

Machine Learning Approach to Predict Mental Distress of IT Workforce in Remote Working Environments

Sanduni Nilushika Gamage*
School of computing
Robert Gordon University, Scotland
nilu.gamage95@gmail.com

P. P. G. Dinesh Asanka
Department of Industrial Management
University of Kelaniya, Sri Lanka
dasanka@kln.ac.lk

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. All the working arrangements are now based at home than before. Since this has been novel to society, the impact caused by this crisis on people is unknown in the short or long term. Since various factors can cause mental distress among online workers, periodic screening for mental distress such as anxiety, depression, and stress is necessary for health and well-being. The causes of mental distress are multifactorial. They include socio-demographic, biological, economic, environmental, occupational, and psychological aspects. This paper proposes a concept of a screening system to predict mental distress given the external features associated with individuals, using supervised machine learning approaches and identifying the employees prone to higher risk and referring them early to professional assistance. 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 work as a part of a software development team. 481 professionals participated in the study and were selected based on selection criteria and appropriate encoding techniques were utilized to encode categorical variables where 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%, and f1 measure is 98.5%.

Keywords – classification, external features, IT employees, Machine Learning, mental distress