Department of Computer Science

St. Francis Xavier University

Presents

A Novel Tweet Dataset and Comparative Analysis of BERT-Based Classification Models for Phobia Subtypes

by

Anik Das

St. Francis Xavier University

M.Sc. Thesis Proposal Presentation

Monday, May 29 at 11am in Annex 23A

Phobias are common anxiety disorders that affect a large portion of the population of the entire world. Timely detection of phobia subtypes can lead to accurate interventions and the betterment of mental health. Machine learning techniques can create a valuable impact in accurate phobia subtypes detection, however, a dataset is required for such tasks. In this study, I proposed a novel Tweet dataset with 102 phobia subtypes and detection models using machine learning algorithms. I discussed the data collection process, pre-processing and manual labeling steps, the data augmentation process, and a BERT-based embedding technique. For the phobia subtypes classification, I will compare the efficacy of different pre-trained variations of BERT-based models (e.g., ALBERT, RoBERTa, DistilBert). To the best of my knowledge, my study will be the first to collect a dataset and classify a vast number of phobias, which will be beneficial for mental healthcare facilities.