Analysis of the risk to make strokes

Analyses dataset to predict whether a patient is likely to get stroke based on the input parameters like gender, age, and smoking.

In this project we look for the factors that increase the risk of strokes.

Data

Content Description

-The data contains 12 columns and 5110 rows

1-Read data

2-Data clean

Now, I'm going to check the data on the following things:

Whether there are empty values in the data?

Whether data features are linked?

drop Nan value

convert text to category

EDA

I tried to answer some questions.

Does the person have a stroke or not?

Does gender or age affect stroke?

Does hypertension, glucose level, bmi, work type have any relation with stroke?

What is the most effective factor on stroke?

Algorithm

I used Decision tree model& Random Forest classifier

Split the dataset into two pieces: a training set **and a** testing set

1. **Train the model on the** training set**. (Used to build model)**
2. **Test the model on the** testing set **and evaluate how well we did.**

Measure scores and results

Tools

-Jupyter Notebook.

- Programming language: python.

Libraries: Pandas -NumPy for data processing

-Matplotlib for Visualization

- sklearn for model

communication

Presentation that includes visuals for communicating the objectives and findings