



IT Salaries 4.0

Lelia Erscoi

Agenda



About the Data

Previous Models

Intermezzo

Decision Tree- Gini

Decision Tree- Entropy

Decision Tree- Best Features

Naive Bayes

Naive Bayes - Best Split

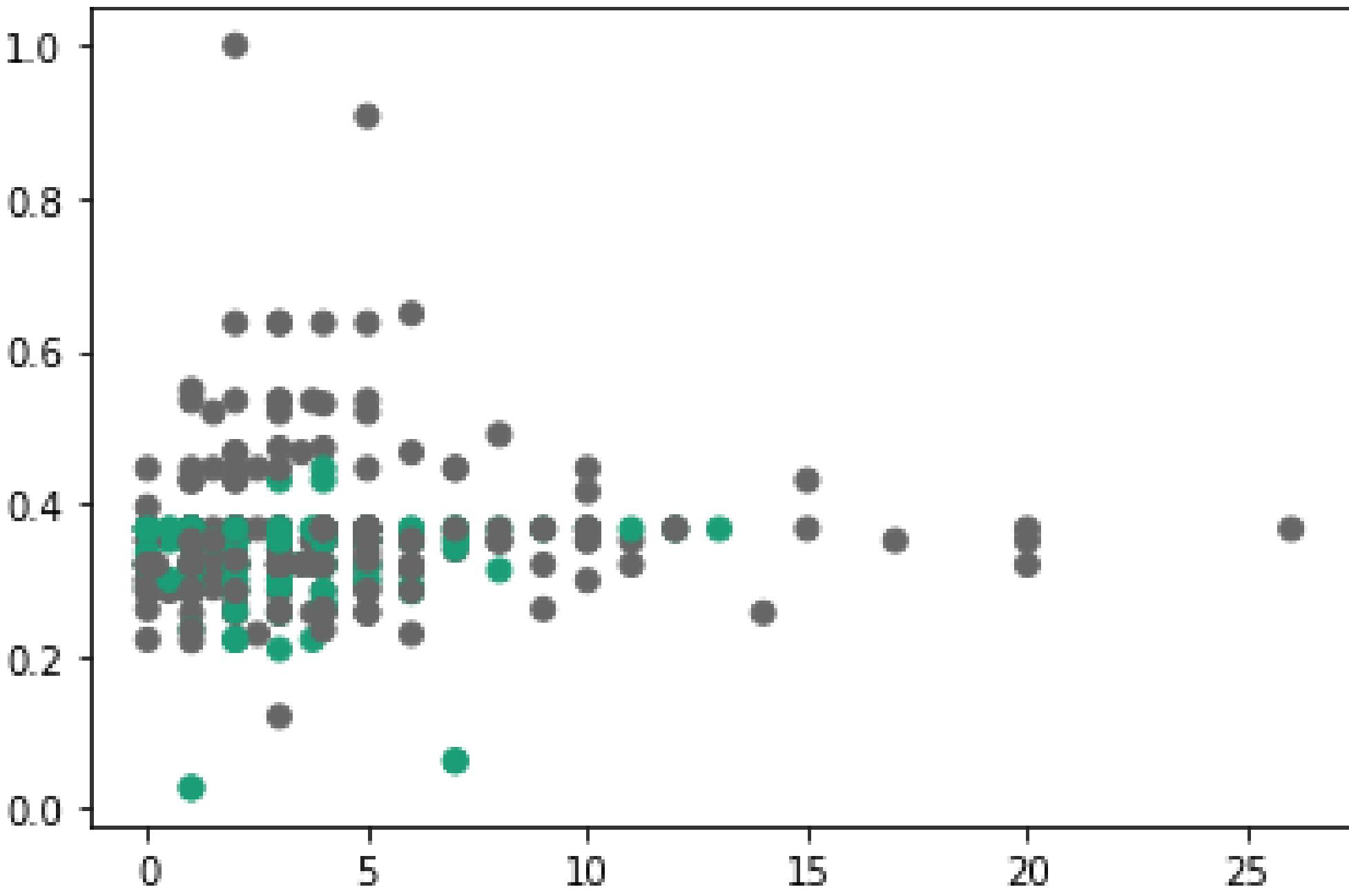
Findings

About the Data

Insights:

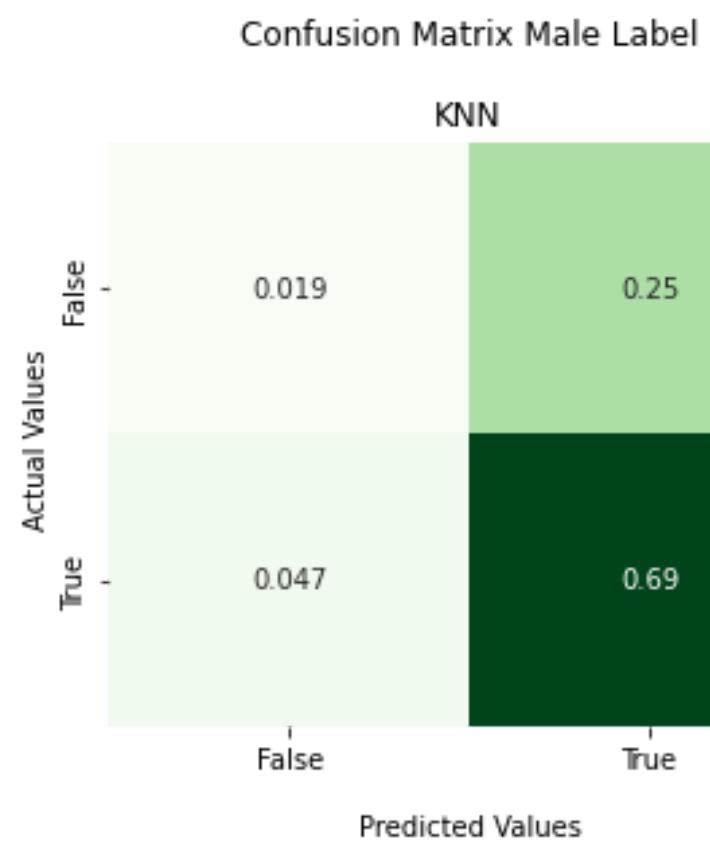
- 431 x 20 dataset
- Focus on the Gender dimension
- 141 "female" and 390 "male"

IT Salaries Data, Predictor = Male

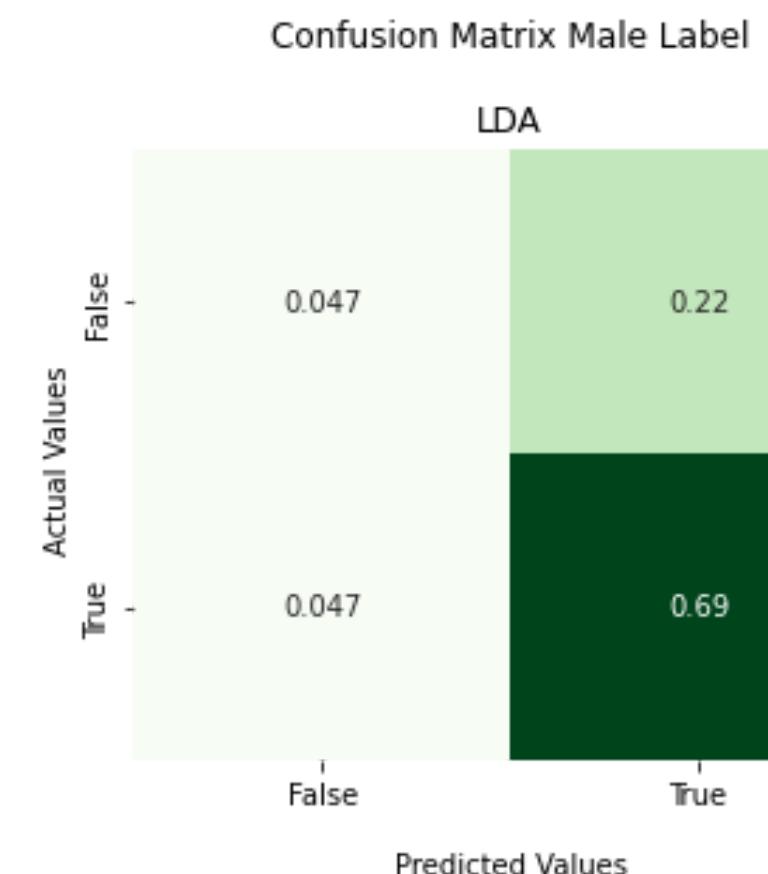


Previous Models

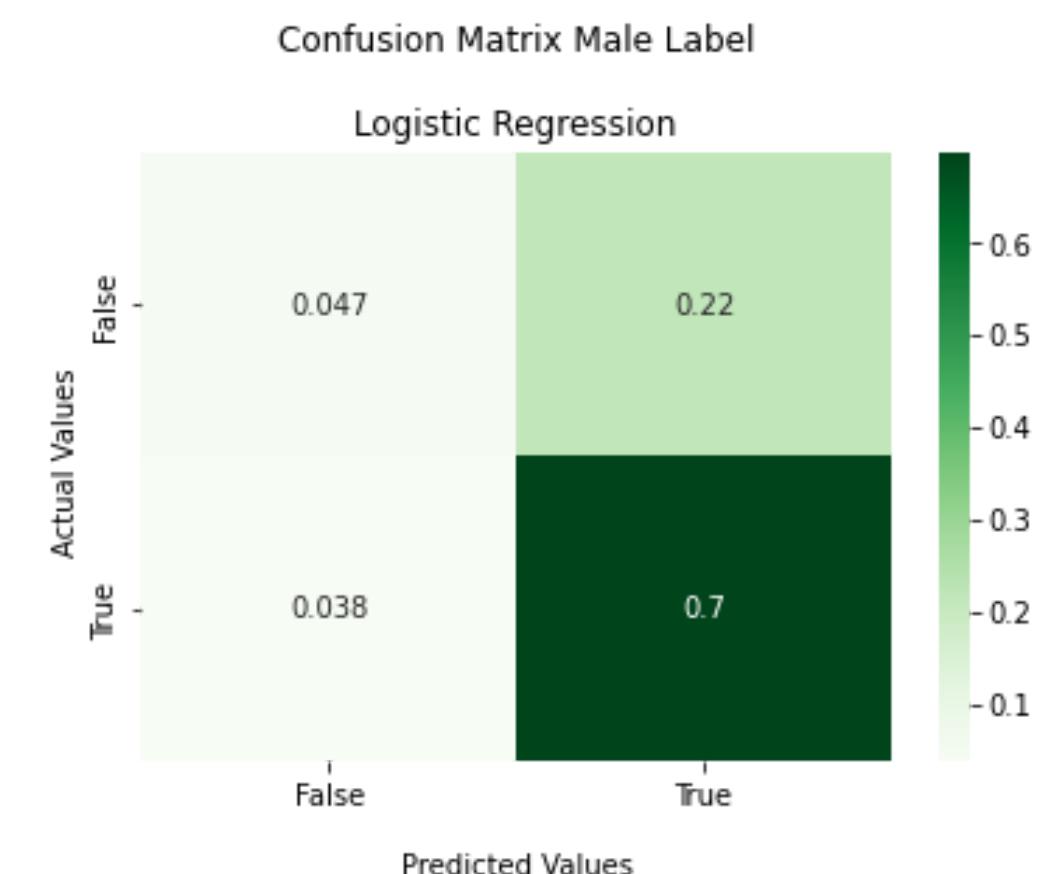
KNN



LDA



Logistic Regression

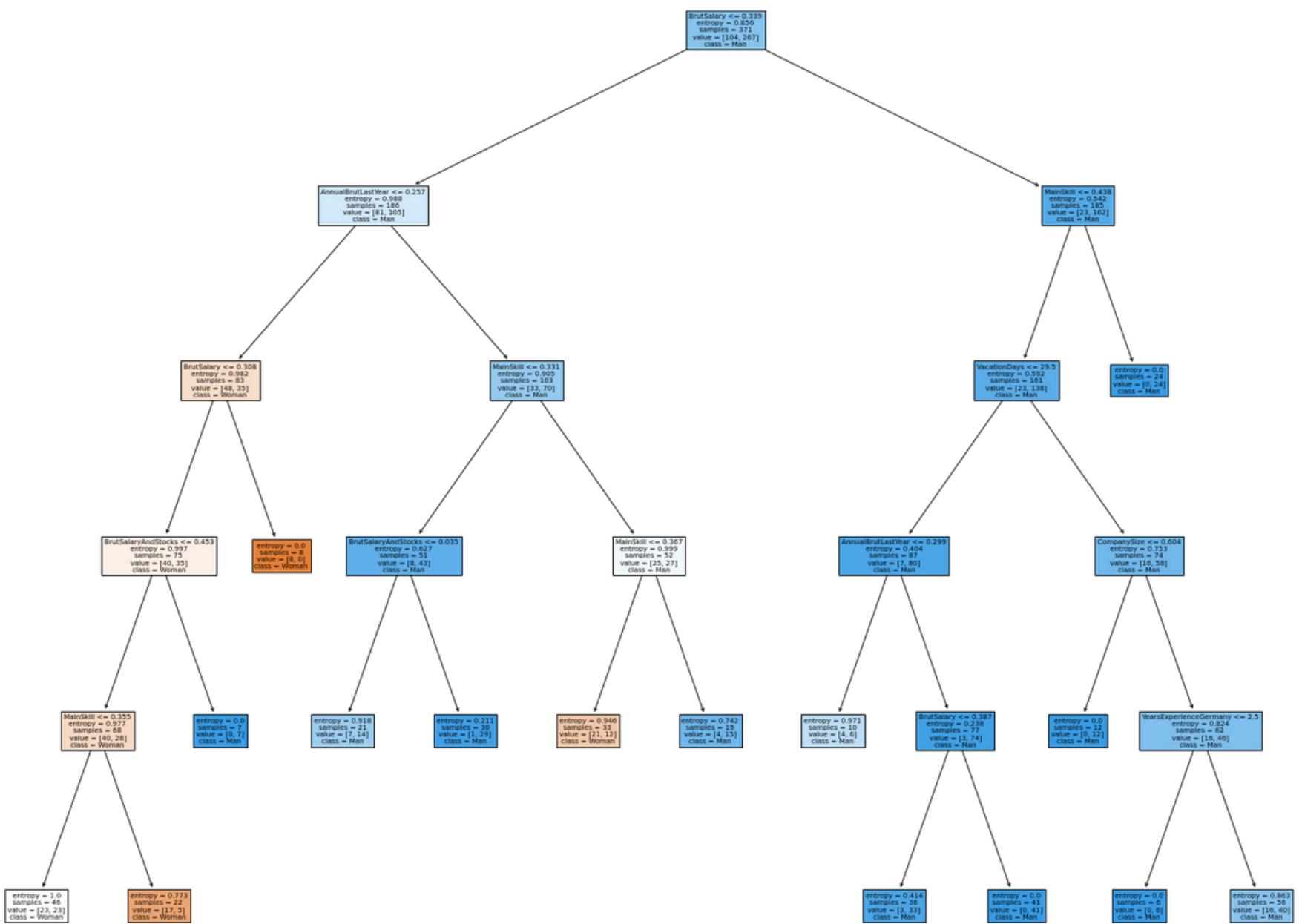


Accuracy: 71%
Recall: 94%

Accuracy: 74%
Recall: 94%

Accuracy: 75%
Recall: 95%

Intermezzo



How to determine the best split?

- measure node impurity

C0: 5
C1: 5

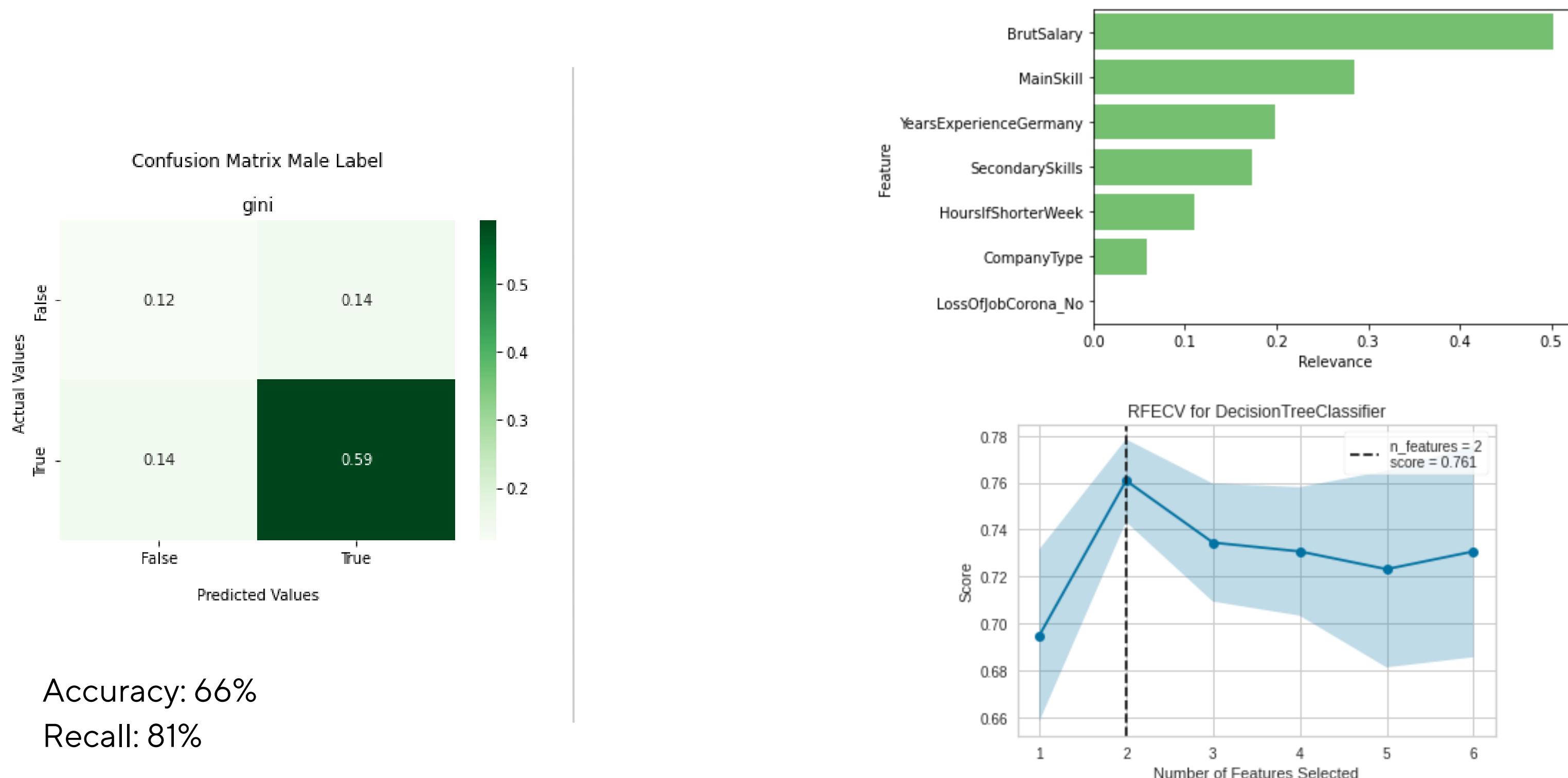
Non-homogeneous,
High degree of impurity

C0: 9
C1: 1

Homogeneous,
Low degree of impurity

Decision Tree- Gini

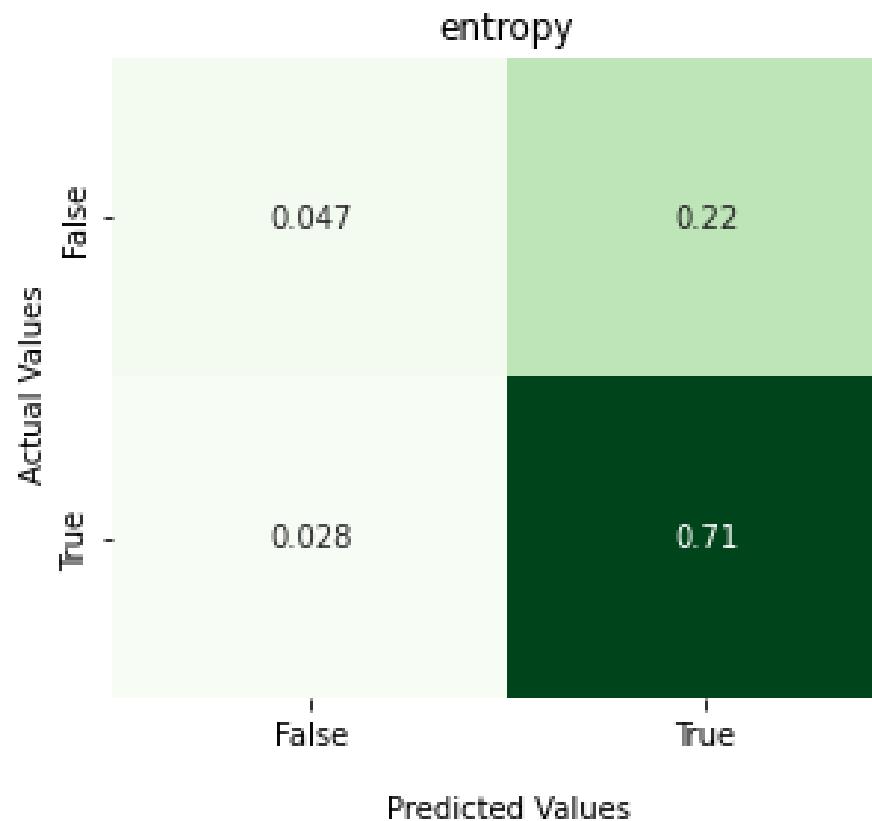
Predicting Male Label



Decision Tree- Entropy

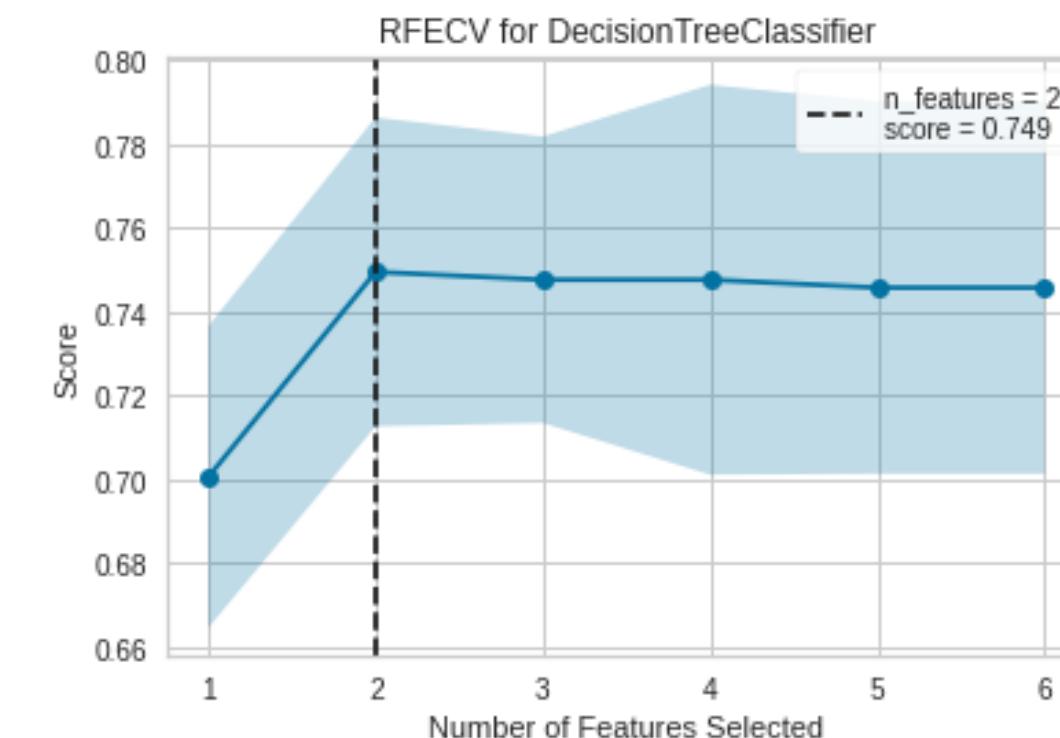
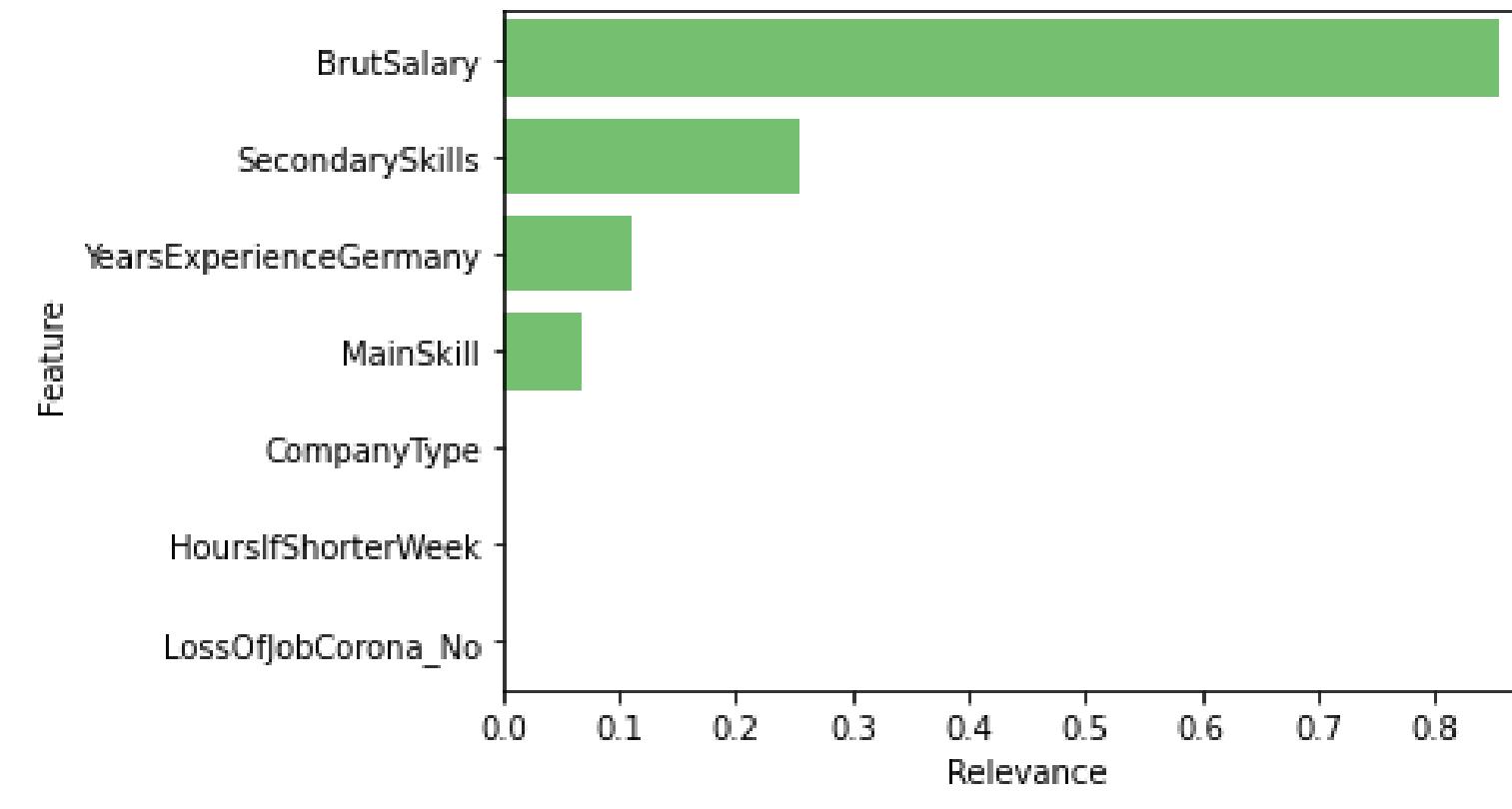
Predicting Male Label

Confusion Matrix Male Label



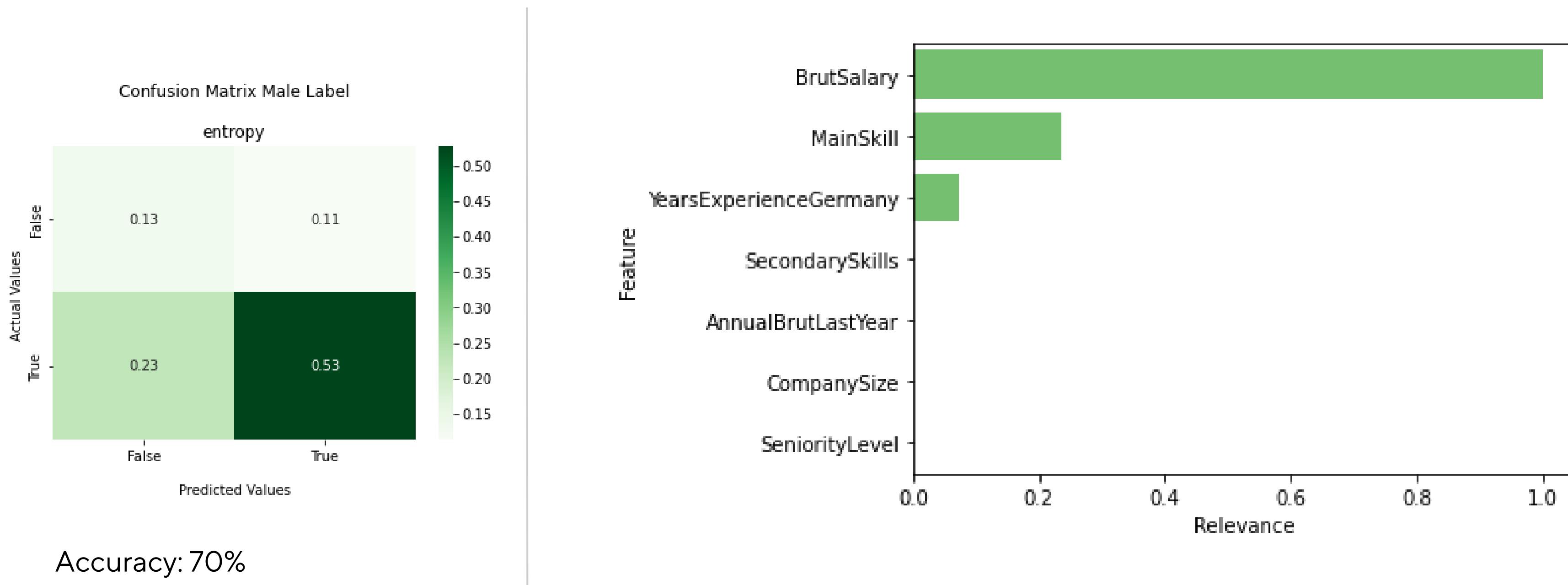
Accuracy: 70%

Recall: 96%



Decision Tree- Best Features Only

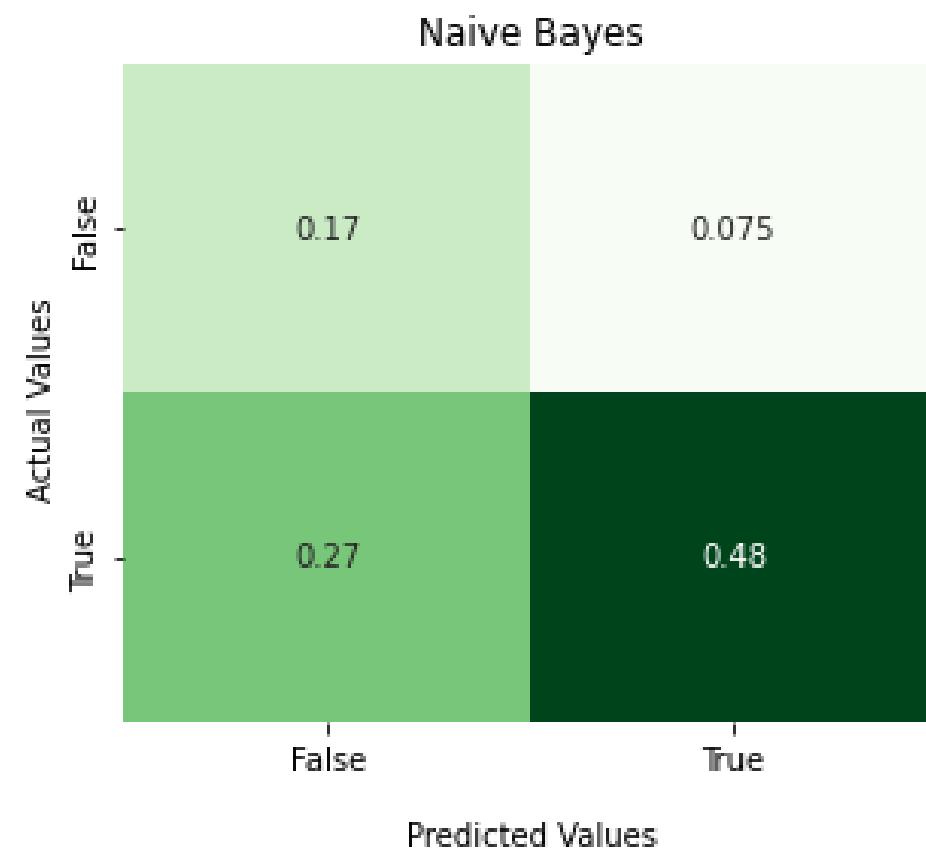
Predicting Male Label



Naive Bayes

Predicting Male Label

Confusion Matrix Male Label

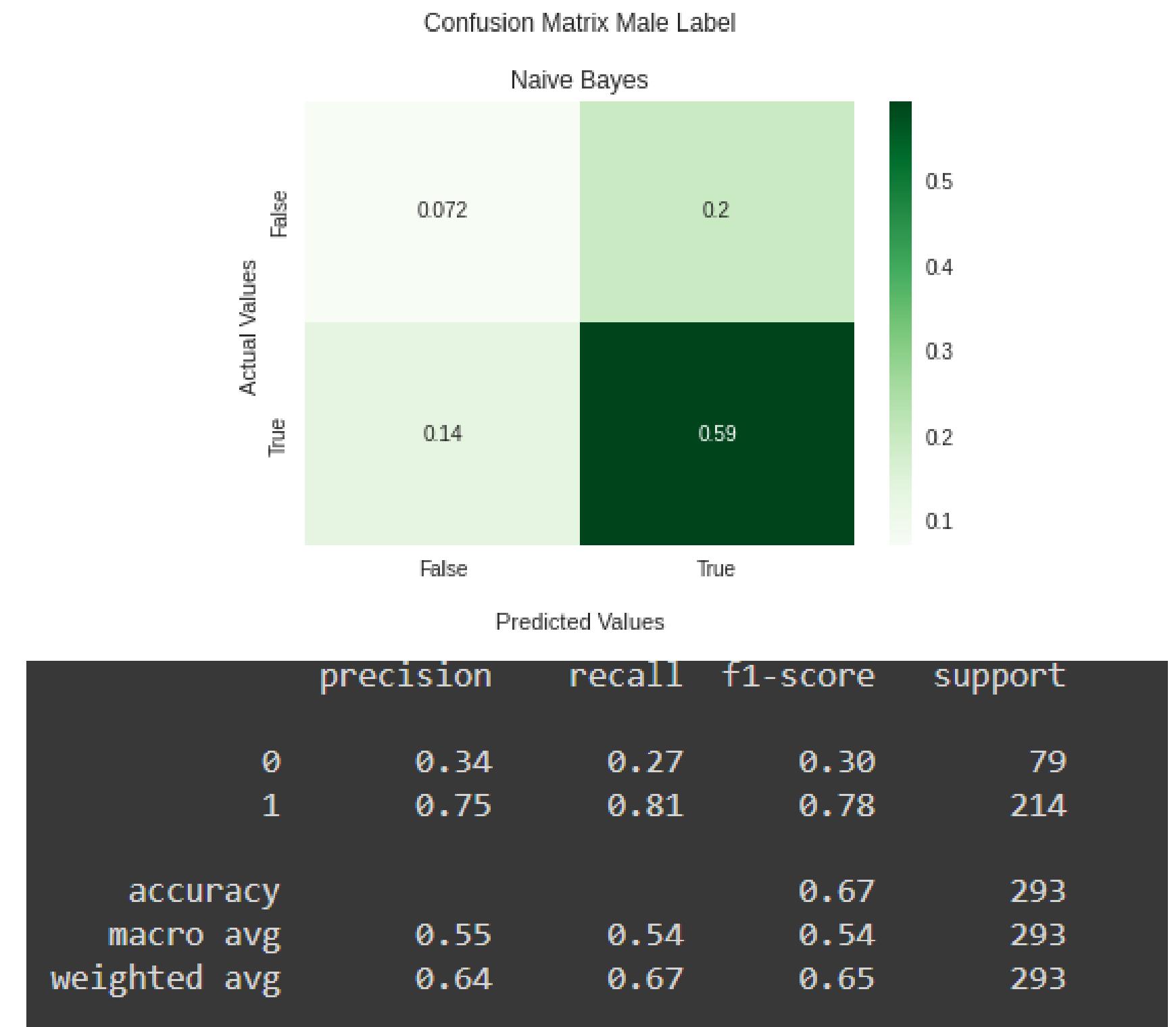
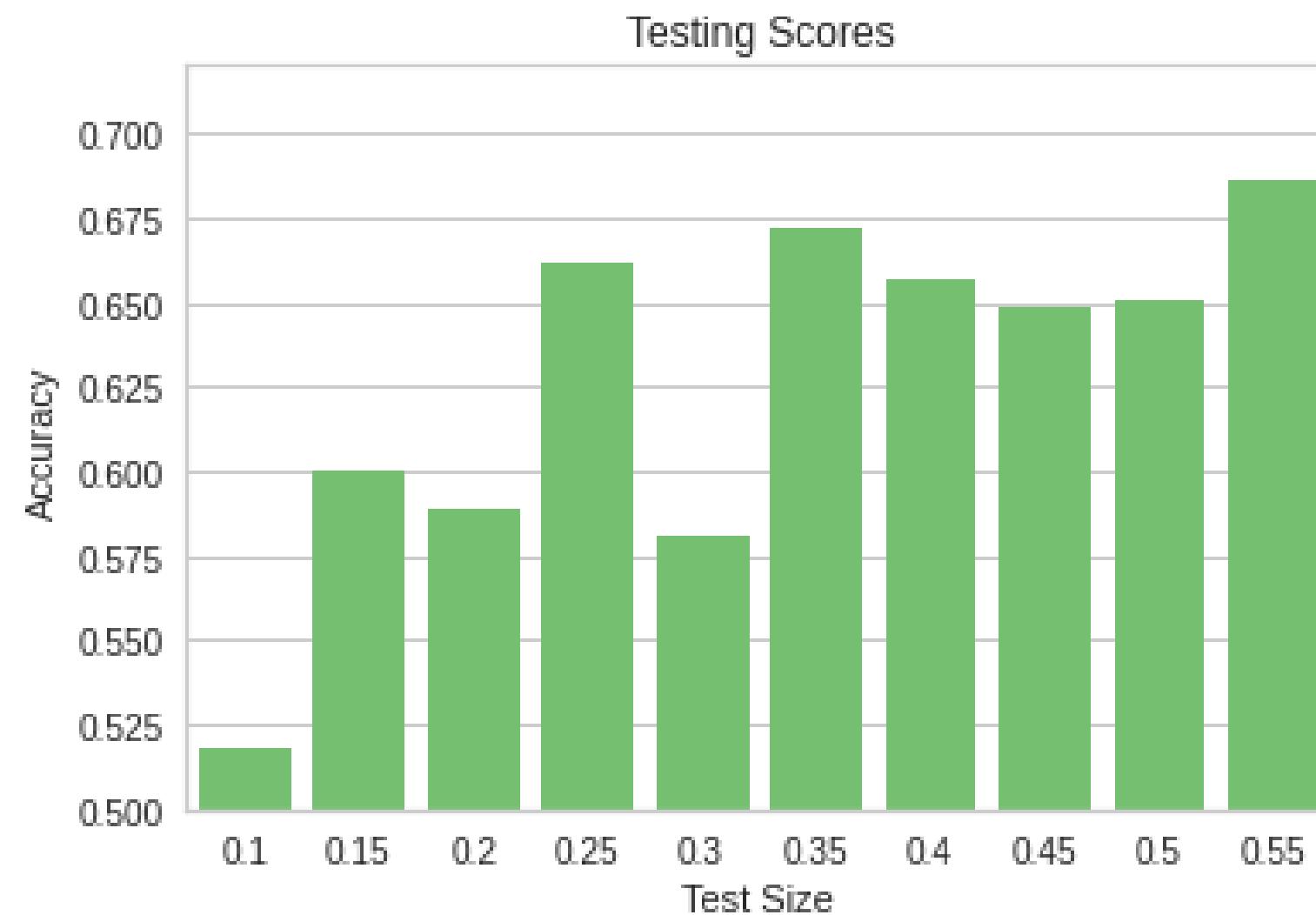


Avg accuracy : 0.6272086051842708

	precision	recall	f1-score	support
0	0.38	0.69	0.49	26
1	0.86	0.64	0.73	80
accuracy				106
macro avg	0.62	0.66	0.61	106
weighted avg	0.75	0.65	0.67	106

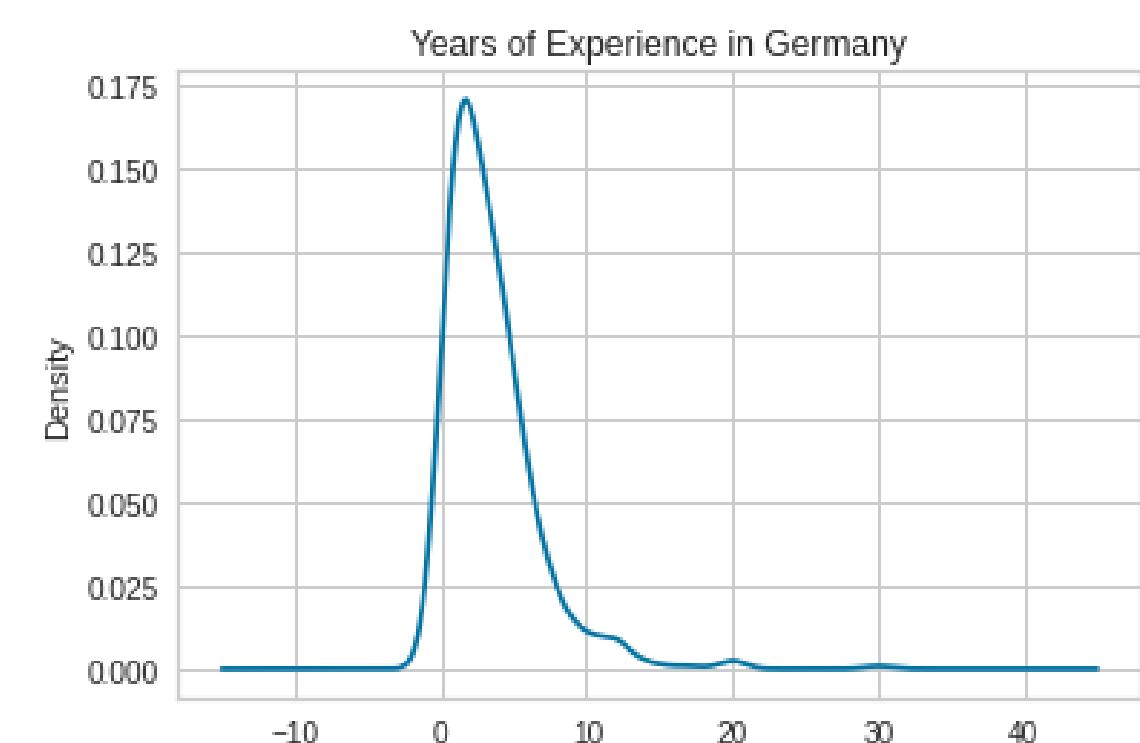
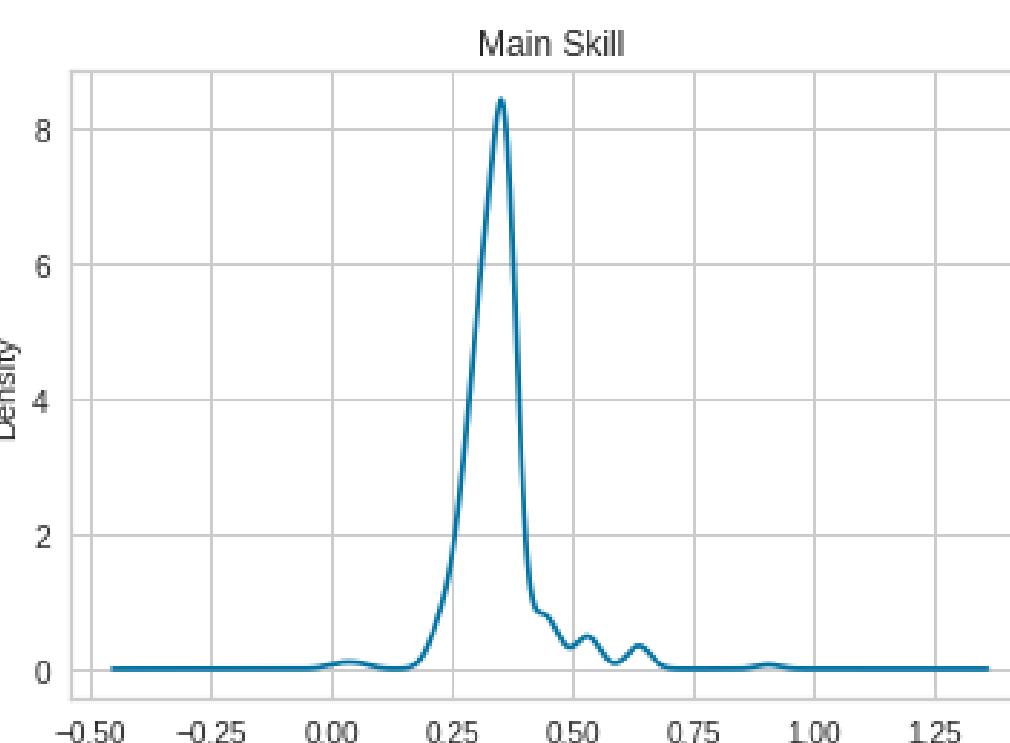
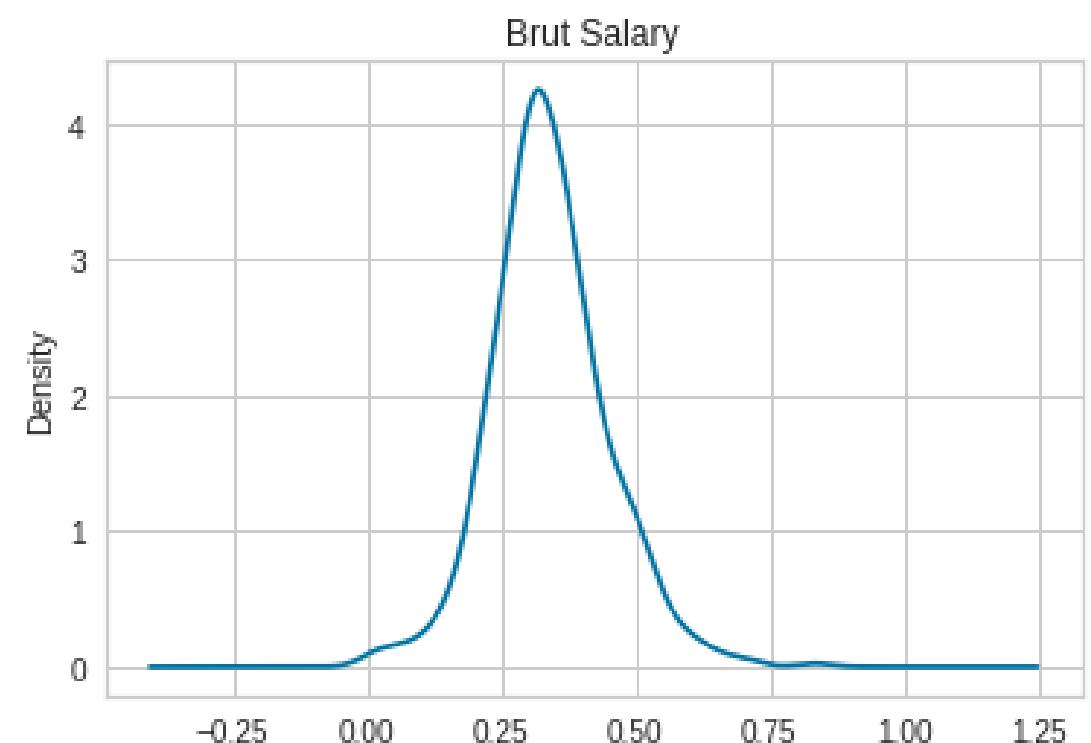
Naive Bayes

Best Split



Naive Bayes

Why Gaussian NB?



Findings

H0: Brut Salary has no difference between men and women.

H1: Brut salary is different between men and women.

Two-Sample T-Test; p-value = 1.61×10^{-11}

=> H0 Rejected

H0: Men and women equally lost their jobs because of COVID-19

H1: There is a difference in how men and women lost their jobs because of COVID-19.

Two-Sample T-Test; p-value = 0.8

=> H0 Failed to be rejected

	Feature	p-value
3	BrutSalary	0.0000
5	AnnualBrutLastYear	0.0000
1	MainSkill	0.0001
12	SeniorityLevel	0.0003
4	BrutSalaryAndStocks	0.0107
2	SecondarySkills	0.0114
0	YearsExperienceGermany	0.0269



Thank You!

