



**INFORMATICS  
INSTITUTE OF  
TECHNOLOGY**

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

**Identification of Phraseological traits in Online Textual  
Representations related to Psychological Disorders (IPT-psy)**

A dissertation by

**Mr. Sachith Sandeep Suraweera**

Supervised by

**Miss. Isuri Anuradha Nanomi Arachchige**

Submitted in partial fulfilment of the requirements for the BEng. (Hons) in Software  
Engineering degree at the University of Westminster.

**May 2022**

## Abstract

Social media has taken all over the world when it comes to connecting people and expressing their feelings with each other. These expressions and feelings can be anything that goes up to extreme happy moments to depressed related posts depending on the human nature. When it comes to language, people can use special features such as sarcasm, irony, and idioms to change the meaning from what they are saying. When it comes to psychologically disabilities related posts most of the time society tends to use these special linguistic features to hide the real meaning what they are truly trying to say. These linguistic features easily can misread the reader with different meaning and criticisms, these also can be used to ridicule, bitter and taunt others. It proves that these features can be easily used for negative impact to both reader and author.

Most of the existing text classification systems are built with the aim of finding one specific language feature in their machine learning models. This project's main aim is to develop the system by training the multiple classes of data and combining them to get the expected result. The system consists of combination of preprocessing, feature engineering and hyperparameter tuning which are recommended for the system. After combing all the core features of the system, model has been trained and evaluated to give the expected solution to the user.

As for the model has been trained with a large dataset and managed to get an accuracy of 99 percent which is a decent accuracy. Proposed approached has been managed to reach the expected outcome and this research can also later be developed to many areas as for the future work.

**Keywords:** Machine learning, Depression, Sarcasm, Irony, Idioms, Text classification