Tech Trek : AI & Data Science Discovery

INDIAN LIVER PATIENT DISEASE

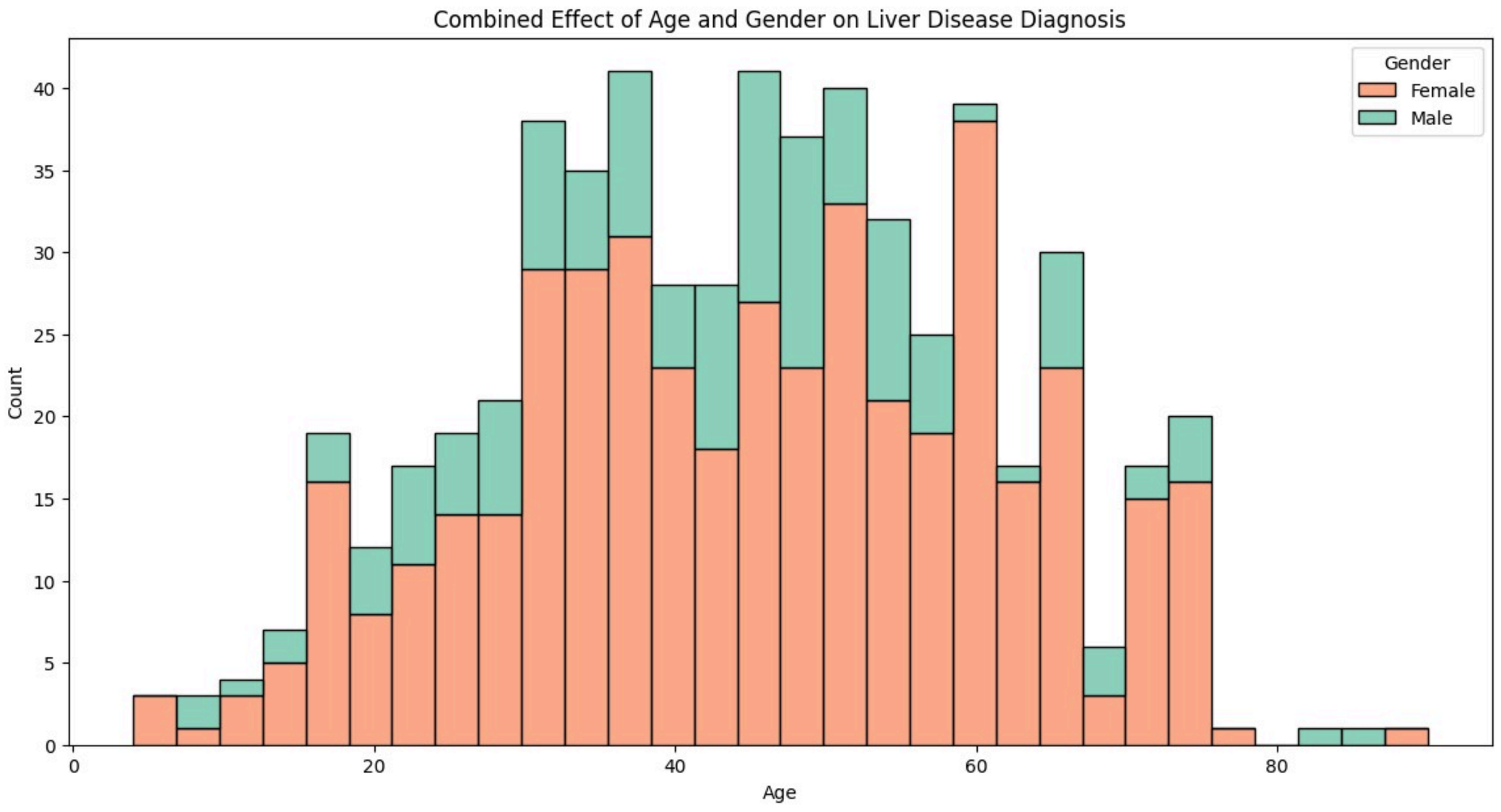
Presented By
Group 3
Dhruvil Makadia
Gagan Jaju
Sachin Guria
Saniya Sapale

PROBLEM STATEMENT

The data set contains 11 clinical and demographic features.

The objective of this project is to develop a predictive model for diagnosing liver disease in Indian patients.

The goal is to accurately predict the presence of liver disease, thereby aiding in early diagnosis and treatment.


Combined Effect of Age and Gender on Liver Disease Diagnosis
Gender
Female
Male
Count
Age
0
5
10
15
20
25
30
35
40
0
20
40
60
80

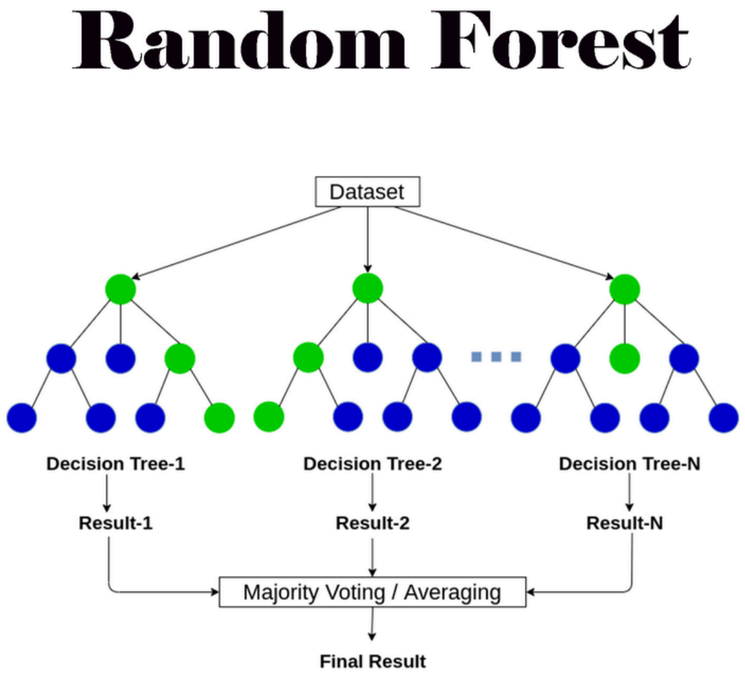

Pearson Correlation of Features
Age 1 0.057 0.012 0.0075 0.08 -0.087 -0.02 -0.19 -0.27 -0.22 0.14
Gender 0.057 1 0.089 0.1 -0.027 0.082 0.08 -0.089 -0.094 -0.03 0.082
TB 0.012 0.089 1 0.87 0.21 0.21 0.24 -0.0081 -0.22 -0.2 0.22
DB 0.0075 0.1 0.87 1 0.23 0.23 0.26 -0.00014 -0.23 -0.19 0.25
Alkphos 0.08 -0.027 0.21 0.23 1 0.13 0.17 -0.029 -0.17 -0.22 0.18
Sgpt -0.087 0.082 0.21 0.23 0.13 1 0.79 -0.043 -0.03 -0.0063 0.16
Sgot -0.02 0.08 0.24 0.26 0.17 0.79 1 -0.026 -0.085 -0.07 0.15
TP -0.19 -0.089 -0.0081 -0.00014 -0.029 -0.043 -0.026 1 0.78 0.24 -0.035
ALB -0.27 -0.094 -0.22 -0.23 -0.17 -0.03 -0.085 0.78 1 0.66 -0.16
A/G Ratio -0.22 -0.03 -0.2 -0.19 -0.22 -0.0063 -0.07 0.24 0.66 1 -0.14
Selector 0.14 0.082 0.22 0.25 0.18 0.16 0.15 -0.035 -0.16 -0.14 1
Age Gender TB DB Alkphos Sgpt Sgot TP ALB A/G Ratio Selector
1.0 0.8 0.6 0.4 0.2 0.0 −0.2

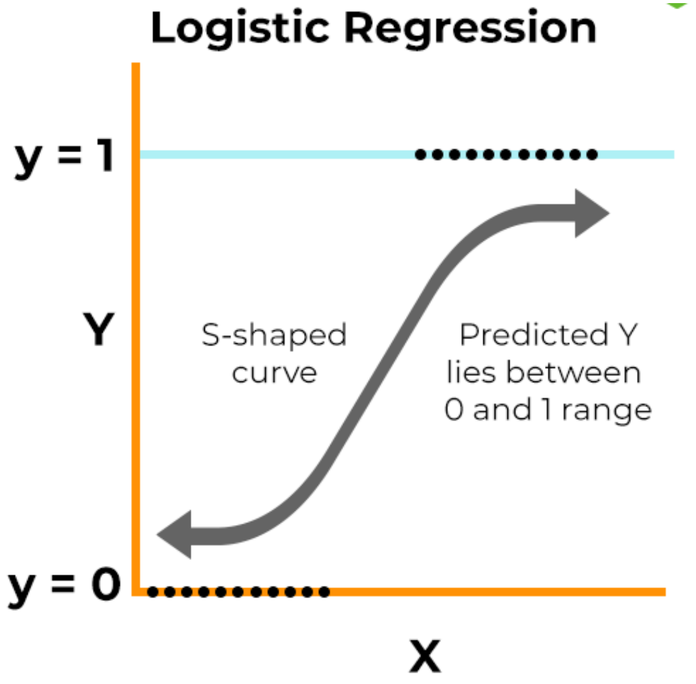
UTILIZED MODEL

RANDOM FOREST ALGORITHM AND LOGISTIC REGRESSION

Random Forest Algorithm is used for classification and regression tasks. It builds multiple decision trees during training and the mode of the classes (classification) or mean prediction (regression) of the individual trees is used for prediction.

Logistic regression is a supervised machine learning algorithm used for binary classification tasks, where the outcome is a categorical variable with two possible values. It models the probability that a given input belongs to a particular category.





PERFORMANCE OF THE MODEL

Random Forest Algorithm

```
#accuracy of the
accuracy= accuracy_score(y_test, y_pred_RF)
print(accuracy)

# confusion matrix
conf_matrix = confusion_matrix(y_test, y_pred_RF)
conf_matrix

# classification report precision recall f1-score
report = classification_report(y_test, y_pred_RF)
print(report)
```

```
0.6923076923076923
              precision    recall  f1-score   support

           0       0.39      0.45      0.42        29
           1       0.81      0.77      0.79        88

    accuracy                           0.69       117
   macro avg       0.60      0.61      0.61       117
weighted avg       0.71      0.69      0.70       117
```

Logistic Regression

```
# accuracy of the Logistic regression
accuracy = accuracy_score(y_test, y_pred)
accuracy
```

```
0.717948717948718
```

```
Start coding or generate with AI.
```

```
# confusion matrix
conf_matrix = confusion_matrix(y_test, y_pred)
conf_matrix
```

```
array([[10, 19],
       [14, 74]])
```

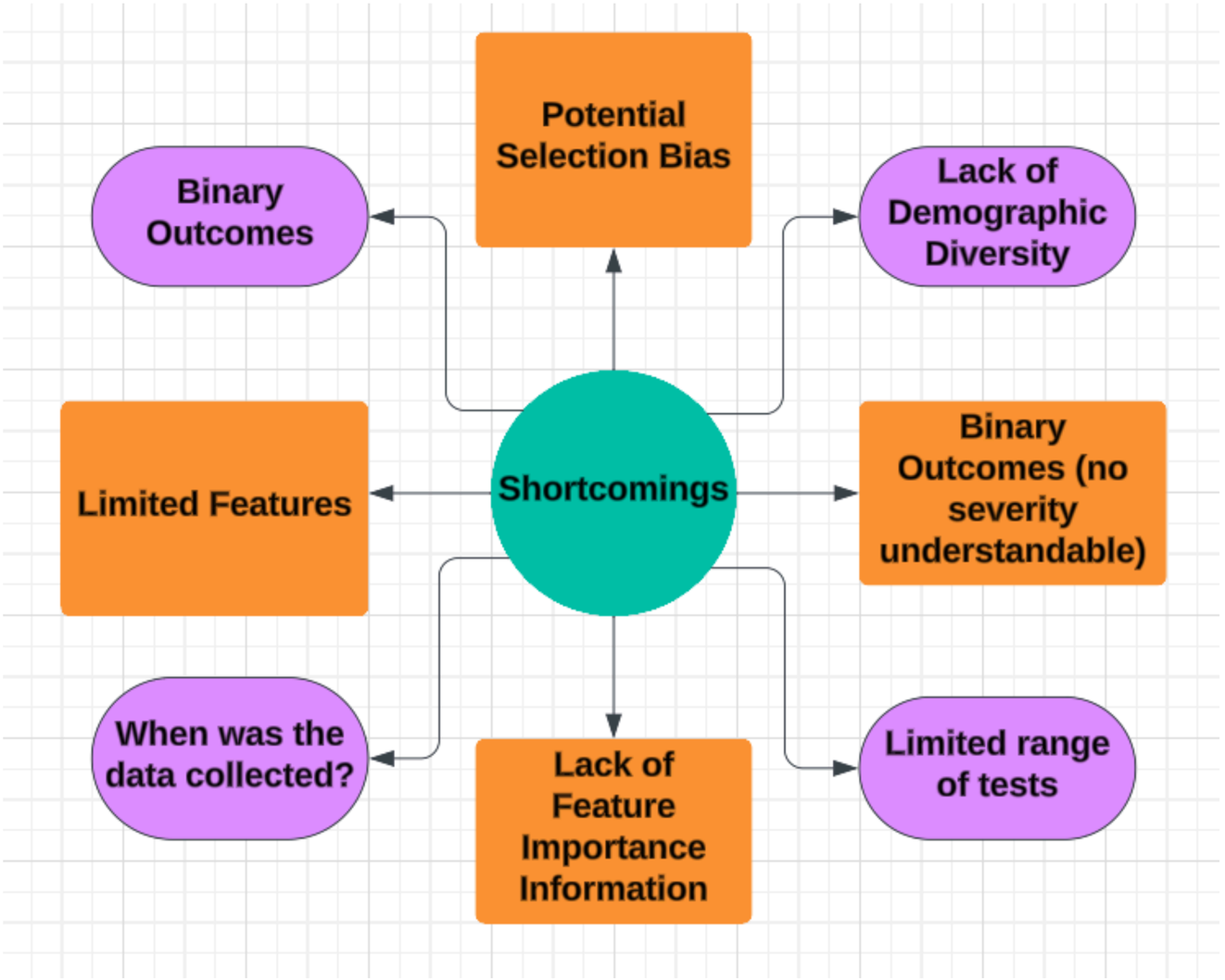
```
# classification report precision recall f1-score
report = classification_report(y_test, y_pred)
print(report)
```

```
              precision    recall  f1-score   support

           0       0.42      0.34      0.38        29
           1       0.80      0.84      0.82        88

    accuracy                           0.72       117
   macro avg       0.61      0.59      0.60       117
weighted avg       0.70      0.72      0.71       117
```

SHORTCOMINGS



THANK YOU

We are now open to questions.