By: Andrew Santa & Waziha Khan

Proposal Report

Due: Aug 16, 2022

**Stroke Risk Prediction**

**Project Summary**

A website to predict stroke risk. The site allows input in the form of a short questionnaire, and outputs a risk profile. User input is added to the dataset.

**Project Description and Technologies:**

This project fits into a medical domain. It allows quick assessment of the user’s risk of stroke, extrapolated from an existing dataset. When it is used, the user’s inputs are added to and extend the dataset, improving its predictive ability.

This is useful to users due to its ease of use and access. Stroke risk can be assessed quickly from the web, without providing any personal identifying information. This information can inform the user and encourage a followup with medical professionals. Several similar services exist, most notably provided by stroke.org and hf.org.The stroke.org service requires manual scoring, increasing the workload on the user. The quizlet provided by hf.org focuses on medical history, and excludes users with low medical literacy. In contrast, our tool focuses on lifestyle factors that everyone knows about themselves (things like gender, employment status, living arrangement, age, etc).

The initial dataset is sourced from Kaggle.com (https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset) with a confidential original source. It consists of 5110 tuples of lifestyle and health values alongside a boolean indicating previous stroke occurrence. Each use of our tool will extend this dataset with the users’ responses.

Planned functionality begins with a quizlet that records the user’s responses and inserts them into the dataset. Based on their responses, the site indicates high or low risk of stroke. The user is given the opportunity to edit their responses (SQL update) or delete their data from the set. Another tab present on the navigation bar will allow searching of the dataset, displaying the demographics of the data (Various SQL SELECT statements). Additionally, we intend to implement several more advanced functions. Using multiple linear regression, we can display percentage-based increases in stroke risk based on a response (ex. “Males are X% more likely to have a stroke than females”). These predictions should improve as the user base increases, as that increases the size and diversity of the data set. Another proposed advanced function is suggesting what a user can change to most reduce their stroke risk. Using the output of the previously described advanced function, the site will suggest changing the highest weighted risk factor for the user. For example, if a given user indicates that they smoke, and the regression indicates that smoking is their highest changeable risk factor, it will suggest that they quit. Lastly, we can implement a feature that automatically detects if a user’s stroke risk is higher than average for their age group, and suggests local medical service providers using the Google Custom Search API.

**Dataset:**

Our dataset source is Kaggle.The data contains 5110 observations with 12 attributes.Our dataset file is in CSV format. (<https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>)

**Schedule:**

The general plan for our application will start by planning and setting up our technologies. Next, we plan to work on the majority of the skeleton of the website. Then, we will apply SQL onto our datasets and embed it on our website. With any time left, we plan on polishing and possibly adding more advanced features in our application. We plan on meeting every week in person and also on Teams.

Specifically, this is what we plan to accomplish each week:

**September 15:** Discussion about our project

**September 16:** Proposal Due

**September 20:** Work on setup of our program; Github set up, Visual studio environment/ IDE, languages downloaded,SQL.

**September 30 :** Development Plan Due

Add information to the database, Shareable database, brainstorm website details, and work on development

**Oct 2:** Complete our basic functions for the demo

**Oct 31:** Presentation and Demo Due

Make sure presentation and demo are complete

**November 1:** Finish up the Front end and back end

**November 06:** Finish up main website programming wise

**November 08:** Implement SQL into the website

**November 11:** Progress Report Due

Think of future plans, polish the application, and finish data integration

**November 28,29:** Final Presentation Due

Make sure the final project and essay are complete with future implementation

**December 03:** Submission of Project Due

**Conclusion:**

Our team is going to create a user-friendly environment for customers by allowing them to keep track of their health and easily predict the symptoms for stroke.As we know, stroke is a leading cause of death in the United States, killing more than 130,000 Americans each year—that’s 1 of every 20 deaths.