Title:-

Depression Classification and Prediction System from User Tweets Using ML Techniques

2nd research paper

BASE PAPER:

**Prediction mechanism of depression tendency among college students under computer intelligent systems**

**AUTHOR NAME : Zhang He**

**Publishing year :December 13, 2023**

**Research Overview**

This research focuses on predicting the tendency of depression using machine learning. **The main goal is to create a smart system that can detect early signs of depression based on people’s habits, emotions, and online behavior.** By using data from surveys and social media, the system can learn patterns that often appear in people experiencing depression. The research tests different machine learning models to find which one gives the best and most accurate results. This helps in finding new ways to support mental health using technology.

**Dataset**

**The data was collected from online surveys. People shared their feelings, lifestyle, habits, and mental health information.** Some data also came from social media behavior. This data helped the model learn what signs may show depression.

**Objective**

**The main objective of this study is to build a machine learning model that can predict if a person might have depression**. By analyzing data like mood, lifestyle habits, and social media use, the goal is to catch signs of depression early. **This can help doctors or support teams give help before the situation gets worse**

**Methodology**

First, **data was collected from online surveys and social media platforms.** This included answers about mood, stress, sleep, and other habits. Then, the data was cleaned and prepared so the machine could understand it.

**Next important features were selected such as how much someone sleeps, how stressed they feel, and how often they post online.** These features helped train the machine learning models.

**Several models like Decision Tree, SVM, Random Forest, Neural Network, and Logistic Regression were trained using this data.** Each model was tested to see how well it could predict if a person might have depression. **The best performing models were Random Forest and Neural Network.**

**Random Forest**

Random Forest is a type of machine learning model that uses many decision trees to make a prediction. Each tree gives its own answer, and the final result is based on the majority of these answers. It works well because it avoids overfitting and handles missing or noisy data better. In this study, it helped find patterns in depression symptoms by combining different answers in smart ways.

**Neural Network**

A Neural Network is a model designed to work like a human brain. It learns through layers of 'neurons' that pass information and adjust themselves to make better predictions. This model is very good at understanding complex and hidden patterns. In this study, it gave the most accurate results by deeply learning from many different features like mood, stress, and online behavior.

**Parameters Used** The model used many features to predict depression, like:

Age and gender, Sleep hours, Stress level, Exercise habits, Mood and emotions, Social media post types and frequency

**Execution Approach**

The execution of this project followed a structured pipeline:

* **Data Collection:** Data was gathered from public surveys and online platforms where people shared information about their mental health.
* **Data Preprocessing:** Unnecessary or missing values were removed. The data was formatted into a clean and structured form.
* **Feature Selection:** Important features like mood, stress, sleep, and social media activity were selected to train the model.
* **Model Building:** Multiple machine learning models were trained using the selected data. Models included Decision Tree, SVM, Random Forest, Neural Network, and Logistic Regression.
* **Model Evaluation:** All models were tested to check their accuracy. The two models that gave the best results were Random Forest and Neural Network.
* **Result Analysis:** The performance of each model was compared and the findings were recorded using a result table.

**Results**

**The models tested in this research showed different levels of accuracy. Among all the models, the Neural Network and Random Forest gave the best outcomes. These two models could correctly identify signs of depression in more than 90% of the cases**. Other models like SVM and Decision Tree also gave good results but were slightly less accurate. This shows that deep learning and ensemble methods are more reliable in detecting mental health patterns from user data.

|  |  |
| --- | --- |
| **MODEL** | **ACCURACY** |
| Decision Tree | 85% |
| SVM | 87% |
| Random Forest | 92% |
| Neural Network | 93% |
| Logistic Regression | 81% |

**Problems and Improvements**

One problem is that the data is self-reported. This means people may not always give honest or accurate answers. This can affect the model’s correctness.

Another issue is privacy. The data includes personal feelings and health information, so it needs to be protected better.

The model should be tested with larger and more real-time data. This will make the system more accurate in real-world use.

Also, the model should support more languages and regions. This will help more people from different places and cultures.

This system is a great start, but it can be improved with better data, more features, and privacy-safe methods.

**Conclusion**

This research successfully shows how machine learning can help in predicting depression. By using different types of data and testing many models, **the study found that Neural Networks and Random Forest are the most accurate.** These models can find early signs of depression based on simple lifestyle and emotional features. Though there are challenges like data accuracy and privacy, the project gives a strong base for future systems that can support mental health. With improvements, this system could be used in real-life to help people get early support and care.

**Research Gap**

Earlier studies used only one type of data like surveys. This project is different because it mixes many types of data—survey answers, social media activity, and health info. This gives better and more accurate results.