

Gender and Stroke Correlation: A Data Mining Research Project

Objective:

Our project aims to utilize data mining tools and algorithms to parse research analyzing attributes and their correlation to the likelihood of an individual experiencing a stroke.

Dataset Utilized:

<https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset/data>

Research Question:

#1 We wish to extrapolate what factor(s), if any, sex plays in the likelihood of having a stroke for certain higher risk individuals.

Do the following Risk Factors increase prevalence of stroke for individuals of a specific sex?

Categorical

- Hypertension
- Marital Status
- Work Type
- Residential Type
- Smoking Status
- Heart Disease
- *Stroke

Quantitative

- Age
- Glucose Level
- BMI

#2 We aim to identify any outliers in the age of people who have previously experienced a stroke using anomaly detection principles.

#3 We wish to determine the association between glucose level and BMI, if any, in persons who have experienced a stroke.

Interim Timeline:

High-level tentative schedule:

Stage 1: Proposal

Due: Sept. 26

Stage 2: Data Cleaning and Processing

Due: Late October

Stage 3: Data Analysis and Implementation

Due: Late November

Stage 4: Term Paper/In-class Presentation

Due: Before December 10

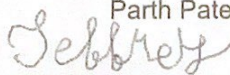
More specifically:

1. Determine attributes required for our research
2. Parse data for noise and anomalies
3. Analyze the data
4. Determine the best format for the data
5. Illustrate correlations between attributes and gender

Group 2 Members:



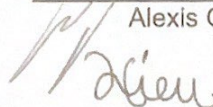
Parth Patel



Jeffrey Su



Alexis Gonzalez



Hieu Do