

Healthcare Predictive Analytics

Naga Siva Poojan Kumar Mittapalli

nmtplli@memphis.edu - U00954242

Problem/Application

With the increasing prevalence of mental health disorders, early detection and intervention are emerging needs. This project will apply machine learning to EHRs and textual data for the prediction of patient mental health outcomes using features like psychological indicators to identify patterns and correlations that could lead to better interventions and support systems for mental health.

Methods

The following project will be making use of machine learning algorithms such as

- **Logistic Regression:** probably the most used model in binary classification and quite ideal for predictions of outcomes based on the features of patients.
- **Decision Trees:** A decision-making model that presents different courses of a decision and their probable outcomes, used in this study to determine the key predictors of mental health indicators.
- **Random Forest:** This is an ensemble of decision trees that improves the generalization of predictions by cancelling out overfitting and manages complex interaction among features.
- **Gradient Boosting:** This is a strong algorithm that creates the models in a sequence, with each model correcting the errors made by the previous one; it usually leads to high predictive performance.

Some of the major steps involved in this project will include cleaning the data, engineering its features, and training the model. Feature selection methods will be used to determine the top factors responsible for mental health.

Possible Challenges

- **Data imbalance:** There may be imbalances in the mental health outcome and strategies such as under sampling or cost-sensitive learning would contribute to enhancing the performance of such models.
- **Feature Engineering:** Feature extraction from text data and EHR will be crucial in improving model performance.

Dataset Used

The dataset to be used for this project is the Mental Health Dataset provided by Kaggle. Among the features of this dataset are psychological aspects, occupation, gender, family history, etc. The data will be taken from the link

<https://www.kaggle.com/datasets/bhavikjikadara/mental-health-dataset/data>

Expected Deliverables

- **Code implementation:** The code implements various machine learning models on data in Logistic Regression, Decision Trees, Random Forest and Gradient Boosting to predict mental health outcomes.
- **Outcome and interpretation:** These include a report on the findings from the models on the major predictors of mental health outcomes that provide actionable insight into early interventions.
- **Comparison of Models:** Comparing the performances for various machine learning models and listing some metrics such as accuracy, precision, recall, F1-score. This will contribute to the growth of mental analytics, since the project applies data-driven insights with predictive tools for a better understanding and addressing of the challenges at hand in mental health.