

Capstone Two - Project Proposal - Hector Sanchez

Proposal

Outcome Prediction for Patients After having a Stroke

- Build a model to predict patient outcomes post-stroke, including recovery time, likelihood of recurrence, and long-term impacts on quality of life. Use data on stroke-severity, treatment methods, and patient demographics

Introduction:

Strokes, which are often referred to as brain attacks, are a medical condition caused by something blocking blood supply to part of the brain, or when a blood vessel bursts in the brain. It is often considered a leading cause of disability and mortality worldwide. Finding a way to predict a patient's outcome after having had a stroke is critical for improving treatment plans, patient care, and quality of life. My goal/focus is to build a predictive model that can forecast recovery time, likelihood of stroke recurrence, and long term impacts on quality of life. This will be done with the use of comprehensive datasets.

Objectives:

- **Build a Predictive Model:** Construct a model to predict patient outcome post-stroke
- **Analyze Recovery Time:** Estimate the time required for patients to recover after having a stroke.
- **Determine Recurrence Likelihood:** Find the probability of stroke recurrence.
- **Examine Long Term Quality of Life:** Assess the long term impacts on a patients' quality of life based on several variables/factors.

Datasets of interest:

- **Patients Treated with Suspected Stroke: Analysis of the Association between Diagnosis and Clinical Outcome**
 - a. <https://b2find.dkrz.de/dataset/50207a53-8f44-55ad-b2b3-a000006d0d77>
 - b. Contains data on stroke severity, treatment types, and patient demographics

- c. Includes detailed medical history and follow up information regarding outcomes such as either death, or hospital discharge.
- **Functional Outcomes Among Stroke Patients in Alagoas, Brazil: Observational Study**
 - a. <https://commons.datacite.org/doi.org/10.6084/m9.figshare.14306380.v1>
 - b. Provides metrics related to patient recovery times, rate of stroke recurrence, and quality of life.
 - c. Includes a diverse patient population with different demographics

Methods/Approach:

1. Data Collection and Preprocessing

- a. Collect and import data from the datasets listed above
- b. Handle missing values, normalize the data, and standardize the data
- c. Feature extraction and engineering to create predictors

2. EDA(Exploratory Data Analysis)

- a. Perform visualization and descriptive statistics to understand data distributions
- b. Identify correlations between variables

3. Model Development

- a. Divide data into training, test, and validation sets
- b. Use machine learning algorithms (linear regression, logistic regression, random forest models) to create predictive models.
- c. Evaluate model performance

4. Model Evaluation and Validation

- a. Use cross validation to determine model performance.
- b. Select the best performing model by comparing multiple model's performance.

5. Outcome and Prediction Analysis

- a. Once a model is selected, use it to predict patient recovery time, recurrence likelihood, and long term quality of life.
- b. See how other factors can impact these outcomes.

Goals/Expected Outcomes of this Project:

- 1. **Creating an Accurate Prediction Model:** A reliable model to help predict a patient's stroke recovery time, probability of stroke recurrence, and their long term quality of life.

2. **Greater Insight/Knowledge into Stroke Outcomes:** Obtaining a better understanding of the factors that influence a patient's outcome after they've had a stroke.
3. **Improved/Enhanced Patient Care:** Our insights that were obtained through data will help inform clinical providers and families regarding treatment decisions and treatment plans.

Conclusion:

The ultimate goal of this project is to use relevant and detailed datasets to build a predictive model for a patient's outcome after they've had a stroke. We strive to accurately forecast recovery times, stroke recurrence likelihood, and the long term impacts on quality of life. Given our success, the model will be instrumental in helping clinicians make better decisions so that patients can receive better care