46477.2019.8977211.

[7] B. Yalamanchili, N. S. Kota, M. S. Abbaraju, V. S. S. Nadella and S. V. Alluri, "Real-time Acoustic based Depression Detection using Machine Learning Techniques," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-6, doi: 10.1109/icETITE47903.2020.394.

[8] Yates, A., Cohan, A., and Goharian, N.: Depression and self-harm risk assessment in online forums. arXiv preprint arXiv:1709.01848 (2017).

[9] Seabrook, E.M., Kern, M.L., Fulcher, B.D., and Rickard, N.S.: Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. Journal of Medical Internet Research 20 (5), e168 (2018).

[10] O'Dea B, et al. Detecting suicidality on Twitter. Internet Interv. 2015;2(2):183–188. doi: 10.1016/j.invent.2015.03.005.