Feature Importance

We have feature importance chart as shown below. We cane see that it accurately predicted the importance of features.

‘Age’ has the most importance amongst the others, because model must have found the correlation of it with output variable(stroke).

Following ‘Age’ ,we can see that ‘average\_glucose\_level’ and ‘BMI’

also impacts more in deciding whether the person has stroke or not.

Graphical user interface, application

Description automatically generated

**BMI Histogram**

Chart

Description automatically generated

X-axis: BMI

Y-axis: Number of people

Here we can see that many people have their BMI from 20 to 40

Those having BMI below 24.5 are considered as healthy.

So we can conclude that there are more number of people who are unhealthy and have sedentary lifestyle.

Age vs Heart\_Disease\_Probability

Graphical user interface, application

Description automatically generated

As we see, heart disease probability increases with age. It increases linearly with age.

Confusion Matrix

Graphical user interface, chart, treemap chart

Description automatically generated

Precision = True Positive / (True Positive + False Positive)

Precision = 715 / (715+252) = 0.74

Recall = True Positive / (True Positive + False Negative)

Recall = 715 / (715+134) = 0.84

**F1\_Score = harmonic mean of precision and recall = 0.79**

F1\_score takes into account how data is distributed

Accuracy does not take that into account

**Here in this case, Our model predicted 134 False negatives.**

**Means Our model predicted 134 people don’t have stroke but actually they have it.**

**This is called as type 2 error.**

**This False negative is dangerous thing in this healthcare related scenario, we can minimise it by increasing sample size.**

Features Correlation

A picture containing graphical user interface

Description automatically generated

We can see that **age** and **bmi** are highly correlated.

After that **hypertension** and **age** are well correlated.

Similarly **heart disease** and **age** , **stroke** and **age** are greatly correlated.