***Project Requirement and Specification***

**On**

***STROKE DISEASE PREDICTION***



**Submitted to: Submitted by:**

Mr. Aniruddha Prabhu Manisha Basera

(CC-CSE-A-VI-Sem) Roll No.:2018468

CSE-A-VI-Sem

Session:2022-2023

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**GRAPHIC ERA HILL UNIVERSITY, DEHRADUN**

**2022-23**

***CERTIFICATE***

Certified that Miss. Manisha Basera (Roll No.-2018468) has develop mini project On “***Stroke Disease Prediction***” for the CS V Semester Mini Project Lab (PCS-404) in Graphic Era Hill University, Dehradun. The project carried out by Student is their own work as best of my knowledge.

Mr. Aniruddha Prabhu

Class Co-ordinator

CSE-A-VI-SEM

(CSE Department)

GEHU Dehradun

***ACKNOWLEDGEMENT***

Here by I am submitting the project report on **“Stroke disease prediction”** as per the scheme of Graphic Era Hill University, Dehradun.

I would like to express our sincere gratitude to My Teachers who helped me for providing a congenial environment to work in and carry out our project, and for the invaluable guidance extended at every stage and in every possible way.

Finally I am very much thankful to all the faculty members of the Department of Computer Science and Technology, friends and our parents for their constant encouragement, support and help throughout the period of project conduction.

**Manisha Basera**

**Roll No.-2018468**

**CSE-A-VI-Sem**

**Session:2022-2023**

**GEHU, Dehradun**

***Table of Contents***

1.Abstract

2.Introduction

3.Problem Statement

4.Methodology

6. Tools & Software used

7. Conclusion

8. Reference

**ABSTRACT**

Stroke is the second leading cause of death worldwide and remains an important health burden both for the individuals and for the national healthcare systems.

our project applies principles of machine learning over large existing data sets to effectively predict the stroke based on potentially modifiable risk factors.

Then it intended to develop the application to provide a personalized warning based on each user’s level of stroke risk and a lifestyle correction message about the stroke risk factors.

***INTRODUCTION***

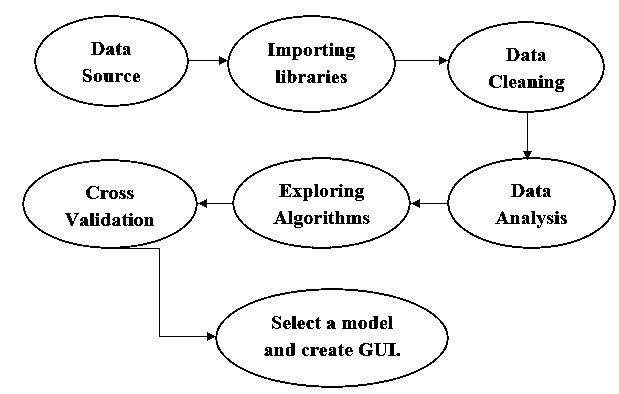
Stroke is the second leading cause of death worldwide and one of the most lifethreatening diseases for persons above 65 years. It injures the brain like “heart attack” which injures the heart. Once a stroke disease occurs, it is not only cost huge medical care and permanent disability but can eventually lead to death. Every 4 minutes someone dies of stroke, but up to 80% of stroke can be prevented if we can identify or predict the occurrence of stroke in its early stage. Stroke is a blood clot or bleed in the brain which can make permanent damage that has an effect on mobility, cognition, sight or communication. Stroke is considered as medical urgent situation and can cause long-term neurological damage, complications and often death. The majority of strokes are classified as ischemic embolic and Hemorrhagic. An ischemic embolic stroke happens when a blood clot forms away from the patient brain usually in the patient heart and travels through the patient bloodstream to lodge in narrower brain arteries. Hemorrhagic stroke is considered another type of brain stroke as it happens when an artery in the brain leaks blood or ruptures. Strokes are sudden but many of the disease processes that precede them take a long time to develop. This is why age is the most clear-cut risk factor for stroke: the chance of blockage or breakage rises with every passing year, so – although it can strike at any age – stroke is much more likely the older we get. The stroke risk factors included in the profile are age, systolic blood pressure, BMI, cholesterol, diabetes, smoking status and intensity, physical activity, alcohol drinking, past history (hypertension, coronary heart disease) and family history (stroke, coronary heart disease). On the basis of the risk factors in the profile, which can be readily determined on routine physical examination in a physician's office, stroke risk can be estimated. An individual's risk can be related to the average risk of stroke for persons of the same age and sex.

**PROBLEM STATEMENT**

Stroke is the second leading cause of death worldwide and remains an important health burden both for the individuals and for the national healthcare systems. Potentially modifiable risk factors for stroke include hypertension, cardiac disease, diabetes, and dysregulation of glucose metabolism, atrial fibrillation, and lifestyle factors.

Therefore, the goal of our project is to apply principles of machine learning over large existing data sets to effectively predict the stroke based on potentially modifiable risk factors. Then it intended to develop the application to provide a personalized warning based on each user’s level of stroke risk and a lifestyle correction message about the stroke risk factors.

***Methodology***



Data source

* *Primary data:-* Collected reference materials and common features/attributes by consulting neurologists and cardiologists

### *Secondary data :-* Obtained from Kaggle website. “Healthcare-dataset-stroke data” which is publicly available. Totally consists of 5110 entries, out of which 2995 were female and 2115 were male. Consists of 12 features.

### Libraries imported

#### NumPy

Python library which deals with arrays, basically used for scientific computations. Used for performing linear algebra, matrix multiplication, Fourier transform.

#### Pandas

Used analyze data. Works on various file formats such as SQL, JSON, Microsoft Excel. Data manipulation operations such as merging, selecting, reshaping and data cleaning in general.

#### Matplotlib Pyplot

Has collection of functions that makes matplotlib works like MATLAB. Basically, used for data visualization, also includes functions such as creating a figure, plotting area etc.

#### Seaborn

It is a data visualization based on matplotlib. It provides high-level interface for drawing informative and attractive graphs.

#### Tkinter

It is a simple and easy to use in-built python model used for designing user interface. Here this module is used to create 1920x1080 window with the necessary widgets for taking inputs.

### Data cleaning

* Data was cleaned for missing data and null values. Missing data was dealt by removing the rows with null values or redundant values.

### Data analysis

* There are three types of data analysis which is performed i.e.,Categorical feature analysis, Numerical feature analysis and Multicollinearity analysis. Data analysis is done to show us the hidden relationships and attributes present in the dataset which help the machine learning model to perform better.

### Implementing algorithms

* Five different algorithms are selected after literature survey. Comparative study is made between these five algorithms - Decision Tree, Logistic Regression, Random Forest, Support Vector Machine and K Nearest Neighbor.

### Cross validation

* Effectiveness of all the models is verified to solve overfitting problems. Overall assessment on how the model will perform for an independent test dataset.
* Finally, the best performing model will be used to predict stroke using the input data given by the user.

### Creating GUI

* The graphical user interface is a user interface which permits users to interact with digital gadgets through graphics and the audio indicator, as opposed to textual content, mostly based on user interfaces, written command labels or textual content. In response to an impressive curve of command-line interfaces that requires directions to be typed on a pc-keyboard, GUIs were provided.

***Tools And Software Used***

In order to accomplish this project following tools were used :-

1. A Code Editor it can be Notepad but here we are using Visual Studio Code Editor.
2. Languages used

* PYTHON
* HTML
* CSS

Conclusion

The completion of the project went quite well, I learned many new things while I was building it, and I get up to know various platforms which help us to learn all this stuff. I was able to learn the practical use of various algorithm. The practice helped me to learn the debugging of code and many more things.

Overall working on this project was great fun as I came up with a great piece of knowledge and understanding of the topic.

Reference

<https://www.sciencedirect.com/science/article/pii/S2772442522000090>

[https://ieeexplore.ieee.org/document/9734197/references#references](https://ieeexplore.ieee.org/document/9734197/references" \l "references)