International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 3, May-June 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

DEPRESSION DETECTION USING MACHINE LEARNING TECHNIQUE ON X DATA

[1]Pamarthi Abhi Ram, [2]Mrs.B.Suneetha

[1](Student, Dept. of Master of Computer Applications, Amrita Sai Institute of Science and Technology, Paritala, Andhra Pradesh, 521180, India.
Email: abhi.pamarthi8592@gmail.com)

[2](Asst.Prof, Dept. of Computer Science & Engineering, Amrita Sai Institute of Science and Technology, Paritala, Andhra Pradesh, 521180, India.
Email: Sunitha.saahu@gmail.com)

----------**********----------

Abstract:

There are more cases of depression affecting people nowadays compared to before. There are people who realise their depression, but for some, it goes unknown. At the same time, individuals often share their feelings and thoughts on social media platforms which acts as a virtual diary to them. Studies done lately have attempted to recognise depression by applying machine learning to content produced by users. Analysing these posts is used to determine if a person is experiencing depression. The proposed research aims to spot depression in individuals by monitoring their X messages. Both Naïve Bayes and NBTree will be used to classify the prepared data. Scientists found that both algorithms show roughly the same high accuracy in detecting depression.

Keywords *— Depression detection, Social media analysis, Machine learning algorithms, Naïve Bayes, NBTree, X data*

----------**********----------

1. Introduction

Depression is a worldwide problem that is experienced by more than 264 million people worldwide. It can result from changes in the environment, alterations in brain chemicals or a person's genetic background. While depression can be helped by therapy or by using medicine, many people still go undiagnosed as not enough people are aware, so they experience issues such as seclusion, increased risk of suicide and becoming reliant on antidepressants.

Since social media is popular, people are expressing their emotions online and their posts can be used to find insights about mental health. X is popularly used for short message posts, giving a good chance to study people's moods through sentiment analysis. Using tone analysis, sentiment analysis helps researchers gather evidence about a person's mood and, if depressed, their need for immediate help.

In this paper, I discuss using sentiment analysis with machine learning algorithms Naïve Bayes and NB Tree to recognise depression through tweets. Their performance is checked by how accurately they can identify depressive text.

2. Requirements

2.1 Literature Survey

1. A Neural Network Approach to Early Risk Detection of Depression and Anorexia on Social Media Text.
   - Yu-Tseng Wang, Hen-Hsen Huang, Hsin-Hsi Chen are the authors.

- In this work, CNNs and TF-IDF are used to recognise depression and anorexia in messages shared on social media. Both depression detection ERDE5 and anorexia detection ERDE5 performed well during the CLEF eRisk 2018 competition, with results of 10.81% and 13.65%, respectively. What we found is that early signs of mental health problems can be recognised through social media.

2. Machine Learning-based Approach for Depression Detection in X Using Content and Activity Features.
   - Hatoon S. AlSagri and Mourad Ykhlef are the authors.
   - The study relies on machine learning and looks at both the posts and online connexions of individuals to identify depression on X. It indicates that using various features in a model improves the accuracy of detecting people who may struggle with depression.

3. Depression Detection and Prevention System by Analyzing Tweets.
   - Mrunal Gaikar, Jayesh Chavan, Kunal Indore and Rajashree Shedge are the authors.
   - This paper presents a system of detecting depression on X by examining depressive language in tweets. The system employs Support Vector Machine and Naïve Bayes classifiers to detect phrases associated with depression, providing a hybrid model for early detection and prevention.

2.2 Existing System

As of today, both sentiment analysis and mental health detection focus on:

- The process should at least cover recognising if sentiment is positive, negative or neutral.
- Little is known due to the unavailability of specialised datasets.
- Traditional NLP does not use deep learning.
- X does not have the ability to process information in real time.

Limitations:

- Lacking the ability to accurately pick up small depression signs.
- Small capacity for handling huge amounts of posts on social media.
- Integrating relevant features such as interactions and multimedia is flawed.

2.3 Proposed System

This plan is designed to address the issues outlined by:

- Building a machine learning solution for recognising depression signs in the data found on X.
- Using recently developed NLP methods (BERT, GPT and Transformers).
- Integrating features involving how people feel, compare various texts and what they do online.
- Getting information from X in real-time through APIs.
- Protecting data privacy and ensuring personal information is anonymous.

Advantages:

- Better detection of signs related to depression.
- You can use it for large datasets and monitor their progress live.
- It can also be used for other types of social media platforms.

2.4 Feasibility Study

Technical Feasibility:

- Takes advantage of TensorFlow, PyTorch and other natural language processing (NLP) libraries.
- With X's API, data is accessible at every moment.
- A technology stack offers the scalability to do complex computing and makes deployment of models possible.

Economic Feasibility:

- The CI/CD model is less expensive with AWS and Google Cloud platforms.
- Available sources for monetary support in studying mental health.

Operational Feasibility:

- Many scientists and programmers find that Python tools are very helpful.
- Psychology research may benefit from cooperation with mental health platforms.

2.5 Hardware And Software Requirements

Section	Details
Project Objective	The project aims to analyze the design of applications for enhanced user-friendliness, with an emphasis on smooth navigation and reducing user typing. The application needs to be compatible with most browsers for accessibility.
Functional Requirements	Graphical User Interface (GUI) with user interaction.
Software Requirements	1. **Python** 2. **Django**
Operating Systems Supported	1. **Windows 10** (64-bit)
Technologies and Languages Used	1. **Python**
Debugger and Emulator	Any Browser (Primarily **Chrome**)
Hardware Requirements	1. **Processor**: Intel i9 2. **RAM**: 32 GB 3. **Hard Disk Space**: Minimum 1 TB

TABLE1: HARDWARE & SOFTWARE-REQ

3. System Analysis

3.1 The document lists the required software for the system.

Its objective is to spot symptoms of depression through the use of machine learning and X observations. Scientists can depend on the system to cheque for depression with efficient and accurate text analysis.

Objective:

Create an instant and flexible system that can recognise depression based on X messages through NLP and machine learning.

Scope:

- Using live tweets to identify those dealing with depression.
- Take advantage of NLP tools for extracting features from the data.
- Carry out machine learning using classification models.

3.1.1 Functional Requirements

- **Data Collection:**
  - ⇨ Use the X API to access live tweets.
  - ⇨ You can filter tweets by looking for words related to depression such as "sad" or "hopeless".
- **Data Preprocessing:**
  - ⇨ Carry out tokenization, clean the text of common words and process the words into their root forms.
  - ⇨ Include the use of emojis and hashtags for analysing sentiments.
- **Model Training:**
  - ⇨ You should use labelled data to train models such as Random Forest, SVM and BERT.
- **Classification:**
  - ⇨ Label the tweets according to whether they reflect depression or not.
- **Visualization:**
  - ⇨ Use charts and graphs to help analyse the data.

3.1.2 Non-Functional Requirements

- **Performance:**
  - ⇨ Analyse at least 10,000 tweets every hour.
- **Scalability:**
  - ⇨ Know that the system has the ability to expand with more data.
- **Security:**
  - ⇨ Anonymize your users' data and assure that you follow privacy rules (for example, GDPR).
- **Usability:**
  - ⇨ Make the website easy for users to navigate.

3.2 MODULES

1. **User:**
   - Before communicating, users are required to register using both their email and mobile.
   - Users can upload datasets after the admin has given them access.
   - Users are able to cheque the outcome of classification and prediction.
2. **Admin:**
   - If someone is not yet a user, Admin will activate their account and explore all system data and outcomes.
3. **Data Preprocessing:**
   - You have to remove unnecessary information and HTML tags from the data.
   - The features are selected and fraud is detected using the Random Forest classifier.
4. **Machine Learning:**
   - KNN, Decision Tree, Random Forest, among others, are some of the classifiers used to perform classification.
   - The measurement results are given for accuracy, average and weighted averages.

4. System Design

4.1 SYSTEM ARCHITECTURE

The system architecture includes a description of the major components and modules in the project.

Overview:

- Input: X API was used to fetch data that had been filtered for content about depression.
- Processing:
  - Preparing and selecting features from the original data.
  - NLP is used in machine learning to classify information.
  - The tool produces an indication of depression in every tweet it analyses.

Architecture Components:

- In the Data Collection Layer, data is gathered through the X API and written in JSON/CSV files.
- Preprocessing Layer: Tokenizes the text, removes unnecessary stopwords and takes care of emojis, hashtags and special characters.
- Here, the text is changed into numbers using TF-IDF and other techniques for text processing.
- This level includes training and testing models, for example, Logistic Regression, Random Forest and BERT.
- Presentation Layer: Allows users to see data in a dashboard through their web browsers and gain real-time updates.

Describe how the components are connected through a system architecture diagram.

4.2 UML/ER DIAGRAMS

- **UML:** Structures, functions and interactions of the system can be diagrammed with UML. They assist in recording and interpreting the way the system is built.
- **Use Case Diagram:** Shows the connexions between users (actors) and the system, emphasising what the system is required to do. Employed to design and define the things the system will handle.

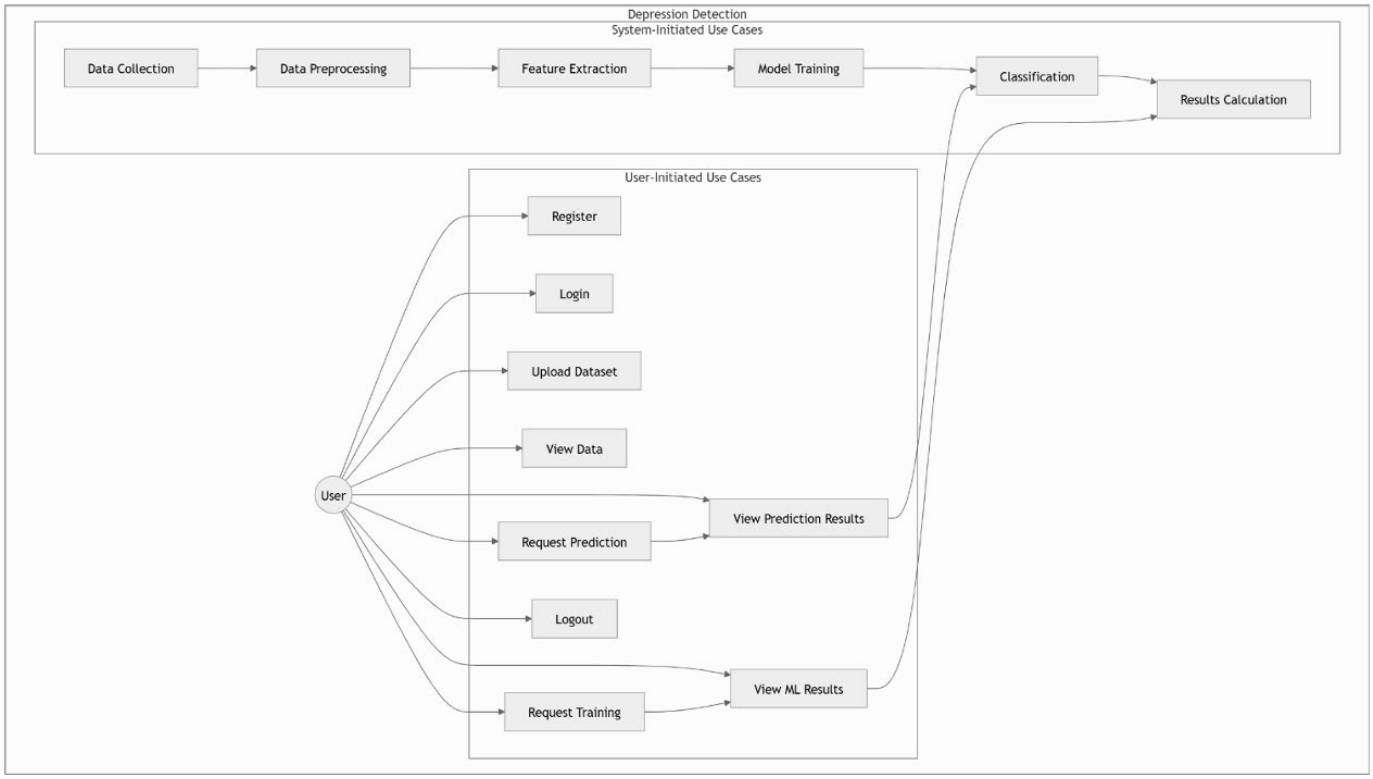

Fig1: UML Diagram

5. System Implementation

5.1 Python Technology

Python is a language used in many fields for being both easy to understand and simple to write. Programmers can use Java to build programmes using procedural, object-oriented or functional paradigms. Python is flexible for different uses because its library is extensive and it works on multiple platforms.

Key Features:

- Beginners find Python easy to understand, since its code is simple.
- **Capable:** Supports programming methods of both data, objects and functions.
- Consequently, Python code is executed piece by piece, making it simpler to notice problems.
- **Large Number of Modules:** Gives you tools for web development, analysing data and other purposes.
- The code can operate on different operating systems without any changes needed.
- Lots of people in the development community support Python.

Cybersecurity is most commonly utilised in the following ways:

- For web development, people use frameworks such as Django and Flask.
- Data Science & Machine Learning: You can use libraries such as Pandas, TensorFlow and Scikit-learn.
- Web scraping and task automation are best done with Automation & Scripting.

- With Pygame, you can create straightforward games.
- This field is used in simulations and computational biology.

The use of data analysis and machine learning involves processing information with automated systems.
Many professionals use Python in statistics, handling data and machine learning. Data visualisations, algorithm building and data analysis can be done through libraries such as Pandas, TensorFlow and Keras.

5.2. Python Platforms

You can use Python on Windows, macOS and Linux computers. The programme uses different libraries based on the operating system it is running on.

There are some key ideas behind each platform you need to consider.

- Python is supported by many systems including Windows, Linux and macOS.
- **Platform Module:** The platform module inside Python can provide information about the system it is running on.
- **Runs on Different OS:** Python allows a programme to function on a variety of operating systems.
- With Python, you can utilise virtual environments to maintain and handle software dependencies.

5.3 The Stages of a Python Programme

A Python programme follows the cycle of:

- **Developing the Programme:** Code the software using an IDE and name the file using .py as the extension.
- Making Bytecode: The interpreter transforms the original Python code to .pyc files.
- In the interpreter execution mode, the bytecode is executed on the computer.
- Python allocates memory as needed by using garbage collection.
- When you execute a programme, it completes actions such as handling data and calling functions.
- Errors are managed using the try-except blocks in my code.
- When the programme has finished operating as expected or due to an error, it comes to an end.
- The Python garbage collector removes unused memory in the programme.
- **Optimization:** After running the code, developers look for ways to improve its speed.

6. Output Screens

Fig2: Homepage

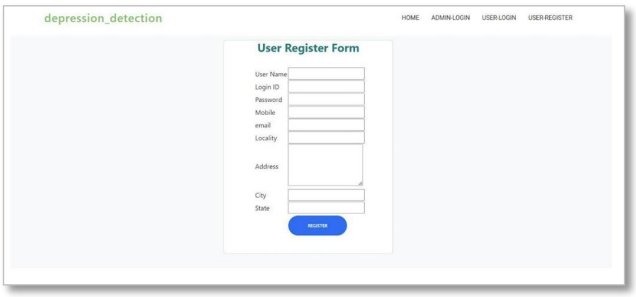

Fig3: User-Reg-Form

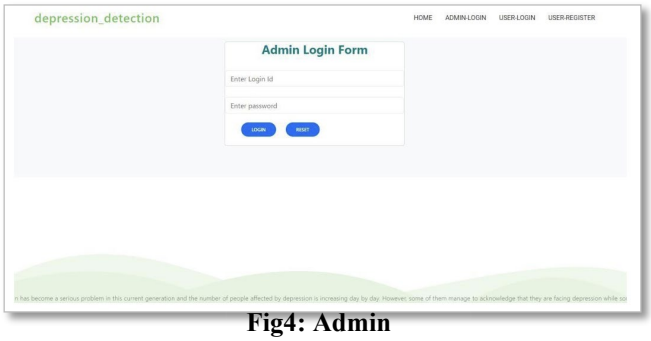

Fig4: Admin

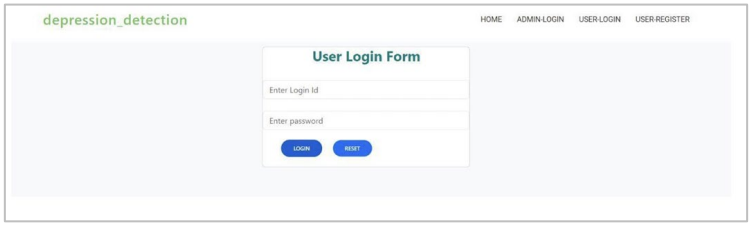

Fig5: Login-Form

7. Conclusion

All things considered, depression is now seen as a major challenge for mental health globally. With the rise of X and other social media, a tremendous amount of data has become available for researchers to study. Since X only allows 140 characters for tweets, users find it easier and faster to express their opinions, making it a valuable tool for analyzing feelings.

Applying sentiment analysis to every tweet allows us to identify whether the sentiment is positive, negative or neutral. After labeling, these tweets are used to teach machines how to classify tweets as either depressive or not. For this project, two sets of tweets were analyzed with two machine learning algorithms called Naive Bayes and NBTree. It was found that both algorithms completed the task at 97.31% on 3000 tweets and 92.34% on 1000 tweets.

Naive Bayes and NBTree achieve similar results in determining if a tweet is depressive or non-depressive. Also, this research only looked at written information. A deeper analysis of each user's tweets at different times could be included in future studies to further understand their depression.

Acknowledgement

I am thankful to the Management of Amrita Sai Institute of Science and Technology for giving me an opportunity to work with his project.
I would like to the thank **Dr. M. Sasidhar**, Principal, Amrita Sai institute of science and technology, for his constant encouragement and support during the progress of this work.

I am deeply grateful to **Dr. P. Chiranjeevi**, Professor and Head of the Department, for his valuable guidance and consistent support during the course of the project.
A special note of thanks to my internal guide, **Mrs.B.Suneetha (M.Tech)**, for her exceptional guidance, constant motivation, and continuous encouragement, which played a crucial role in the successful completion of this project.

PAMARTHI ABHI RAM

References

[1] World Health Organization, "Depression," [Online]. Available: https://www.who.int/health-topics/depression#tab=tab_1. [Accessed: Sep. 09, 2020].

[2] L. Goldman, "Depression: What it is, symptoms, causes, treatment, types, and more," 2019. [Online]. Available: https://www.medicalnewstoday.com/articles/8933. [Accessed: Oct. 10, 2020].

[3] S. Kemp, "Digital 2020: 3.8 billion people use social media - We Are Social UK - Global Socially-Led Creative Agency," Jan. 30, 2020. [Online]. Available: https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media. [Accessed: Oct. 10, 2020].

[4] Y.-T. Wang, H.-H. Huang, and H.-H. Chen, "A Neural Network Approach to Early Risk Detection of Depression and Anorexia on Social Media Text."

[5] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," Health Inf. Sci. Syst., vol. 6, no. 1, pp. 1–12, 2018, doi: 10.1007/s13755-018-0046-0.

[6] A. Sistilli, "Twitter Data Mining: Analysing Big Data Using Python | Toptal," [Online]. Available: https://www.toptal.com/python/twitter-data-mining-using-python. [Accessed: Feb. 04, 2021].

[7] Fontanella Clint, "How to Get, Use, & Benefit From Twitter's API," [Online]. Available: https://blog.hubspot.com/website/how-to-use-twitter-api. [Accessed: Mar. 05, 2021].

[8] World Health Organization, "Depression," [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression. [Accessed: Oct. 10, 2020].

[9] N. Schimelpfening, "7 Most Common Types of Depression," Aug. 26, 2020. [Online]. Available: https://www.verywellmind.com/common-types-of-depression-1067313. [Accessed: Sep. 09, 2020].

[10] M. R. Islam, A. R. M. Kamal, N. Sultana, R. Islam, M. A. Moni, and A. Ulhaq, "Detecting Depression Using K-Nearest Neighbours (KNN) Classification Technique," in Proc. Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2, 2018, pp. 4–7, doi: 10.1109/IC4ME2.2018.8465641.

