Department of Electrical and Computer Engineering

Senior Design Project

Depression analysis using Deep learning and Natural language Processing

Submitted By

Name	ID
Faysal Ibna Shahadat	1512243043
Md Khorshed Alam	1612834043
Sajid Ansari	1731802642

Submitted To

Dr. Mohammad Monirujjaman Khan

Associate Professor

Department of ECE

1. Background Study

Depression, often known as depressive disorder, is a very prevalent illness. According to the World Health Organization (WHO), there are more than 300 million people worldwide who suffer from depression. Depression can have a significant impact on one's well-being and ability to function at job, school, and in the family, and can even lead to self-harm. In adulthood, adolescent depression is linked to mood disorders and serious mental illness. Suicide is the fourth highest cause of mortality in 15–19-year-olds, according to WHO, with about 0.8 million people dying each year. Five of the top major disorders that cause disability or incapacity are mental illnesses, with depression being the most common. As a result, the disease burden associated with depression is enormous. Depression affects about 5% of the adult population throughout cultures, and 20% of the population in its lesser forms (partial symptoms, moderate depression, and probable depression). The middle-aged group is the most vulnerable among adults. In addition, depression is becoming more common over the world, with an 18 percent increase between 2005 and 2015. However, early professional intervention can improve mental symptoms (e.g., absence of self-confidence and rumination) and resolve somatic problems (e.g., gastrointestinal problems and sleeping disorders) in most of the cases.

Depression can make a person suffer severely and cause them to perform poorly at job, at school, and in social situations. Regardless of what you nearly did on your phone or laptop computer recently, it appears that social media was involved. Did you use Facebook to check up with pals, or Instagram to share images of your cat or a video of your toddler walking for the first time? Perhaps you arrived here via a Twitter link. People nowadays use a growing number of social media platforms such as Twitter, Facebook, and Instagram to express their emotions, thoughts, and reveal their daily lives.These expressions square measure generally through photos, videos and primarily through text.

Early recognition of depressed symptoms, as well as assessment and therapy, can greatly improve the odds of controlling symptoms and the underlying condition, as well as minimizing negative consequences for one's well-being and health, as well as personal, economic, and social life. However, detecting depression symptoms is difficult and time-consuming. Current methods rely mostly on clinical interviews and questionnaire surveys conducted by hospitals or agencies, with psychological evaluation tables used to create mental disorder predictions.

Depression analysis is the method of determining whether or not someone is depressed based on their social media textual activity. The problem of detecting depression from social media has been posed as a classification problem in the Natural Language Processing domain (NLP). In this paper, we look at NLP techniques that can successfully extract information from textual material in order to improve depression detection. To create document representations, these NLP techniques extract different features.

2.Description Of The Problem Being Solved (Problem Statement)

One of the most dangerous and widely diagnosed mental diseases is depression. It has an impact on not just the victims, but also their families, friends, and society as a whole. There have been some new advances that try to predict the severity of depression in an individual by analyzing certain parameters collected from their twitter Data, thanks to significant advancements in Artificial Intelligence and Deep Learning. This study analyzes the different frameworks (CNNs, LSTM, etc.) and algorithms utilized in some of the most current and notable researches in the subject of depression analysis.Aside from the comparative analysis, this research also presents a few of ways that look into certain non-traditional factors that could be useful in the task of detecting and predicting sadness from Twitter data.

3. Review of Existing Systems

Several studies on text-based depression detection have been conducted. In this part, we'll look at some of the previous years' work.

Farig Sadeque, et al. had tried to detect early depression from a user's post on Reddit [6]. The authors used depressive words and concept unique identifiers from the Unified Medical Language System as features. They used both deep learning and traditional machine learning for the prediction task. He also used the ERDE metric to tackle a variety of difficulties [3]. The authors devised a Flatency metric to address the issues. Then they applied this metric to a number of previously created model eRisk 2017 depression tasks and found that it performed better. They conducted studies using data from the pilot project Early Detection of Depression in CLEF eRisk 2017.
Words, Depressive Words, Depressive Embed, and the Unified Medical Language System were among the four feature sets used.
To classify data, they employed both Deep Learning and standard machine learning approaches.

Tsugawa et al. [15] considered user activities in social media such as frequencies of words related to melancholy in a tweet, topics tweeted on, posting regularity etc.to check for the manifestation of depression. Nonetheless, by employing techniques like principle component analysis the feature set used could be improved. The methods such as deep learning and ensemble methods are expected to offer better results than SVM.

Quan Hu et al. [10] used classification and regression models to look for signs of depression in behavioral and dialectal twitter data. Their results can be greatly improved by including a wider range of people and lengthening the observation duration for better analysis.

4.Objective of the project

To automatically detect depression symptoms in text for decision support, correctly classify depressed users but also to reduce the amount of time to predict the state of the users
The project's main goal is to get the highest possible prediction accuracy for the input. This ensures that the final product is reputable and trustworthy. The accuracy rate is the key goal while using this dataset. As mentioned later in this work, various strategies have been tried to improve accuracy.

5. Output of the project or expected results of the project

Two algorithms will be used for deep learning. The advantage of using these two models is that they provide comparative analysis. These comparisons enable us to determine which model provides the best level of accuracy. Finally, we can determine which system performs better at detecting intrusions.

6. Feasibility study indicating at least two possible solutions

As potential solutions in this study, we will consider Deep learning and Natural learning possibilities. These two strategies will assist us in obtaining the output result of our study as well as identifying intrusions

Deep learning: We're going to start with deep learning because we already have the dataset and can preprocess, train, and test it with deep learning models to reach the desired outcome.

Natural Language Processing: Many downstream applications, such as text analytics, benefit from NLP because it helps clarify difficulty in language and adds helpful analytical structure to the data.

7. Working steps (Work plan)

1. Background Study: We read some research papers regarding our project topic We also studied deep learning and some suitable programming languages for it. Moreover, we gathered some knowledge about different types of deep learning algorithms. Through a couple of weeks thus we studied these things.

2.System Design Requirement : For a number of days, we talked about the needs for our project's system design. We've decided to use Python as our programming language in the next few days. As a result, our system will necessitate the Python Programming Language as well as several of its libraries. A dataset will also be required for our system design.

3. System Design Analysis:
For data analysis, the ideal programming language is Python. With Python's access to large libraries, it makes it the suitable language to deal with deep learning methods. For moderately smaller dataset, personal GPU usage will be done for the data pre-processing.

4. Dataset Collection: We have collected a suitable dataset (available online) related to our work. This will be a dataset of tweets which we can use to train our model.

5. Training: For this phase we will need to apply the deep learning models.
During this stage, we will be required to complete a number of tasks
i. Dataset process : The dataset we collect will need to be processed to be trained. This is a a very crucial part of training.

6. Testing: After collecting and training the dataset, we will start testing the dataset by applying our model. We will work on various testing methods so that we can get the result we desire. Thus, a couple of weeks will have passed.

7. Predict: In this phase, we will finally predict the accuracy of our trained and tested models.This is the most important part since our whole work was to predict the accuracy of our trained models. We will spend some days on it.

8. Implementation: This will be the implementation part. We will be working on this step for some weeks.

8. The major milestones of our project

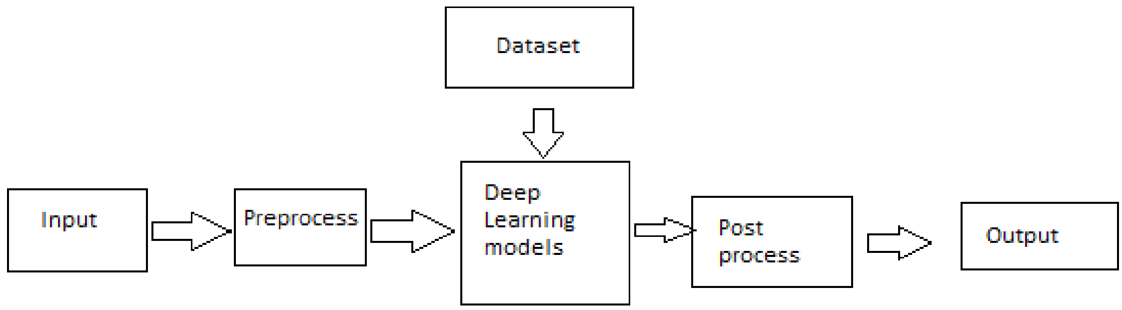
- Background Study
- System Design Requirements
- System Design Analysis
- Dataset Collection
- Training
- Testing

- Predict
- Implementation

9.Research Methodology:

9.1 Dataset: Depressive and non-depressive tweets between December 2019 and December 2020 originated largely from India and parts of the Indian subcontinent. Sentiment scores are allotted using text blobs. Tweets are extracted with the top 250 most frequently used negative and positive lexicons in mind, as determined by SentiWord and other research publications.It has about 1.4 million tweets.

9.2 Block Diagrams:- The system was depicted as a block diagram. As input, we use the dataset. Pre-processing on the data is done with the system, which includes loading datasets of a particular size, splitting the dataset.



9.3 Data Preprocessing:-

The raw data is prepared in data preparation so that it can be used by a deep learning model without further processing. It is the first and most crucial step in the development of a deep learning model.

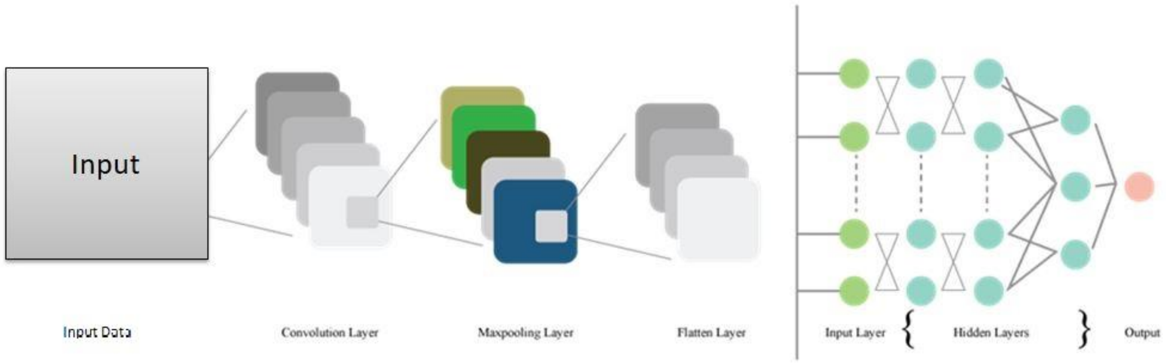
Data preparation tasks such as cleaning and preparing data for use in a deep learning model can increase the accuracy and efficiency of the model. We need to import several predefined Python libraries in order to execute data preparation with Python. These libraries are used for a variety of tasks.

9.4 Algorithms

Depression Analysis will be done using the following deep learning algorithm.

1) CNN
2) LSTM

CNN:- A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning system that can take an input image, assign relevance (learnable weights and biases) to various aspects/objects in the image, and distinguish between them.



Convolutional Layer:- This is CNN's basic layer. A filter is applied to the input data. The function map is then used to derive values. It is in place to determine the qualities. The filter is made up of a two-dimensional array of weights that are multiplied by an array of data input. Between the filter's patch and the input's patch, a dot is produced. In most cases, the filter is smaller than the input. The input from multiple points is multiplied using the same filter.

Assume the NN input is $V \in RA\times B$, where A is the number of features indicating an input frequency band and B represents the total number of input frequency bands. In the case of filter bank features, B represents the size of the filter bank function vector. Assume v = [v1 v2... vB], where vb is the band b function vector. The activations of the convolution layer can be calculated as:

$$h_{j,k} = \theta\left(\sum_{b=1}^{s} w_{b,j}^{T} v_{b+k-1} + a_j\right) \quad (1)$$

where $h_{j,k}$ denotes the j^{th} feature map's convolution layer output of the k^{th} convolution layer band, s denotes filter scale, $w_{b,j}$ denotes a weight vector for the j^{th} filter's b^{th} band, a_j denotes the j^{th} feature map's bias, and (x) is the activation function [19].

Pooling Layer:- This layer is in charge of simplifying the sampling process by aggregating the existence of characteristics. The pooling layer is applied after the convolution layer has been applied. This layer primarily employs two approaches. The average pooling is one of them, while the maximum pooling is the other. These sum up the average presence of a function and the highest activated presence of a function, respectively.

The data is transformed into usable data via the pooling layer, which removes redundant characteristics. When you use Average Pooling, the layer takes the value of its current view and averages it. When Max Pooling is enabled, the layer selects the highest value from the filter's current view at each iteration. The max-pooling strategy picks only the maximum value, resulting in fewer output neurons, by utilizing the matrix size defined in each feature map. The image shrinks drastically as a result, yet the scenario remains unchanged. A pooling layer reduces the amount of feature mappings and network parameters. A dropout layer is employed to prevent overfitting.

$$p_{j,m} = max_{k=1}^{r}\left(h_{j,(m-1)(n+k)}\right) \quad (2)$$

where $P_{j,m}$ is the performance of the pooling layer of the j^{th} function map and the m^{th} pooling layer band, n is the sub-sampling factor, and r is the pooling scale, which is the number of bands to be pooled together, and n is the sub-sampling factor, and r is the pooling scale, which is the number of bands to be pooled together, and n is the sub-sampling factor.

The Flatten layer:- The Flatten layer turns the matrix's data into a one-dimensional array that may then be used in the fully linked layer. To make a single one-dimensional feature that is long and thin. If desired, vectors can be flattened. It then connects the single vector to the fully linked layer, often known as the final classification model.

Fully Connected Layer:- CNNs rely heavily on fully connected layers, which have proven to be extremely useful in text recognition and classification in computer vision. The CNN method begins with convolution and pooling, which splits apart the data into properties and examines them separately.

Each input is connected to all neurons in a fully connected layer, and the inputs are flattened. As a completely connected layer, the ReLu activation function is frequently utilized. In the last layer of the fully linked layer, the Softmax activation function is utilized to forecast output results. At the end of the Convolutional neural network architecture, a fully linked layer is used.

LSTM:- LSTM networks are a sort of recurrent neural network that may learn order dependence in sequence prediction challenges.

This is a requirement in a variety of complicated issue domains, including machine translation, speech recognition, and others.

Tanh:- Tanh is a non-linear activation function. It regulates the values flowing through the network, maintaining the values between -1 and 1. To avoid information fading, a function is needed whose second derivative can survive for longer. There might be a case where some values become enormous, further causing values to be insignificant. You can see how the value 5 remains between the boundaries because of the function.

Sigmoid:- The sigmoid function is a type of non-linear activation function. The gate keeps it contained. Sigmoid, unlike tanh, keeps the values between 0 and 1. It aids the network's ability to update or forget data. If the multiplication yields a zero, the data is considered lost. Similarly, if the value is 1, the information remains.

There are three different gates in an LSTM cell: a forget gate, an input gate, and an output gate.

Natural language processing:- Natural language processing (NLP) is the ability of a computer software to interpret spoken and written human language, often known as natural language. It's a part of AI (artificial intelligence).

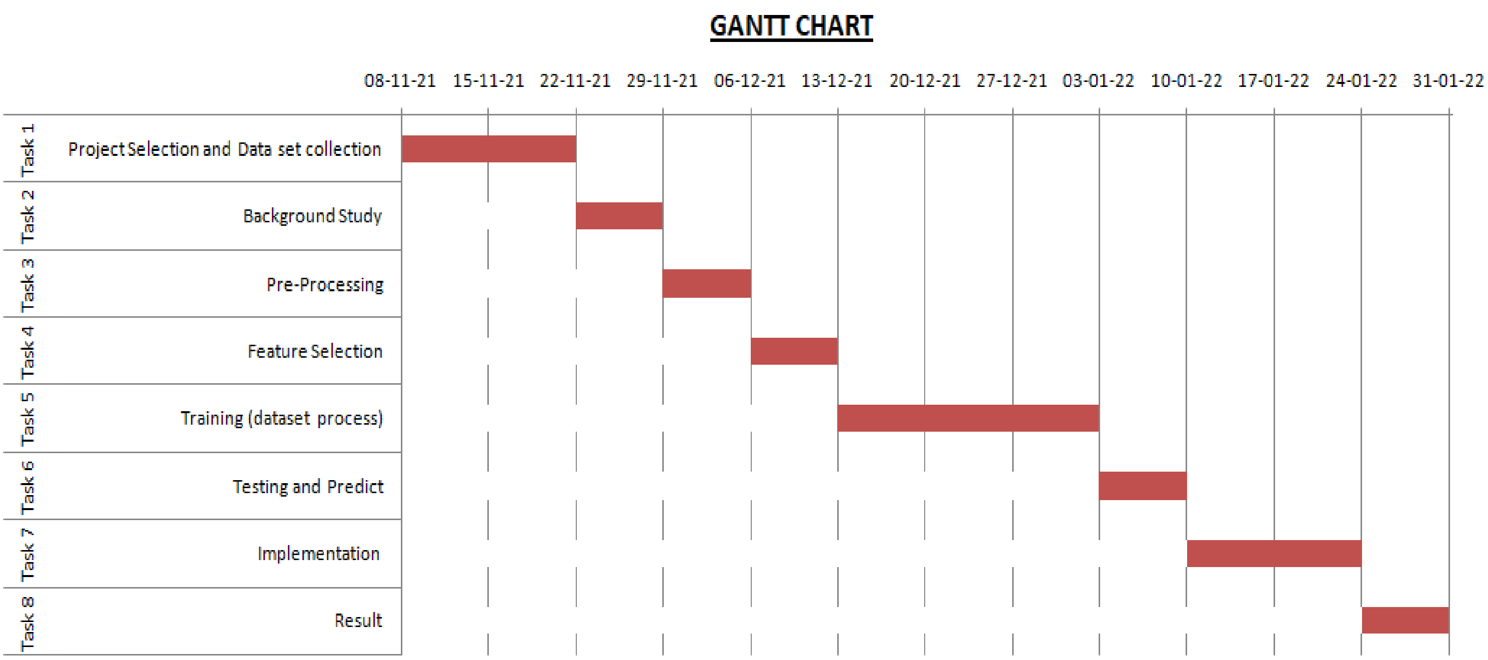
Data preprocessing involves preparing and "cleaning" text data for machines to be able to analyze it. preprocessing puts data in workable form and highlights features in the text that an algorithm can work with.

<u>Confusion Matrix:-</u> Confusion matrices are calculated using the predictions of a model on a data set. By looking at a confusion matrix, you can gain a better understanding of the strengths and weaknesses of your model, and you can better compare two alternative models to understand which one is better for your application. Traditionally, a confusion matrix is calculated using a model's predictions on a held-out test set.

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Figure : Confusion matrix block diagram



10. Project Timeline

11. Software tools

Basically two software applications will be used. The first is named Anaconda Navigator, and the second is called Jupyter Notebook. Python is going to be the programming language of choice. The model training and validation will be carried out using Anaconda Navigator and Jupyter Notebook.

Anaconda is a Python and R programming language distribution aimed at simplifying package management and deployment in scientific computing.Anaconda Navigator is a desktop graphical user interface that comes with Anaconda that allows you to run programs and manage CONDA packages, environments, and channels without having to use command line commands.The Jupyter Notebook is an open-source web tool that lets experts create and share documents with live code, equations, computational output, visualizations, and other multimedia elements, as well as explanatory text.

Our project's programming language will be Python. Data may be modified in Python in a variety of ways. Python may also be used to easily visualize data. These Python features will help us with our project. We chose Anaconda for our project because it is a Python programming language distribution.

12. Target Population (Users of your system).

Our target would be people of Bangladesh but priority will be given to younger generation cause they tend to fall in depression due to pressure from school

13. What makes the solution an 'innovation'

Yes Innovative, all the papers we have read, all the papers' accuracy is below 90%. Our dataset is an efficient dataset, so we will make it to 100% accuracy.

Our key goal was to not only correctly diagnose depressed users, but also shorten the time it took to predict their state.

14. Sustainability of the project.

Yes, this is a long-term project because it is a software project written in Python that is easy to understand and maintain, unlike hardware projects.

15.Project Scalability .

The project can be expanded to handle a bigger number of datasets. This will enhance the accuracy rate even further. It can also be taught to recognize various types of depression. If scaled up and other types of depression predictions are added, this project might become the one-stop shop for all potential Depression forecasts.

16. Income Generation.

The model can be utilized by hospitals and medical workers to improve their diagnosis of depression patients using the techniques used, such as trained models and algorithms. Human errors are less likely when the model does all of the forecasting. This might be an intriguing concept that encourages users to employ prediction algorithms in this way. It can be marketed or licensed to hospitals, or even offered to the general public, if the market is ready to accept it.

17.Funding

Currently, no funding is required to complete the analysis for this study.

18. Benefit .

The advantage is that such models can help diagnose depression much more quickly. In a couple of minutes, a prediction can be made. This reduces the time it takes to get a diagnosis. The likelihood of misdiagnosis due to human mistake is also reduced when the accuracy rate is considered. As a result, the user will profit in terms of both time and accuracy.

19. Risk Factor.

Medical forecasts are constantly at risk of being incorrect, which might lead to a misdiagnosis. However, by working with huge datasets and pre-trained models to improve the accuracy rate, this will be lowered.

20. Environmental Impact.

Our project has no negative environmental consequences. It will not hurt the environment in any way; It will be environmentally friendly.

21. Existing Research Publications

- Predicting depression using deep learning and ensemble algorithms on raw twitter data (Nisha P. Shetty, Balachandra Muniyal, Arshia Anand, Sushant Kumar, Sushant Prabhu)

  This research investigates and analyzes both machine learning (Random Forest) and deep learning LSTM and CNN algorithms used in this system. Prediction accuracies of 72% for Random Forest and of 93% for LSTM were obtained from this system.

- Deep Learning for Depression Detection of Twitter Users
  (Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, Diana Inkpen)

  This research uses 4 models, the first three models use CNN and the last one uses RNN. To optimize word-embedding for classification tasks. They performed a comparative evaluation on some of the widely used deep learning models for depression detection from tweets on the user level.

22. Conclusion

We shall have to determine whether the person is depressed or not in this report. Efficiency will be given top priority. It'll help users to get medical attention in a much shorter time.

23. Bibliography

[1] Fariq Sadeque and Dongfang Xu and Steven Bethard, "UArizona at the CLEF eRisk 2017 Pilot Task: Linear and Recurrent Models for Early Depression Detection",CEUR Workshop Proc. Author manuscript; PMC 2017. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5654552/

[2] Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki, “Recognizing Depression from Twitter Activity,” In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). ACM, New York, NY, USA, pp. 3187-3196, 2015. DOI: https://doi.org/10.1145/2702123.2702280.

[3] Q. Hu, A. Li, F. Heng, J. Li and T. Zhu, “Predicting Depression of Social Media User on Different Observation Windows,” 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Singapore, pp. 361-364, 2015. DOI: 10.1109/WI-IAT.2015.166.

[4] Shetty, N., Muniyal, B., Anand, A., Kumar, S. and Prabhu, S., 2020. Predicting depression using deep learning and ensemble algorithms on raw twitter data. [online] Available at: <http://doi.org/10.11591/ijece.v10i4.pp3751-3756>