Name: Sovannara Chou

Faculty advisor: Professor Wooyoung Kim

Project title: Early Prediction of Depression in Individuals with Hypertension

In this research, I have collaborated closely with Professor Wooyoung Kim and Dr. Chiyoung Lee to explore datasets centered on depression in individuals with hypertension. The goal of this project aims to develop and compare the effectiveness of different machine learning models with two data balancing techniques to analyze the data.

My tasks involve following the ML processes to collect, organize, and clean the data, selecting the appropriate model, conducting training and testing, and finally interpreting the results. The project utilizes various types of machine-learning models, ranging from interpretability models to complex deep-learning models, implemented in Python. The models I employ include Logistic Regression, Support Vector Machine, and Multi-layer Perceptron neural network. Data imbalance is managed through over-sampling and random down-sampling, allowing for a comparison of performance between these two techniques. Model performance is assessed using accuracy rate and F1-score metrics. Additionally, I implement a confusion matrix to demonstrate how well the model predicts data, displaying True Positive, True Negative, False Positive, and False Negative outcomes.

In the analysis of the three models, the support vector machine achieves the best performance, attaining an accuracy of 92–94% with a 90% F1-score, in contrast to Dr. Lee’s previous research using random under-sampling, which yielded a top result of 77%. Both logistic regression and multi-layer perceptron models yielded results around 75–80%, along with corresponding F1 scores. The most favorable outcome was achieved using oversampling data balancing. For random under-sampling, the highest accuracy across all models was 77%, consistent with Dr. Lee's analysis. Overall, my observations and analysis lead to the conclusion that the support vector machine model with the oversampling data balancing technique, achieves the highest performance accuracy among the three models.