

Computer Science Capstone Topic Approval Form

The purpose of this document is to help you clearly explain your capstone topic, project scope, and timeline. Identify each of these areas to have a complete and realistic overview of your project. Your instructor cannot sign off on your project topic without this information.

Note: You must fill out and submit this form. The space beneath each number will expand as needed.

Note: Any costs associated with developing the application will be the student's responsibility.

INFORM INSTRUCTOR:

Potential use of proprietary company information: (Y/N) N

ANALYSIS:

1. Project topic and description:

My project will be an exercise in Supervised Learning using a logistic regression model to model binomial outcomes with multiple exploratory parameters. The model will predict whether a patient is likely to get a stroke based on various input parameters such as gender, age, numerous diseases, and smoking status. The project uses a stroke prediction dataset with a sample size of over 5,000 patients. The result will return a probability of whether a patient is likely to have a stroke; a probability of 0.5 or greater will return an answer of potential to have a stroke, while a probability of under 0.5 will return an answer of unlikely to have a stroke.

2. Project purpose and goals:

This project aims to use our training data to develop a model that assigns a numerical value to the importance of contributing factors when deciding if a person is likely to have a stroke. By the end of the project, we will have a working model trained on over 5,000 people, where many of the patients received strokes. We will use this model to return a probability on new data on whether or not this person will likely have a stroke.

3. Descriptive method:

The descriptive analysis will aim to observe the data set's main characteristics. Calculating averages of factors such as age will give us a more comprehensive understanding of the data set. Displaying minimums and maximums will allow us to understand outliers within the data.



4. Predictive or prescriptive method:

The predictive analysis will use a logistic regression model using the scikit-learn python library to come up with an estimate for stroke likeliness. We have a large enough sample size, over 5,000 patients, allowing us to split the data into training and testing purposes. A majority of the data will be used for testing purposes, at least 2,500 patient entries. Some data, such as gender, will need to be transformed from its existing state as a string to a numerical value in order to fit our mathematical model. The logistic regression model aims at assigning a value to the binomial coefficients. At a high level, we assign a weighted value to the various input parameters regarding their impact on the occurrence of a stroke.

DESIGN and DEVELOPMENT:

1. Computer science application type (select one):

- Mobile (indicate Apple or Android)
- **Web**
- Stand-alone

2. Programming/development language(s) you will use: Python

3. Operating system(s) or platform(s) you will use: macOS Sequoia 15.1.1

4. Database Management System you will use:

The database management system will be a Google Collaboratory hosting a Jupyter Notebook for an easy new user setup.

5. Estimated number of hours for the following:

- i. Planning and design:3
- ii. Development:12
- iii. Documentation:15
- iv. Total: 30

6. Projected completion date: December 25th, 2024

IMPLEMENTATION and EVALUATION:

1. Describe how you will approach the execution of your project.

I will start by downloading the CSV format of my dataset. I can use the scikit-learn library with python to implement the logistic regression model. I will use a majority of the data to train the model, roughly 75%, with the remain being used to test the model since it will be considered new data. Next I will create visuals using matplotlib.



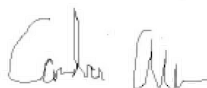
- ✓ **This project does not involve human subjects research and is exempt from WGU IRB review.**
- o **Confirmed. My project does not involve research on a human subject.**

STUDENT'S SIGNATURE

Zachary Trani

By signing and submitting this form, you acknowledge that any costs associated with the development and execution of the application will be your (the student's) responsibility.

INSTRUCTOR'S SIGNATURE:



INSTRUCTOR APPROVAL DATE:

12/22/2024

