- Topic: Stroke Disease Detection and Prediction Using Robust Learning Approaches
- MOHAMMAD NUR (21-44540-1)
- SANDIP MISRA (21-44506-1)
- MD.RESAT AHMED (21-44542-1)
- SIRAJUS SALEHIN (21-44543-1)

- *Problem Statement:*

*A stroke is a severe medical condition caused by disrupted blood flow to the brain, leading to brain damage. It's a major global cause of death and disability. Early identification of stroke warning signs can help minimize its impact. Researchers have employed various machine learning methods to predict stroke risk using physiological data. This study focuses on using four models (Logistic Regression, Decision Tree, Random Forest, and Voting Classifier) to predict strokes accurately. The best-performing model, Random Forest, achieved around 96% accuracy. The research utilized the Stroke Prediction dataset and outperformed previous studies in accuracy. This investigation underscores the effectiveness and reliability of these models in stroke predictio*n.

- Objective:
- This study aims to create a computer program to predict if someone might have a stroke by looking at their body information. We want to find out which computer method works best for this. Also, we want to see if we can make the program even better by using more information and different computer techniques.

- **Proposed Algorithms**

- (i) Logistic Regression
- (i) SVM.
- (ii) Random Forest
- (iii) Decision Tree
- (IV)Naive Bayes Classifier

Data Collection

1)Identify Relevant Features

2)Public Datasets: World Health Organization (WHO).

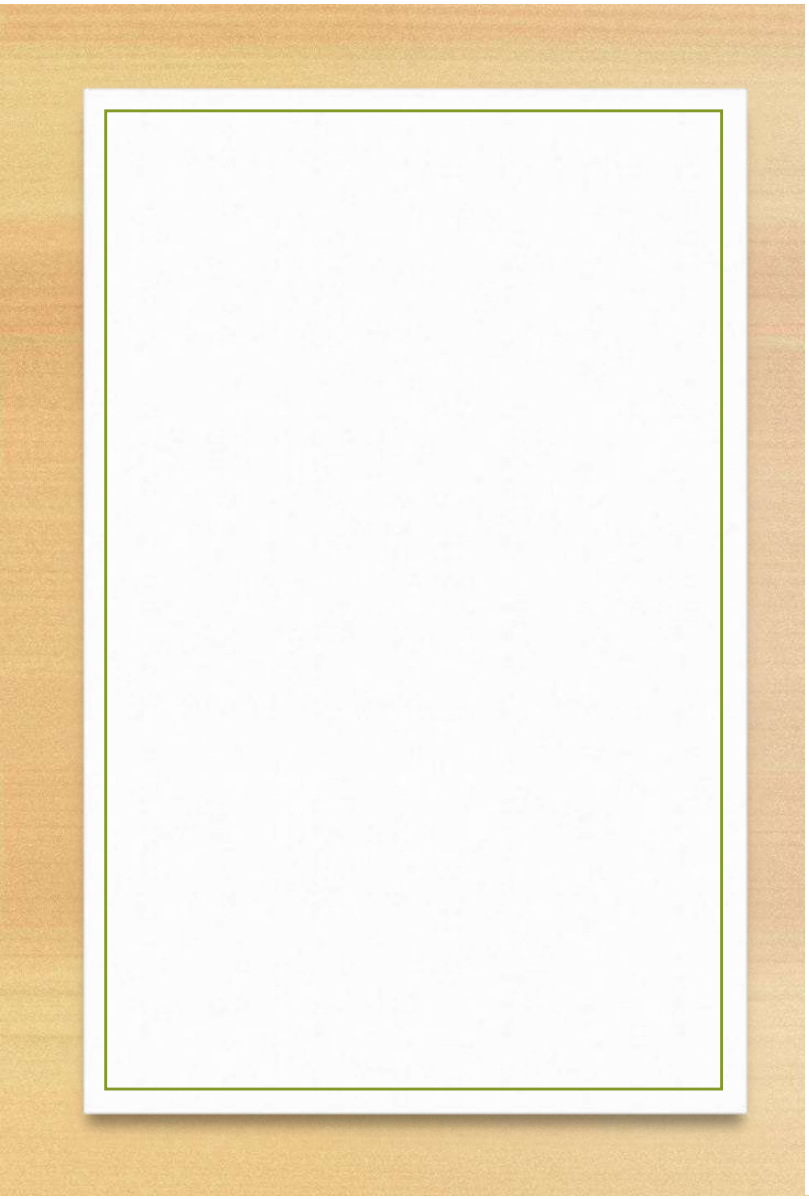
National Health and Nutrition Examination S urvey (NHANES).

3)Medical Records

4)Surveys

5)Diagnostic Imaging

6)Historical Data

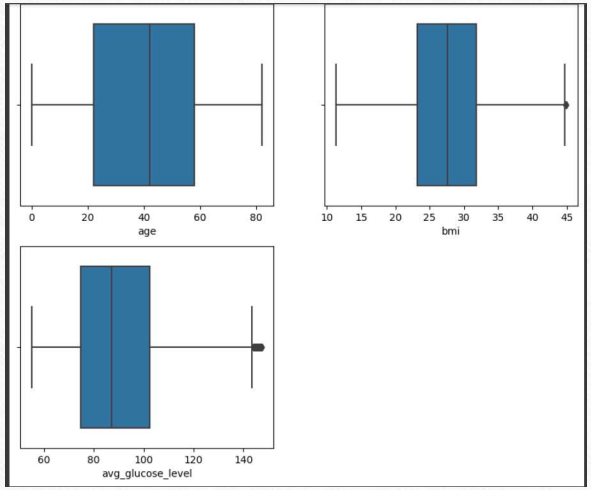


Data Nature and Classification

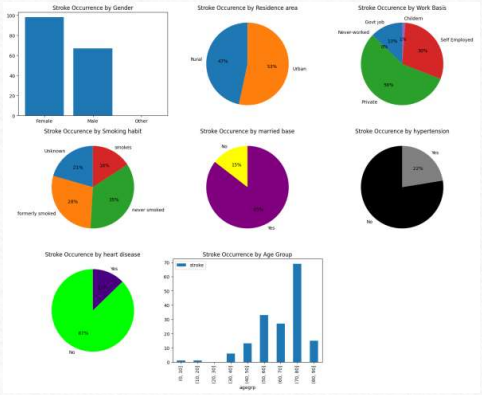
Physiological Variables, Dataset Size, Algorithm Performance, Cross-Validation(Random Forest Algorithm), Model Enhancement

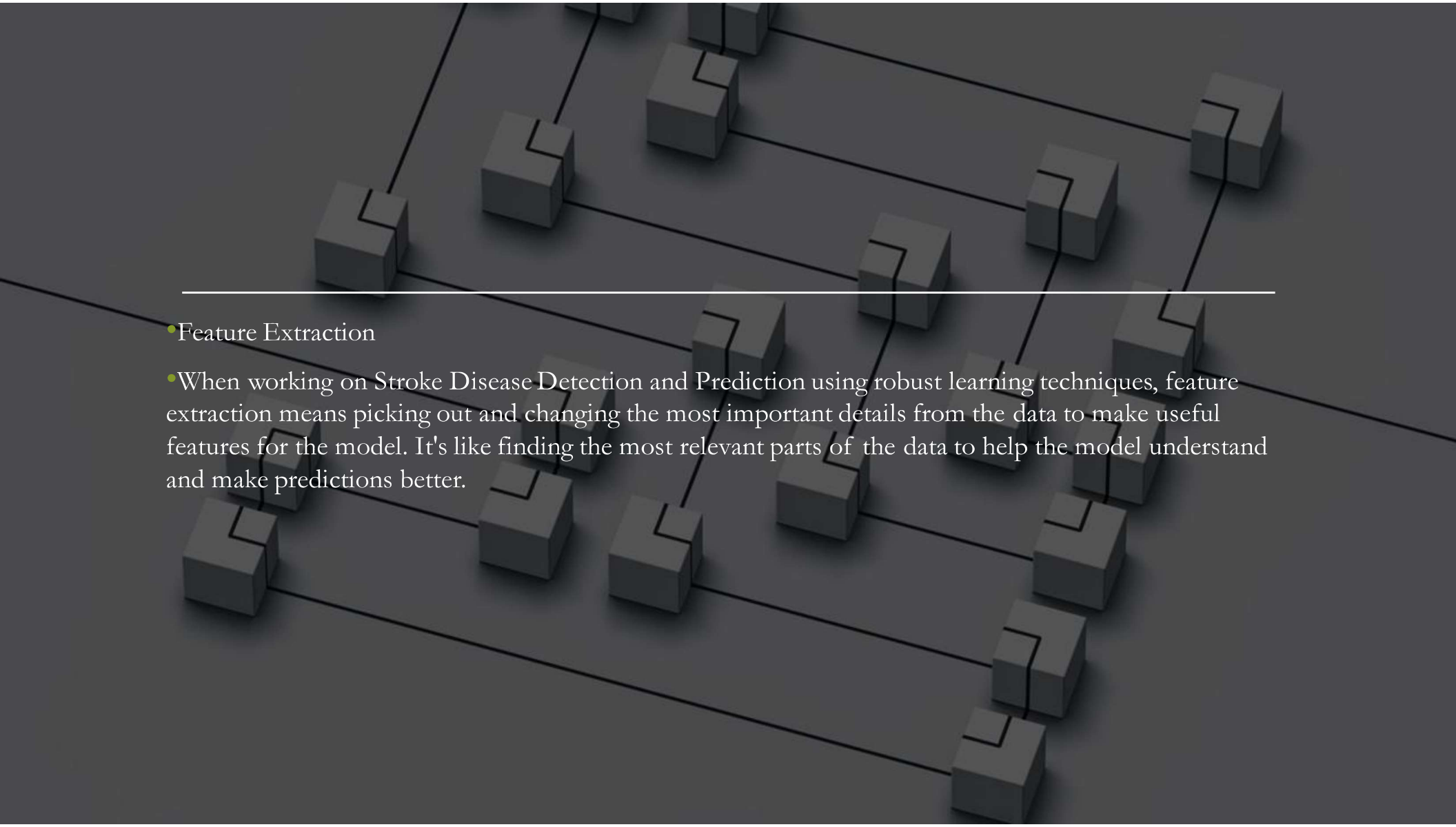
- Data preprocessing
- Removing Unnecessary Column.
- Handling Null Values.
- Label Encoding.
- Deal with Imbalanced Data.
- SMOTE Technique.
- Visualization.

Data out layer removing



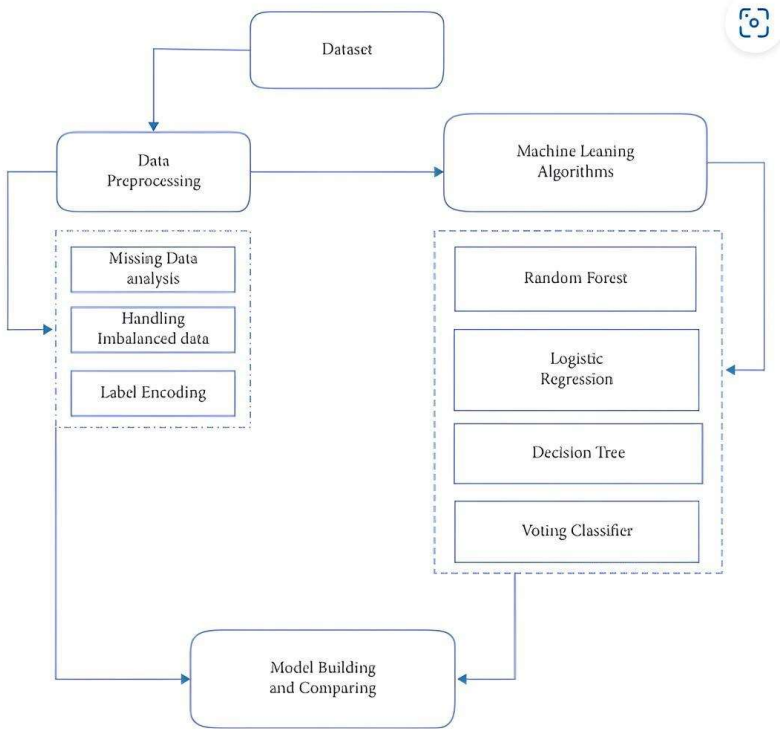
Several Types of people stoke prediction





- Feature Extraction
- When working on Stroke Disease Detection and Prediction using robust learning techniques, feature extraction means picking out and changing the most important details from the data to make useful features for the model. It's like finding the most relevant parts of the data to help the model understand and make predictions better.

ARCHITUC TUTE MODEL



Result AN ALYSIS 01

Naive Bayes Classifier

```
[71] from sklearn.naive_bayes import GaussianNB
     nb_model = GaussianNB()
     nb_model.fit(x_train , y_train)
```

```
GaussianNB
GaussianNB()
```

```
[72] y_pred = nb_model.predict(x_test)
     accuracy(y_test,y_pred)
```

```
0.3416856492027335
```

Decision Tree

```
[73] from sklearn.tree import DecisionTreeClassifier
     tree_model = DecisionTreeClassifier(criterion='entropy')
     tree_model.fit(x_train,y_train)
```

```
DecisionTreeClassifier
DecisionTreeClassifier(criterion='entropy')
```

```
[74] y_pred = tree_model.predict(x_test)
     accuracy(y_test, y_pred)
```

```
0.9328018223234624
```

Result ANALY SIS 02

Support vector machine

```
[75] from sklearn.svm import SVC
     svm = SVC()
     svm.fit(x_train , y_train)
```

```
SVC()
```

```
[76] y_pred = svm.predict(x_test)
     accuracy(y_test,y_pred)
```

```
0.9635535307517085
```

Random Forest

```
[77] from sklearn .ensemble import RandomForestClassifier
     rf_model = RandomForestClassifier(n_estimators=150,criterion='entropy',random_state = 123)
     rf_model.fit(x_train,y_train)
```

```
RandomForestClassifier(criterion='entropy', n_estimators=150, random_state=123)
```

```
[78] y_pred = rf_model.predict(x_test)
     accuracy(y_test,y_pred)
```

```
0.9635535307517085
```

- Conclusion: We learned that a computer method called "Random Forest" is good at telling if someone might have a stroke. It was right about 96 out of 100 times. This means the program could help know if someone might get a stroke. We also think that if we use more information and try different computer ways, we could make the program even better. The main goal is to help people find out about stroke risks early to get help and recover well if needed.

Thank you

Question Answer?