Detection Of Depression from Textual Data in Social Media Using Machine Learning

Abstract

Depression is one of the leading causes of suicide worldwide. However, a large percentage of cases of depression go undiagnosed and, thus, untreated. Previous studies have found that messages posted by individuals with major depressive disorder on social media platforms can be analysed to predict if they are suffering, or likely to suffer, from depression. This study aims to detect depression from textual data posted on social media using machine learning classifiers and advanced feature extraction methods.

Referred to 3 studies: The 1st paper shows that integrating LIWC, LDA, and bigrams with a multilayer perceptron enhances Reddit depression detection to 91% accuracy. The 2nd paper demonstrates that bag-of-words and n-grams with LR and RF can detect depression in social media posts with over 90% accuracy without keywords. The 3rd paper finds Decision Trees most effective for detecting depression in Facebook comments with 60-80% accuracy, while also tracking depression patterns over time.

The study uses N-gram, and TF-IDF (Term Frequency-Inverse Document Frequency) vectorization technique to extract features, focusing on emotional process, temporal process, linguistic style. The system leverages Multi-Layer Perceptron (MLP) and Random Forest (RF) for ensemble methods, along with Decision Trees (DT) to achieve high accuracy. Additionally, it includes a feature suggesting remedies for users identified as having depression. The system will analyse text from social media posts or comments to detect depressive content and provide supportive suggestions via a web interface developed using Flask.

To train and test the model, we utilize the below dataset: posts from datasets focusing on depression analysis by B.B.eye. The dataset is obtained from Kaggle, the dataset contains 10314 rows and 3 columns.

References:

1. Tadesse MM, Lin H, Xu B, Yang L. Detection of depression-related posts in reddit social media forum. Ieee Access. 2019 Apr 4;7:44883-93.

2. Chiong R, Budhi GS, Dhakal S, Chiong F. A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. Computers in Biology and Medicine. 2021 Aug 1;135:104499.

3. Islam MR, Kabir MA, Ahmed A, Kamal AR, Wang H, Ulhaq A. Depression detection from social network data using machine learning techniques. Health information science and systems. 2018 Dec;6:1-2.

Malavika.P

MAC23MCA-2037

Signature of the Guide

Signature of the Project Coordinator