

**MACHINE LEARNING APPROACH TO PREDICT
MENTAL DISTRESS OF IT WORKFORCE IN REMOTE
WORKING ENVIRONMENTS**

SANDUNI NILUSHIKA GAMAGE

A dissertation submitted in partial fulfilment of the requirement for

Master of Science degree in Business Analytics

Department of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

Robert Gordon University Aberdeen, Scotland

2022

Abstract

When considering online workers, due to the emergence of the coronavirus pandemic prevailing in the world, employees have been restricted to work remotely for a prolonged period. Since this has been novel to society, the impact caused by this crisis on people is unknown in the short or long term. Especially it has been identified that mental distress among remote workers has increased over time due to the change in usual working patterns and minimization of work-life separation. Even though it prevails, employers don't provide any special attention to the mental well-being of their employees, and employees themselves are not aware of any distress they are going through. These affect how people feel, think, and act. Since these mental distresses can cause severe consequences if not diagnosed early and correctly, it is essential to implement a mechanism to detect early signs of mental distress and take necessary measures to address them. Since various factors can cause mental distress among remote workers, periodic screening for common mental distresses such as anxiety, depression, and stress is necessary for health and well-being. Nonetheless in a country like Srilanka where mental health care resources are minimal, it is difficult to conduct mental health screening frequently. Therefore, a standalone system to early detect any prevailing mental distress or possible mental distress without any medical assistance is much needed. As a result, this study was conducted to implement a screening system to predict the mental distress of an individual based on the associated external features using machine learning techniques. The system is designed so that any individual can assess their mental condition by themselves and proceed to any medical care if necessary.

As the causes of mental distress are multifactorial, it was designed on the hypothesis that socio-demographic, biological, economical, environmental, occupational, and psychological aspects directly affect the mental health of an individual. The study was conducted concerning the circumstances in a pandemic era considering COVID-19 as the case study. The study was done with remote IT workers in Sri Lanka who works as a part of a software development team. The data was collected through a questionnaire specially designed to capture any prevailing mental distress and associated external features. 481

professionals participated in the study and were selected based on selection criteria and 370 participants were selected for the final analysis. Of them, 358 were identified as distressed and 12 were not distressed. Then appropriate encoding techniques were utilized to encode categorical variables where the most important 25 features were detected among 60 features using feature selection. Finally, classification techniques such as Random Forest, SVM, XGBoost, CatBoost, decision tree, and Naïve Bayes were used for modeling by which the CatBoost algorithm in overall measures outperformed other algorithms with a predictive accuracy of 97.1%, precision of 97.4%, recall of 99.7%, f1 measure is 98.5% and ROC/AUC score of the model is 99%.

This prediction model is more efficient since it scored higher values for the f1 measure and ROC/AUC score. learning has been done on highly imbalanced data and higher interest is in deriving distressed individuals correctly. Therefore, with regard to the main performance evaluation methods of F1 score and ROC/AUC score CatBoost outperformed other models in both measures.

Keywords – Classification, external features, IT employees, Machine Learning, Mental distress