

GUEST SPEAKER



ML Tlachac is a Data Science PhD candidate at Worcester Polytechnic Institute (WPI). This is Tlachac's fourth year as a graduate student at WPI. Tlachac's graduate research and publications have been within health informatics. Research interests include machine learning, visualization, unsupervised learning, and natural language processing.

HEALTH INFORMATICS DATA SCIENCE

Depression is the leading cause of disability, often undiagnosed, and one of the most treatable mood disorders. As such, unobtrusively diagnosing depression is important. Many studies are starting to utilize machine learning for depression sensing from social media and Smartphone data to replace the survey instruments currently employed to screen for depression. In this study, we compare the ability of a privately versus a publicly available modality to screen for depression. Specifically, we leverage between two weeks and a year of text messages and tweets to predict scores from the Patient Health Questionnaire-9, a prevalent depression screening instrument. This is the first study to leverage the retrospectively-harvested crowd-sourced texts and tweets within the combined Moodable and EMU datasets. Our approach involves comprehensive feature engineering, feature selection, and machine learning. Our 245 features encompass word category frequencies, part of speech tag frequencies, sentiment, and volume. The best model is Logistic Regression built on the top ten features from two weeks of text data. This model achieves an average F1 score of 0.806, AUC of 0.832, and recall of 0.925. We discuss the implications of the selected features, temporal quantity of data, and modality.

THURSDAY, MAY 7TH

8:30 EST VIA ZOOM

FLORIDA STATE ONLINE COMPUTER CLUB

FOR MORE INFO, EMAIL KAREN WORKS KEWORKS@FSU.EDU

