Harnessing Machine Learning for Mental Health Prediction from Social Media Activity

Authors : S. Mohanraj, R. Ramesh Krishna, M. Shan Adams, C. Nallusamy

Abstract : The study in question investigates applying machine learning techniques to text data from Naive Bayes classifiers in predicting depression symptomology. The dataset encompasses the text and its relevant labels, while other variables include LIWC variables such as sentence length, post ID, subreddit, and lexical attributes. After data preprocessing-cleaning and feature selection-the topic modeling process involves Latent Dirichlet Allocation (LDA). The main purpose is to detect latent themes in the textual data, as well as to assess the influence of individual words among different publications. Then, the model is trained on a labeled dataset and evaluated against an independent one. These findings show good accuracy and F1 score to validate the performance to successfully detect depression symptoms on textual data. This work precisely shows how machine learning algorithms can be used in mental health research, especially in early detection and treatment of those who may manifest depressive symptoms with their online behavior.

Keywords : prediction, mental health, machine learning, social media analysis

Conference Title : ICMLC 2025 : International Conference on Machine Learning and Cybernetics

Conference Location : Bengaluru, India

Conference Dates : January 30-31, 2025