

Project Proposal: *Stroke Prediction*

Background:

Stroke is a serious disease that affects millions of people all around the world. It is a cardiovascular disease that blocks oxygen from arteries leading to and from the brain. There are three main types of strokes the first one is an ischemic stroke that blocks the flow of oxygen to a part of the brain, the second one is a hemorrhagic stroke which happens when the artery in the brain ruptures or leaks blood causing pressure on the cells around it, the last one is a transient ischemic attack (TIA) which is considered a mini-stroke where the block happens in a short time, not more than 5 minutes [1]. Stroke affects 1 in 4 people that are above the age of 25. Every year, approximately 13 million people experience their first stroke, with approximately 500 thousand people dying as a result [2]. It is considered the second leading cause of death globally [3].

Problems:

- Predict whether the patient will have stroke or not?
- Does the gender impact on strokes?
- Does the Hypertension impact on strokes?
- Does the marriage impact on strokes?
- Does the smoking impact on strokes?
- Does the age impact on strokes?
- What effects more BMI or the glucose level on strokes?

Data:

In this project, I will use a dataset from Kaggle [4]. It contains 11 features and one target, which is whether the patient has a stroke or not.

ID	Unique identifier
Gender	Male, female, other
Age	The age of the patient
hypertension	0 if the patient doesn't have hypertension, 1 if the patient has hypertension
heart_disease	0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease
ever_married	"No" or "Yes"
work_type	"children", "Govt_jov", "Never_worked", "Private" or "Self-employed"
Residence_type	"Rural" or "Urban"
avg_glucose_level	average glucose level in blood
bmi	body mass index
smoking_status	"formerly smoked", "never smoked", "smokes" or "Unknown"
Stroke	1 if the patient had a stroke or 0 if not

Tools:

In this project I will use these tools to help me to achieve the requirement of this project

Scikit learn I will use a couple of algorithms and compare it with each other

NumPy

Pandas

Seaborn

Matplotlib

Bokeh

SciPy

Keras

Project Proposal: *Stroke Prediction*

Reference:

- [1]: https://www.cdc.gov/stroke/types_of_stroke.htm#hemorrhagic
- [2]: https://www.world-stroke.org/assets/downloads/WSO_Fact-sheet_15.01.2020.pdf
- [3]: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- [4]: <https://www.kaggle.com/fedesoriano/stroke-prediction-dataset>