Capstone Three Project Proposal: Predicting the Length of Stay in Hospital among Schizophrenic and Other Psychotic Disorders Patients Using Machine Learning:

Problem Identification

Problem statement formation

The length of hospital stay (LOS) is a critical measure in healthcare, particularly in psychiatry, as it impacts both the quality of care and resource allocation. Prolonged hospitalizations among patients with schizophrenia and psychotic disorders can increase healthcare costs and affect patients' mental health outcomes. This project aims to apply machine learning techniques to predict the LOS of such patients, enabling healthcare professionals to manage resources and develop personalized discharge plans proactively.

Context

Predicting LOS is crucial for hospital resource management, personalized care planning, and improving patient outcomes. Early identification of patients with potentially extended stays could help healthcare providers plan interventions and allocate resources more efficiently.

Criteria for success

The following criteria will measure the success of this project:

- *Predictive Accuracy*: The model should accurately predict the length of hospital stay with high accuracy, accounting for varying patient conditions.

- *Actionability*: The model should provide healthcare providers with actionable insights, such as identifying the most significant factors contributing to longer hospital stays.

- *Usability:* The final appropriate model should be user-friendly and easily incorporated in the existing healthcare system for practical use

Scope of solution space

This project will focus on developing a machine learning model to predict LOS among patients with schizophrenia and other psychotic disorders. The project will involve data collection, preprocessing, feature engineering, model development, and evaluation. Various machine learning algorithms, including decision trees, random forests, will be explored. Feature selection techniques will be used to identify the most relevant predictors of LOS. The model will be trained and validated using a comprehensive dataset of patients with schizophrenia and psychotic disorders.

Constraints

The availability and quality of data will constrain the project. Access to a comprehensive dataset with relevant health and lifestyle factors is essential. Another constraint is that the model must comply with healthcare regulations and data privacy laws, protecting the confidentiality and security of patient's healthcare information.

Stakeholders

The primary stakeholders for this project include:

- *Healthcare providers:* Physicians, psychiatrists, and nurses who can use the model to plan patient care and hospital resources
- *Healthcare Organizations*: Hospitals and clinics looking to optimize patient flow and reduce operational costs.
- *Researchers*: Academics and medical professionals who focus on improving psychiatric care through predictive analytics
- *Patients*: Individuals with schizophrenia and other psychotic disorders who may benefit from improved care management and outcomes

Data Sources

This project will use a large open healthcare dataset provided by the New York State Statewide Planning and Research Cooperative System (SPARCS) containing 2.3 million de-identified patient records.

https://health.data.ny.gov/Health/Hospital-Inpatient-Discharges-SPARCS-De-Identified/tg3i-cinn/about_data

Proposed solution approach

1. *Data Collection and Preprocessing:* Gather and preprocess data from the identified sources and perform data cleaning, handling missing values and transforming variables as needed.

2. *Exploratory Data Analysis (EDA):* Perform EDA to understand the distribution of variables and identify potential patterns related to LOS. This will help in feature selection and understanding correlations between patient factors and LOS.

3. *Feature engineering:* Extract relevant features, such as patient demographics, clinical history, diagnosis, medication, and comorbidities, to improve model performance. Use feature selection techniques to refine the most important predictors of LOS.

4. *Model selection and training:* Explore various machine learning algorithms (e.g., logistic regression, decision trees, random forests) and train models on the preprocessed data.

5. *Model evaluation and selection:* Evaluate the performance of different models using appropriate metrics (e.g., accuracy, sensitivity, specificity). Then, select the best-performing model based on these metrics and its interpretability.

6. *Model Interpretation*: Analyze the model to understand the significance of various predictors and provide actionable insights for healthcare providers regarding the factors contributing to longer hospital stays.