



**INFORMATICS
INSTITUTE OF
TECHNOLOGY**

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Polarize-AI

Ensemble Machine Learning for Early Diagnosis of Bipolar Disorder through Social Media Analysis

A dissertation by

Nisandi Jayasuriya

Supervised by

Mr. Iresh Bandara

Submitted in partial fulfilment of the requirements for the BSc (Hons) Computer Science
degree at the University of Westminster.

May 2023

ABSTRACT

Bipolar disorder is a complex mental illness that poses significant challenges for accurate diagnosis, particularly in its early stages as there are presently no reliable and effective diagnostic techniques available. Limited attention has been placed on the potential advantages of ensemble machine learning techniques for this purpose, despite previous research exploring the use of conventional machine learning algorithms to detect signs of bipolar disorder in social media data.

This work suggests an ensemble machine learning technique to identify people with bipolar illness using Twitter data to address this knowledge gap in past literature. The author gathered a sizable dataset of tweets from people with and without bipolar disorder, and then they utilized a variety of data preparation approaches to identify linguistic and behavioral traits linked to the illness. To increase the precision and caliber of bipolar illness detection, the retrieved characteristics were utilized to create an ensemble machine learning model, which integrates multiple distinct machine learning methods. To attain the best performance, the model was adjusted utilizing hyperparameter tuning methods.

In comparison to other research that employed traditional machine learning algorithms, the suggested ensemble machine learning technique obtained an accuracy of 95.83% in detecting bipolar illness using Twitter user profiles. Transformer based models were also explored in the study, for detecting bipolar illness symptoms in tweets.

Keywords: Bipolar Disorder, Ensemble Machine Learning, Deep Learning, Transformers, Natural Language Processing, Twitter

Subject Descriptors:

- Computing methodologies → Machine learning → Machine learning algorithms → Ensemble methods
- Computing methodologies → Artificial intelligence → Natural language processing
- Human-centered computing → Collaborative and social computing → Collaborative and social computing design and evaluation methods → Social network analysis
- Applied computing → Life and medical sciences → Health care information systems