Vivek A

Click here to view the github

Heart Stroke Prediction

Data Science project

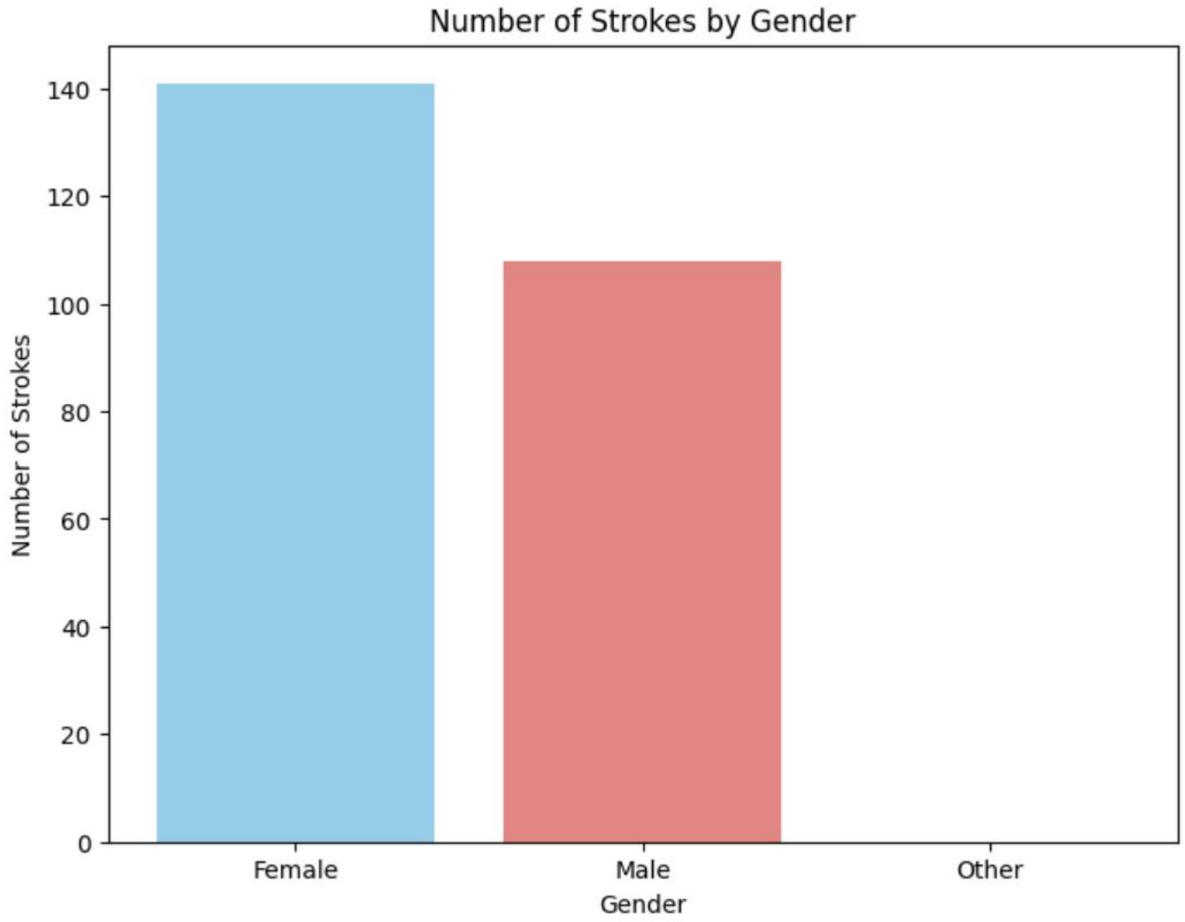
Problem statement

In this project, we aim to develop and compare machine learning models to predict the likelihood of a stroke occurrence based on various health-related features. Using a dataset that includes information on factors such as age, gender, medical history, and lifestyle choices, we will build models to help identify individuals at higher risk of stroke.

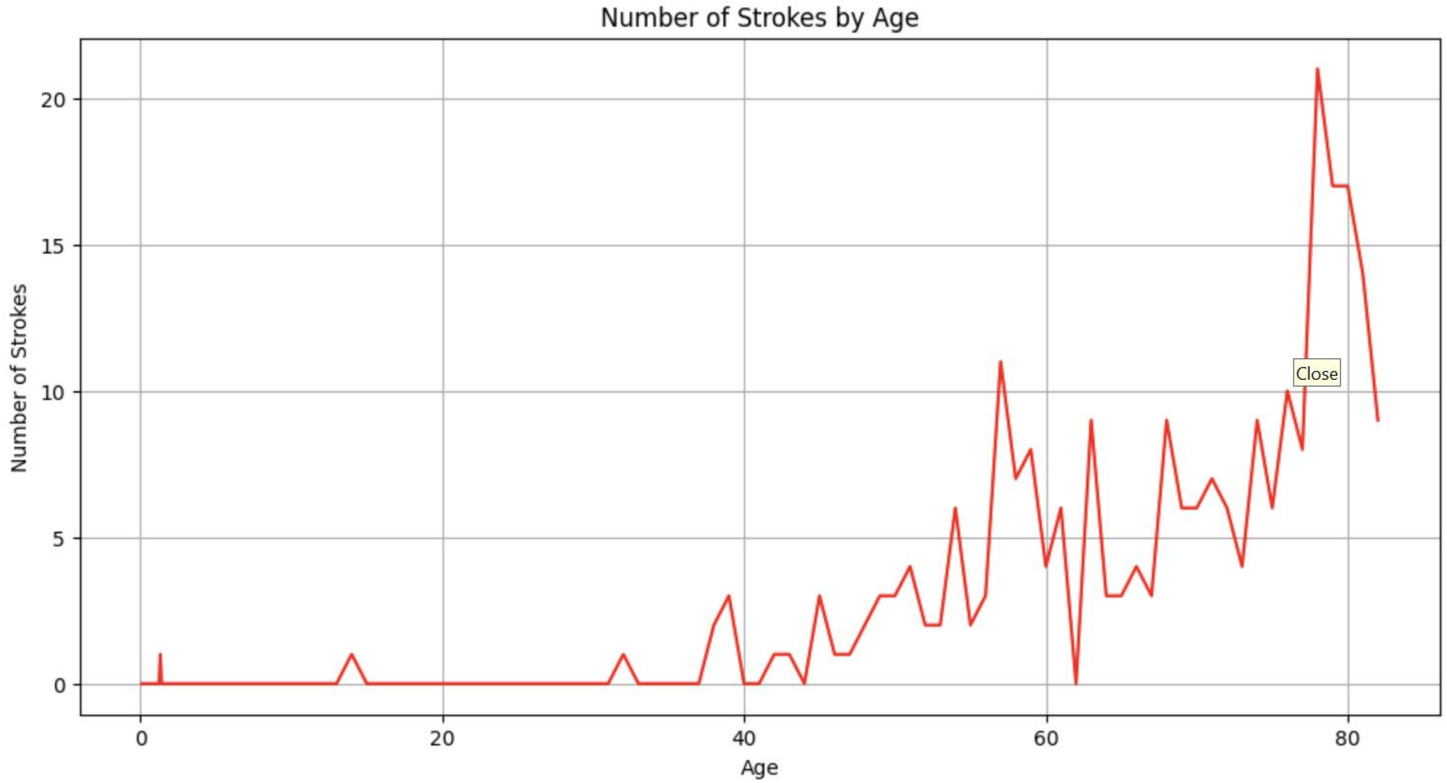
Insights from EDA

Number of strokes by gender

- females recorded more stroke cases than males and others



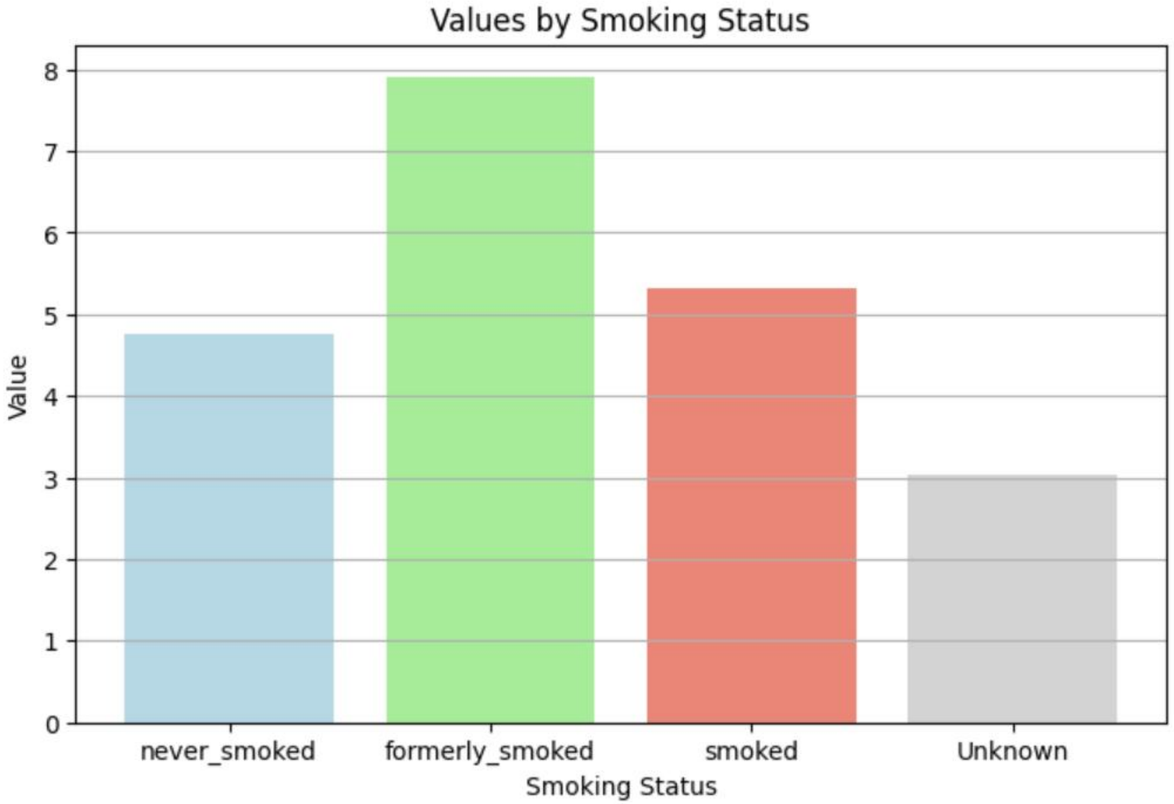
Number of strokes by Age



- stroke rate have a gradual increase after 50 years old

stroke per 100 people by smoking category

- formaly smoked reported more stroke cases for 100 people
- never smoked have lowest because we dont have data on unknown



Predictive Models

Logistic Regression

- Logistic regression model has 93.54 % accuracy

```
Accuracy: 0.9354207436399217
Confusion Matrix:
 [[954   6]
 [ 60   2]]
Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.99      0.97       960
           1       0.25      0.03      0.06        62

    accuracy                           0.94      1022
   macro avg       0.60      0.51      0.51      1022
weighted avg       0.90      0.94      0.91      1022

ROC AUC Score: 0.8496975806451613
```

- The model showing an accuracy of 93 percent.

Decision Tree

- Decision Tree model has 91.38 % accuracy

```
Accuracy: 0.913894324853229
Confusion Matrix:
 [[924  36]
 [ 52  10]]
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.96      0.95       960
           1       0.22      0.16      0.19        62

    accuracy                           0.91      1022
   macro avg       0.58      0.56      0.57      1022
weighted avg       0.90      0.91      0.91      1022

ROC AUC Score: 0.5618951612903227
```

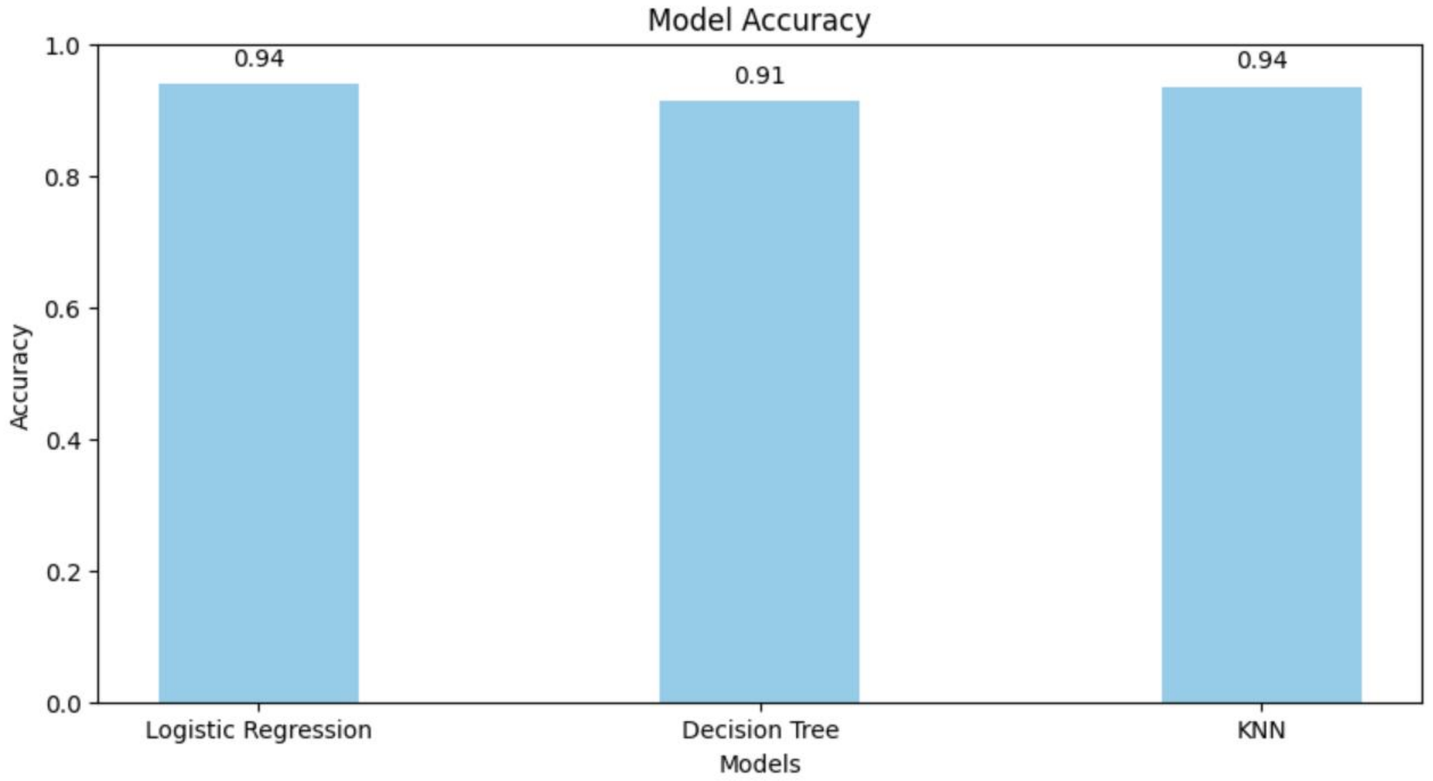
KNN

- KNN model has 93.54 % accuracy

```
Accuracy: 0.9354207436399217
Confusion Matrix:
 [[954   6]
 [ 60   2]]
Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.99      0.97       960
           1       0.25      0.03      0.06        62

    accuracy                           0.94      1022
   macro avg       0.60      0.51      0.51      1022
weighted avg       0.90      0.94      0.91      1022

ROC AUC Score: 0.6853578629032258
```

Model Accuracy

- In the model building phase all the models showed high accuracy with logistic regression and KNN with 94 percent.



Thanks

Thank you everyone who been with me with this journey.
Check the github repository for more info about the project