

# **CLASSIFICATION OF DEPRESSION ON SOCIAL MEDIA USING TEXT MINING**

**A Project Work**

*Submitted in the partial fulfilment for the award of the degree of*

**BACHELOR OF ENGINEERING**

**IN**

**CSE(AI&ML)**

**Submitted by:**

**Pranay Reddy(18BCS6100)  
Karishma Kansal(18BCS6103)  
Chitrang Juneja(18BCS6115)  
Sejal Gulati(18BCS6116)  
Sachin Kaushik(18BCS6119)**

**Under the Supervision of:**

**Mr. Gurpreet singh**



**CHANDIGARH  
UNIVERSITY**  
*Discover. Learn. Empower.*

**CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,  
PUNJAB**

**April, 2020**

## **DECLARATION**

I, **Sachin Kaushik**, student of ‘**Bachelor of Engineering in Computer Science**’, session: **2018-2022**, Department of Computer Science and Engineering, Apex Institute of Technology, Chandigarh University, Punjab, hereby declare that the work presented in this Project Work entitled ‘**Classification of depression on social media using text mining**’ is the outcome of our own bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

**(Sachin Kaushik)**

**Candidate UID: 18BCS6119**

**Date: 24 April,2021**

**Place: Chandigarh University**

*Annexure-3 (A typical specimen of table of contents)*

## **Table of Contents**

Title Page	i
Declaration of the Student	ii
Abstract	iii
Acknowledgement	iv
List of Figures	v
List of Tables (optional)	vi
Timeline / Gantt Chart	vii
<b>1. INTRODUCTION*</b>	<b>1</b>
1.1 Problem Definition	1
1.2 Project Overview/Specifications* (page-1 and 3)	2
1.3 Hardware Specification	3
1.4 Software Specification	4
1.3.1	4
1.3.2	
...	
<b>2. LITERATURE SURVEY</b>	<b>5</b>
2.1 Existing System	5
2.2 Proposed System	6
2.3 Feasibility Study* (page-4)	7
<b>3. PROBLEM FORMULATION</b>	<b>9</b>
<b>4. OBJECTIVES</b>	<b>16</b>
<b>5. METHODOLOGY</b>	<b>18</b>
<b>6. CONCLUSIONS AND DISCUSSION</b>	<b>19</b>
<b>7. REFERENCES</b>	<b>22</b>

## **ACKNOWLEDGEMENT**

This project is made under the shadow of Mr. Gurpreet Singh Panesar. It is a precise project regarding the **Classification of depression on social media using text mining**. I would like to express my special gratitude to our mentor who gave me the golden opportunity to do this wonderful project on such a topic which is a beneficial one for our society- that we will be able to detect Depression in people, which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to them. Secondly I would also like to thank my team mates for their dedicated efforts and for working with me on this educational project. This project was made for educational purpose as well as for personal research work which was carried individually at personal level or whosoever it may concern.

## **ABSTRACT**

Mental illness has been prevalent in the world; depression is one of the most common psychological problems. With the use of the large amount of data tweets and Facebook post online we can use machine learning to data mine it and be able to produce a meaningful and useful outcome.

Social media generates countless data every day because of millions of active users share and communicate in entire community, it changes human interaction. For this project, we will be using Python and various modules and libraries.

Twitter rapidly has become one of the most popular social media since it launched, it advises 313 million active users who produce 6,000 tweets on Twitter every second as June, 2016<sup>1</sup>. In favour of gathering the depression related data, we keep monitoring each streaming tweet that includes the word “depression” in entire Twitter community. Totally, we roughly have gathered 54-million of tweets that discussed the depression relevant field.

The aim of the project is to predict early signs of depression through **Social Media** text mining. Below are the steps to run the python codes using the data sets uploaded in the repositories or you can download your own.

# 1 INTRODUCTION

## 1.1. Problem Definition

Mental illness has been prevalent in the world; depression is one of the most common Psychological problem. Untreated depression increases the chance of dangerous behaviors. The significant challenge of detecting depression is the recognition that Depressive symptoms may differ from patients' behavior and personality.

For clinic depression, doctors may evaluate the patient via the depression test taken by patients. Apparently, these clinical records are restricted due to many factors, such as age, sex; moreover, they are private and expensive.

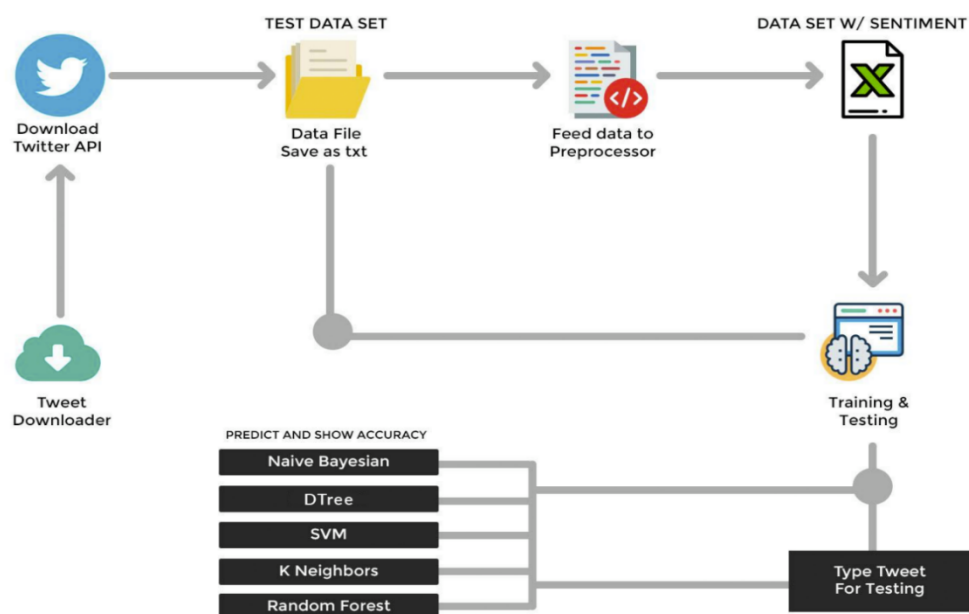
To overcome such limitations of clinical data, it would be beneficial to use text mining tools to extract and analyse depression symptoms from social media, such as Twitter. Social media generates countless data every day because of millions of active users share and communicate in entire community, it changes human interaction.

Other than the traditional data, such as literatures, social media data is richer and more accessible. However, investigating this new fast-growth of data requires advanced development tool to discover useful insights. These advanced technologies include Natural Language Processing (NLP) ,data mining, machine learning, social media analysis and so on.

In our research, the goal is to extract and summarize the uncommon but potentially helpful factors that depressive symptom performed from the social media data.

Finally, the extracted depression symptoms will be used as references when manually recognizing the clinical depression by humans.

Below is the flow chart that we will be following to build our project.



## **1.2. Project Specifications**

- This project mainly consists of social media post. Various social media posts are used for detecting whether there is depression or not.
- We have used Deep learning to take those as input and process them and give respective output as the task.
- We have used CSS, HTML, Pycharm Software, jQuery python to design UI. Python provides several different options for writing GUI based programs.

## **1.3. Hardware and Software Specification**

- Python 3.6.1 or Higher
- Twitter developer account
- A bunch of modules (Keras, TF, Numpy, sklearn, pandas and itertools)
- CSS, HTML for the GUI.
- Pycharm Software
- jQuery

# **2.LITERATURE REVIEW**

## **2.1 EXISTING SYSTEM:**

There is growing body of literature that analyses the properties of depression. Choudhury argue that depression constitutes a genuine test in individual and general wellbeing. Considerable number of individuals experiences the ill-effects of despondency and just a division gets sufficient treatment every year. They also investigated the possibility to utilize online networking to identify and analyze any sign of significant depression issue in people. Through their web-based social networking postings, they quantified behavioral credits identifying with social engagement, feeling, dialect and semantic styles, sense of the self-system, and notices of antidepressant medications.

Choudhury had considered online networking as a promising instrument for public health, concentrating on the utilization of Twitter presents on fabricating predictive models about the forthcoming impact of childbirth on the conduct and disposition of new mothers.

Utilizing Twitter posts, they measured postpartum changes in 376 mothers along measurements of social engagement, feeling, informal community, and phonetic style. O'Dea et al. examined that Twitter is progressively researched as methods for recognizing psychological well-being status, including depression and suicidality in the population.

Their investigation revealed that it is conceivable to recognize the level of worry among suicide-related tweets, utilizing both human coders and a programmed machine classifier

## **2.2. PROPOSED SYSTEM:**

The aim of the project is to predict early signs of depression through Social Media text mining. With the use of the large amount of data tweets and online Facebook posts, we can use machine

learning to mine the data and be able to produce a meaningful and useful outcome. For this project, we will be using Python and various modules and libraries.

What will be done-

- Natural Language Processing and Data Mining based Depression Classifier.

What will not be done-

- Use of any hardware/physical devices.

## **FUTURE GOAL:**

- Use Contextual Semantic segmentation
- Use Stopwords to increase accuracy of model
- Eliminating features with extremely low frequency
- Use Complex Features: n-grams and part of speech tags

## **2. PROBLEM FORMULATION**

Mental illness has been prevalent in the world; depression is one of the most common Psychological problem. Untreated depression increases the chance of dangerous behaviours. The significant challenge of detecting depression is the recognition that Depressive symptoms may differ from patients' behaviour and personality.

For clinic depression, doctors may evaluate the patient via the depression test taken by patients. Apparently, these clinical records are restricted due to many factors, such as age, sex; moreover, they are private and expensive. Its goal is to extract opinions, emotions or attitudes towards different objects of interest.

To overcome such limitations of clinical data, it was beneficial to use text mining tools to extract and analyze depression symptoms from social media, such as Twitter.

Social media generates countless data every day because of millions of active users share and communicate in entire community, it changes human interaction.

## **3. Research Objectives**

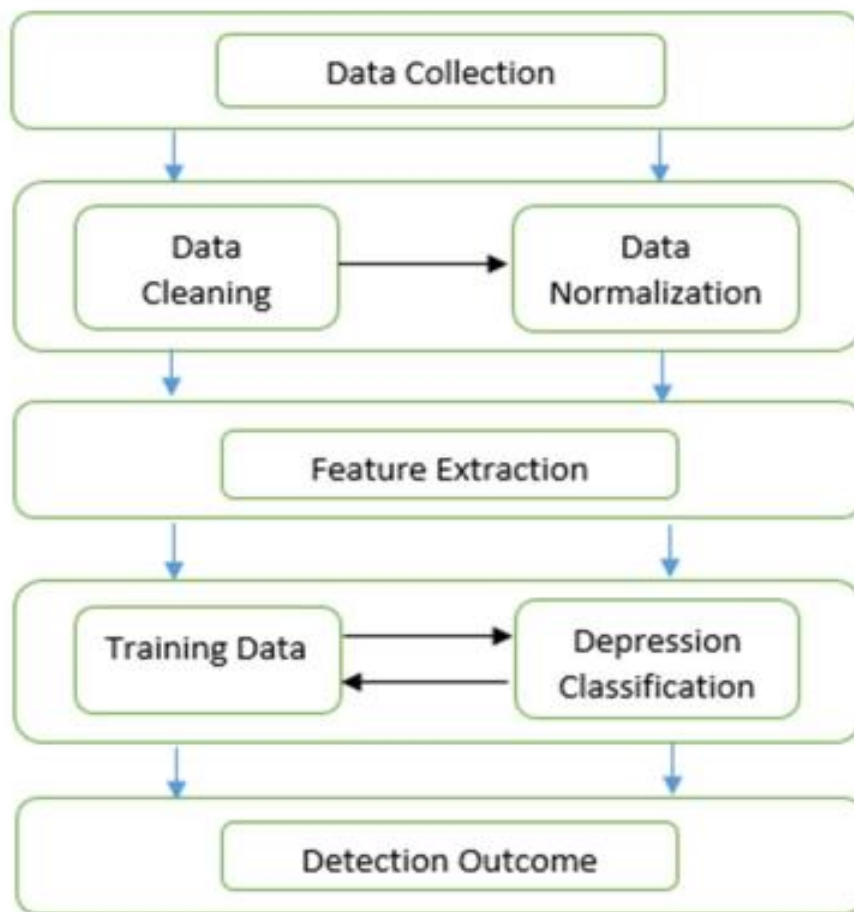
- We have studied different model combinations so that we could make use of ML and feature sets as well as NLP.
- We learnt a lot about neural networks and natural language processing
- We learnt how to make use of various machine learning libraries
- We learnt how to use data mining in order to extract the social media posts



## 4. Methodology

In this study, we first focused on different types of factors such as emotional process:- Temporal process, Linguistic style, etc. (emotional, temporal, linguistic style) features together for the detection and processing of depressive data received as Facebook posts.

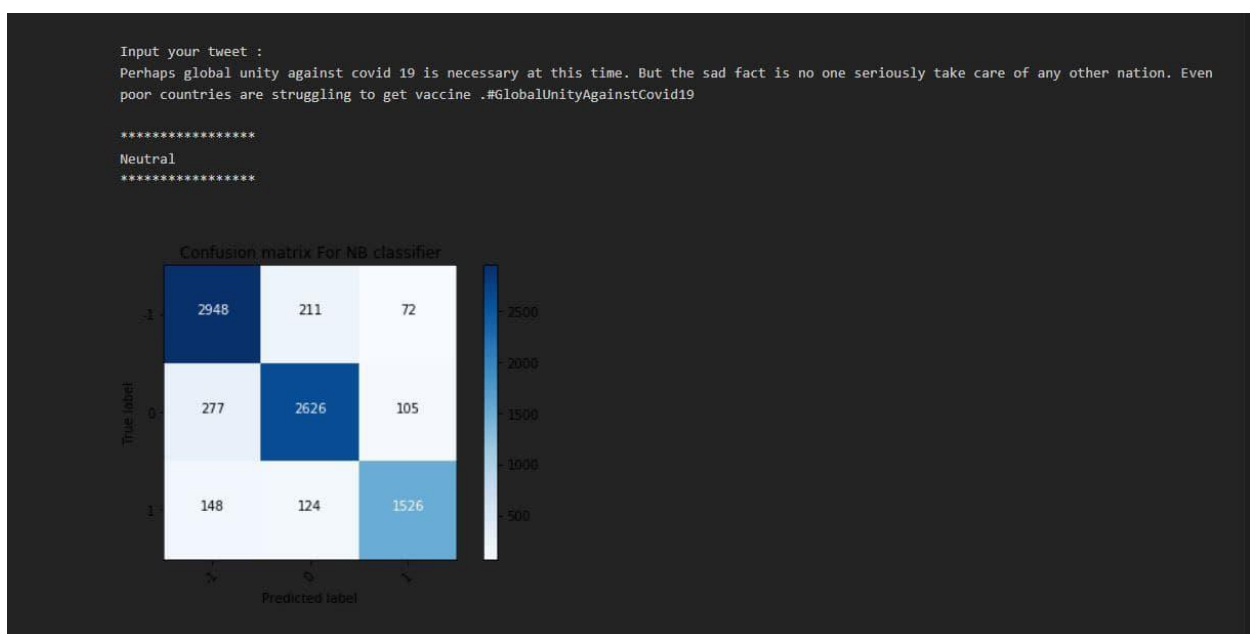
We then apply supervised machine learning approaches to study each factor types independently. Various classification techniques such as ‘decision tree’, ‘k-Nearest Neighbour’, ‘Support Vector Machine’, and ‘ensemble’ are used for this project.



## 5. RESULTS AND DISCUSSION

Clinic depression has been a serious mental illness since past decades which negatively affects human's health. it is difficult to confirm human's depression symptoms from their behaviours via restricted clinic records. Our proposed methods and experiments illustrate that social network and web blogs provide rich information for depression symptoms extraction from a distinctive perspective. Current advanced natural language processing approach, like Word2Vec, can be helpful for medical uses. In the future, we will collect other types of data, e.g. image and video from other social networks. Additionally, advanced entity selection technique would be used to select more accurate and meaningful depression symptoms.

We were successfully able to detect the right results in our project i.e.- Classifying if the person who posted that post, has the symptoms of depression or not.





Document x Depression\_SentimentAnalysis.ip: x +

localhost63342/PycharmProjects/flaskProject/templates/home.html?\_ijt=oqtaddfrt5o5e0orpqjbcknin2l

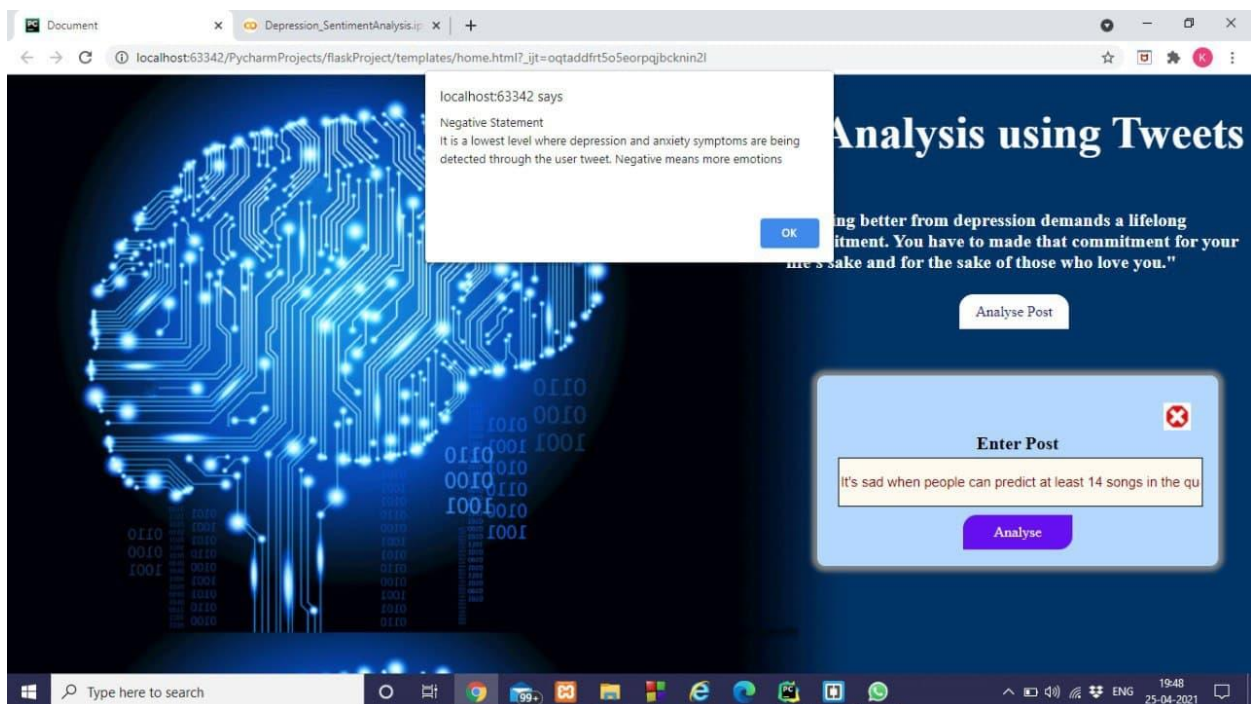
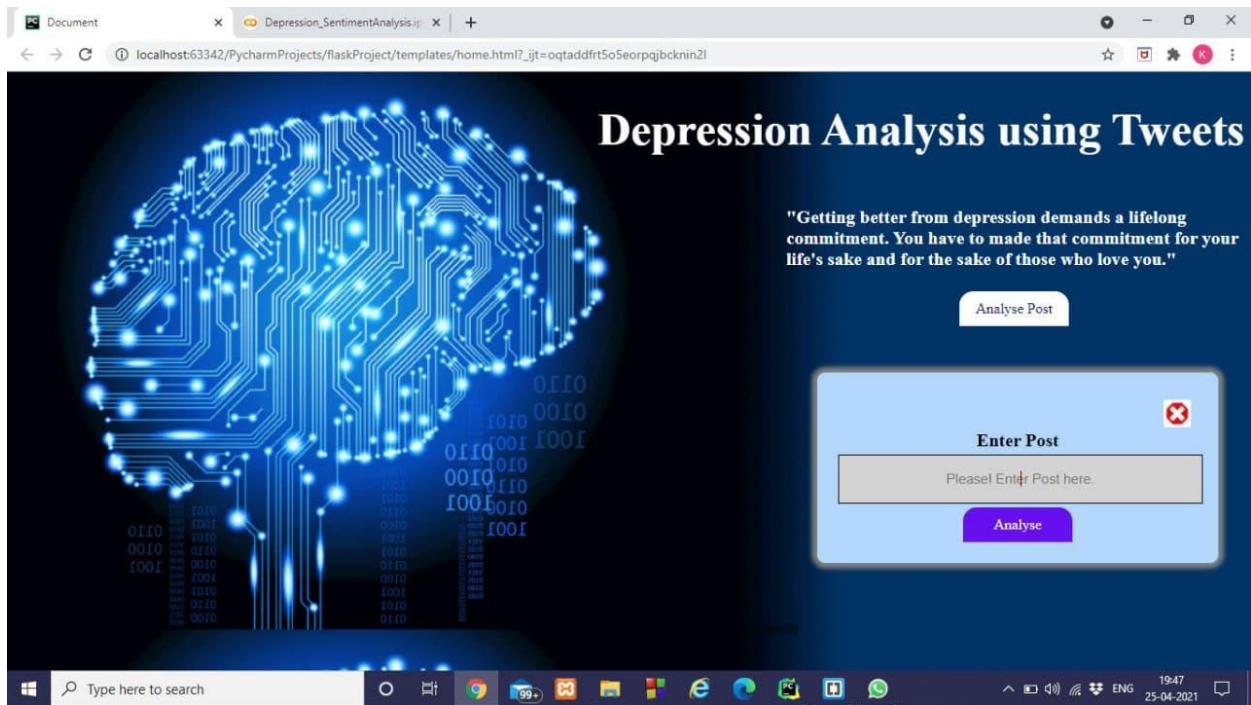
# Depression Analysis using Tweets

"Getting better from depression demands a lifelong commitment. You have to made that commitment for your life's sake and for the sake of those who love you."

Analyse Post

Type here to search

99% ENG 19:47 25-04-2021



## 6. REFERENCES

1. [https://www.researchgate.net/publication/318136574\\_Extracting\\_Depression\\_Symptoms\\_from\\_Social\\_Networks\\_and\\_Web\\_Blogs\\_via\\_Text\\_Mining](https://www.researchgate.net/publication/318136574_Extracting_Depression_Symptoms_from_Social_Networks_and_Web_Blogs_via_Text_Mining)
2. <https://vgpena.github.io/classifying-tweets-with-keras-and-tensorflow/>
3. <https://github.com/AshwanthRamji/Depression-Sentiment-Analysis-with-Twitter-Data>
4. [Griffin, J. M., Fuhrer, R., Stansfeld, S. A., Marmot, M.: The importance of low control at work and home on depression and anxiety: do these effects vary by gender and social class?. Social science and medicine, vol. 54\(5\), pp. 783--798. \(2002\)](#)
5. [Cai, Z., He, Z., Guan, X. Li, Y.: Collective data-sanitization for preventing sensitive information inference attacks in social networks. IEEE Transactions on Dependable and Secure Computing. \(2016\)](#)
6. [A Hybrid Document Feature Extraction Method Using Latent Dirichlet Allocation and Word2Vec | Request PDF \(researchgate.net\)](#)
7. <https://ieeexplore.ieee.org/document/9277251>
8. <https://github.com/laugustyniak/awesome-sentiment-analysis>
9. <https://github.com/arnab39/Sentiment-Analysis-of-Citations>
10. [A Hybrid Document Feature Extraction Method Using Latent Dirichlet Allocation and Word2Vec | Request PDF \(researchgate.net\)](#)
11. [DEEP LEARNING METHODS FOR DISEASE DETECTION AND CLASSIFICATION | Jyoti Islam | Research Project \(researchgate.net\)](#)