

|  |  |
| --- | --- |
| BRAIN STROKE PREDICTION  Report |  |
| Name: Imran Ali Soomro  Submitted to: Dr. Sanam Narejo & Ms. Roshni Mustafa  CS NAVTTC Batch-III, Big Data Analytics. |  |

Table of Contents

[ABSTRACT: 3](#_Toc114071466)

[BACKGROUND STUDY: 3](#_Toc114071467)

[INTRODUCTION: 3](#_Toc114071468)

[AIM & OBJECT: 4](#_Toc114071469)

[ALGORITHM USED: 4](#_Toc114071470)

[DATASET USED: 4](#_Toc114071471)

[TOOLS & TECHNOLOGY: 4](#_Toc114071472)

[RESULT: 5](#_Toc114071473)

[CONCLUSION: 7](#_Toc114071474)

[REFERENCES: 8](#_Toc114071475)

# ABSTRACT:

Machine Learning (ML) delivers an accurate and quick prediction outcome, and it has become a powerful tool in health settings, offering personalized clinical care for stroke patients. Stroke is one of the major reasons for adult deaths around the globe, impacting 6.2 million people every year. To analyze dataset, we used different EDA approaches and applied ML different algorithms for brain stroke prediction, but here the performance of KNN is best, it provides highest accuracy. To get insights from data different visualization techniques are used: histogram, boxplot, scatter plot etc. that are enough to conclude my result. To avoid brain stroke one must follow the precautions to save the precious life.

# BACKGROUND STUDY:

ML is a crucial branch of Artificial Intelligence (AI). ML is the study of algorithms that could learn and improve from past data without being explicitly programmed, it is seen as a subfield of **Artificial Intelligence (AI)**, which aims to make machines behave intelligently like humans. The area of ML is divided into 3 parts, as given next:

* Supervised Learning: Type of machine learning that works on labeled data, and makes a function that maps the input to the given output.
* Unsupervised Learning: Type of machine learning that works on unlabeled data to draw useful patterns from it and generate decisions.
* Reinforcement Learning: Type of machine learning that aims to train an agent in an interactive environment to achieve its goal.

However, in this review, we focus on supervised learning, Supervised learning trains a model that maps an input to an output based on observations and predicts the output. It is categorized into classification and regression.  Classification is to classify discrete target variable using predictors. Logistic Regression & Support Vector Machine are example of Classification algorithms. Regression investigates the relationship between numerical target variable and predictors.

# INTRODUCTION:

The brain is one of our body’s largest and most complex organs. From scratching your nose to solving an intrinsic problem, we have our brains to thank for our every thought and our every action. [2] Even before you leave the womb, your brain works throughout your life to control your body’s functions and helps you understand and interact with the world around you. Maintaining a healthy brain will help your mind stay clear and active, so that you can continue to work, rest and play. [1]

Chronic conditions like diabetes and high blood pressure together with family history and the way we live our lives have an impact on the healthy function of our brains. All these factors can increase the risk of developing diseases like Alzheimer’s and vascular dementia. We can’t change our genetic heritage, but we can make lifestyle changes that can reduce the risk of developing dementia and mild cognitive decline. Caroline Abrahams, director of Age UK said: “*The changes that we need to make to keep our brains healthy are already proven to be good for the heart and overall health, so it’s common sense for us all to try to build them into our lives.*” [1]

Stroke is one of the major reasons for adult deaths around the globe, impacting 6.2 million people every year. Over the past two decades, there has been a 26 percent increase in stroke deaths, worldwide. Stroke is the second leading cause of death across the globe. Brain stroke occurs when the blood flow to the brain is stopped or when the brain doesn't get enough blood. As a result, the part of the brain drained of blood supply experiences a shortage of oxygen and becomes unresponsive [3].

# AIM & OBJECT:

The aims and objectives of writing this document are, to know what parameters are affecting our healthy brain, and to do the brain stroke predictions, so that we can follow precautions related to brain health to save healthy lives. For that a dataset is downloaded from Kaggle, to make data analysis and apply different machine learning model with proper steps of data preprocessing (if any), and to generate brain stroke predictions. The objective also includes accessing the accuracy of the model(s) so that its skill can be defined.

# ALGORITHM USED:

There are so many ML used by peoples, but the performance of model depends on the data that is used. This dataset falls in the category of classification. So, for this project classification algorithms (models) are used:

* Logistic Regression
* Support Vector Machine (SVM)
* K-Nearest Neighbor (KNN)
* Artificial Neural Network (ANN)
* Decision Tree
* Random Forest Regression (RFR)

# DATASET USED:

A stroke is a medical condition in which poor blood flow to the brain causes cell death. There are two main types of strokes: ischemic, due to lack of blood flow, and hemorrhagic, due to bleeding. Both cause parts of the brain to stop functioning properly. The dataset consists of one csv file, having 11 columns and a total 4981 rows. Out of which 11 are independent variables and One is a target which is stroke. It used a training dataset to train the machine learning model to predict the brain stroke.

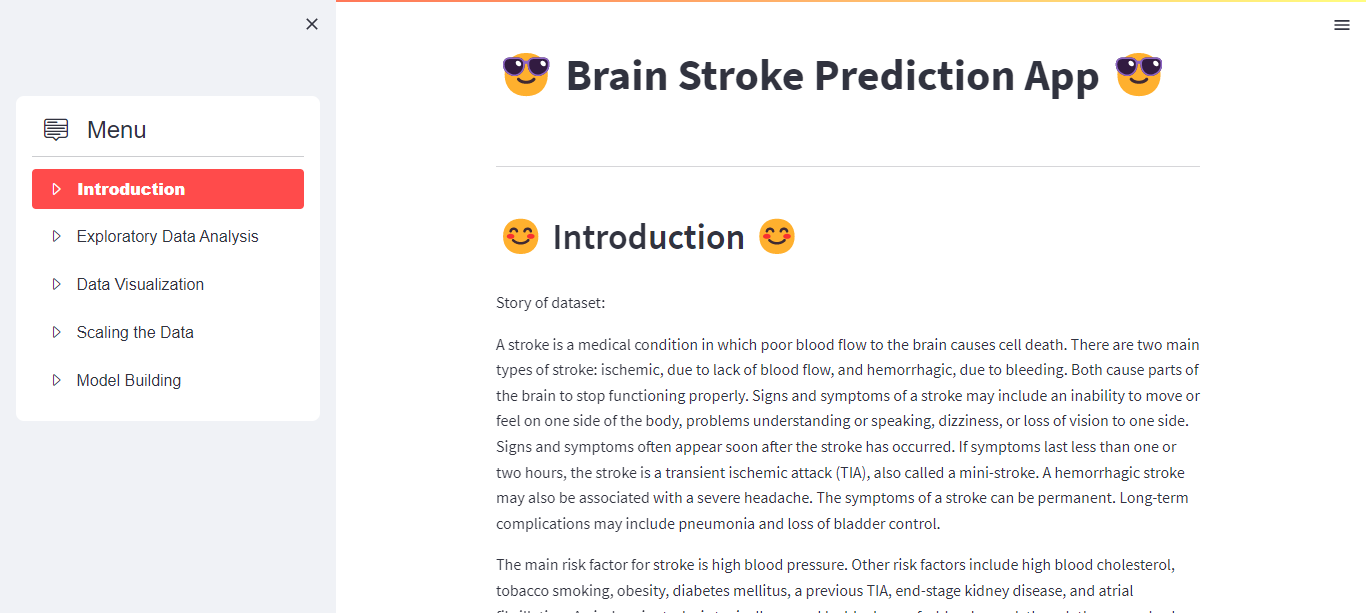
# TOOLS & TECHNOLOGY:

The data analysis, modeling, and preprocessing will be done by using Python Programming Language. Python has many libraries for data science and machine learning that we can effectively use to make our model. Following python libraries will be used:

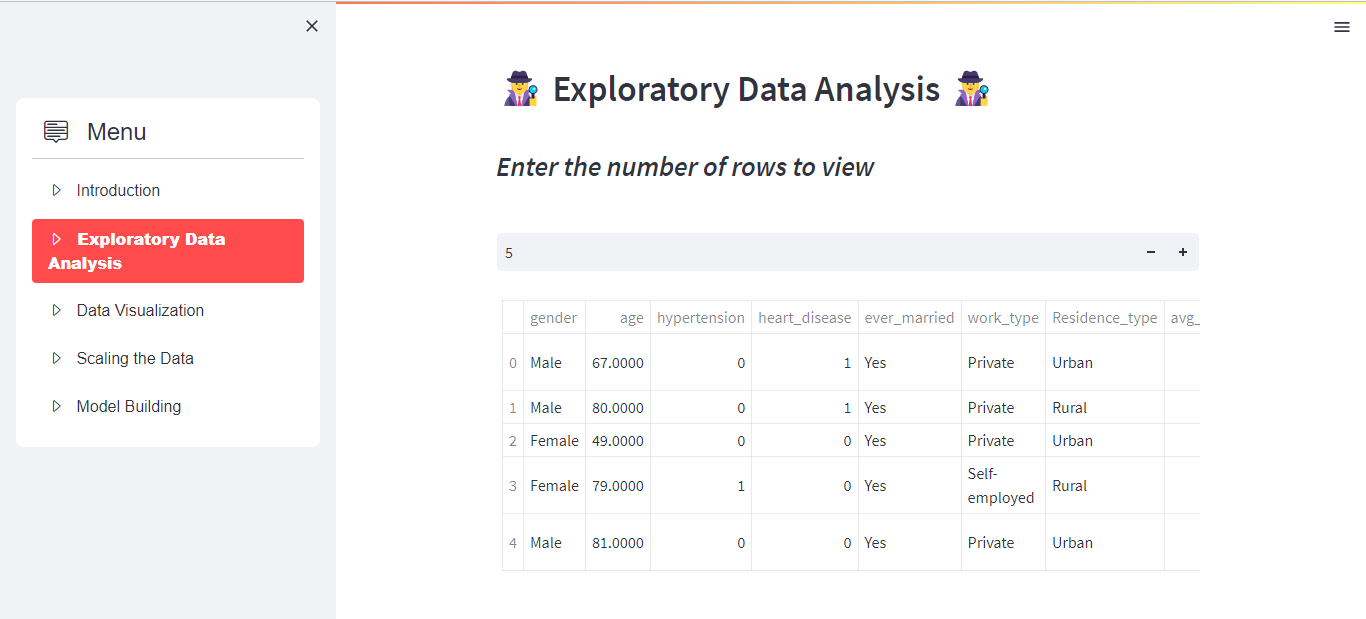
* Streamlit (for Creating Web App)
* NumPy (for mathematical formulation)
* Pandas (for data loading, preprocessing)
* Matplotlib (for data visualization)
* Seaborn (for data visualization)
* Plotly (for data visualization)
* PIL (for images)
* Scikit-learn (for machine learning)

# RESULT:

Page 1:



Page 2:



Page 3:

Graphical user interface

Description automatically generated with medium confidence

Page 4:

A picture containing table

Description automatically generated

Page 5:

Graphical user interface

Description automatically generated

# CONCLUSION:

After doing Exploratory data analysis & applying Logistic Regression, SVM, KNN, ANN, Decision Tree, Random Forest, it has been concluded:

* The Accuracy is above than 90% for all algorithms, it gives best in KNN.
* According to dataset most of the peoples having stroke lies in age 40 to 80.
* The Ratio of stroke is less in male then the females.
* The person is predicted having brain stroke it should do following precautions:
  + Brain training Keeping the brain active is an important aspect of brain health.
  + Live an active life – Regular exercise boost your muscles.
  + Eat well – A balanced diet can nourish your mind and your body.
  + Give up smoking – Smoking increases your risk of heart disease.
  + Stay social – Friends and family can be good for your brain health.
  + Get quality sleep – Sleep is a chance for our bodies to rest and repair the damage inflicted by daily life.

# 

# REFERENCES:

* [Why is brain health important? | hometouch | Expert Care Articles (myhometouch.com)](https://myhometouch.com/articles/why-is-brain-health-important) [1]
* [Why Brain Health is Important | Memory Loss and More (memoryhealth.com)](https://www.memoryhealth.com/blogs/the-brain-blog/why-brain-health-is-important) [2]
* [Neuroimaging and deep learning for brain stroke detection - A review of recent advancements and future prospects - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S0169260720315613#abs0001) [3]
* [Machine Learning for Brain Stroke: A Review - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S1052305720305802#abs0001) [4]