FairHire

Fairness-driven recruitment optimization model

Summary

Motivation

Project objectives

Exploratory analysis

Approach explanation

Results and conclusion

Motivation

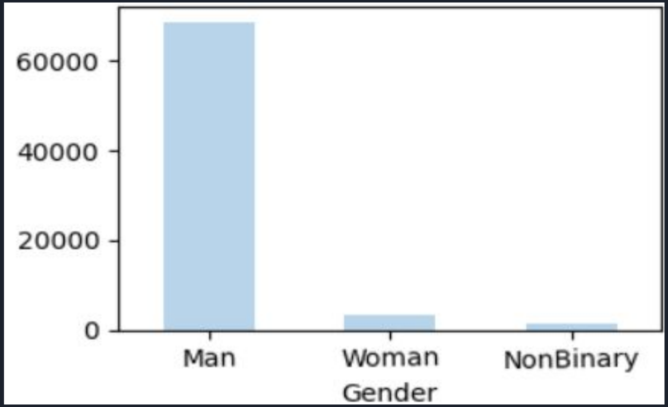
Project objectives

01 Provide a solution to facilitate the recruitment in the company

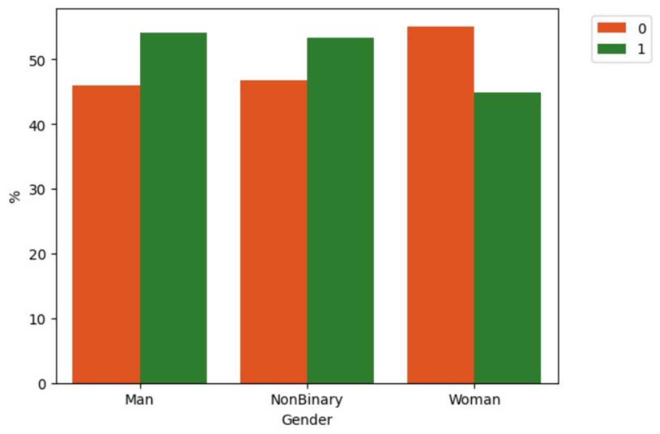
02 Ensure the solution does not discriminate against candidates

03 Ensure the selection contributes positively to the company’s growth and success

Exploratory Analysis



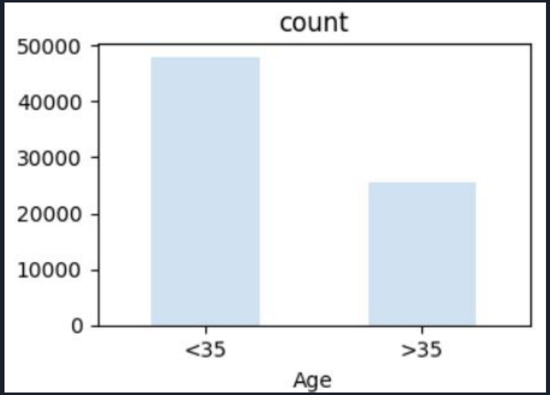
Number of Male applicants far exceeds the other genders.

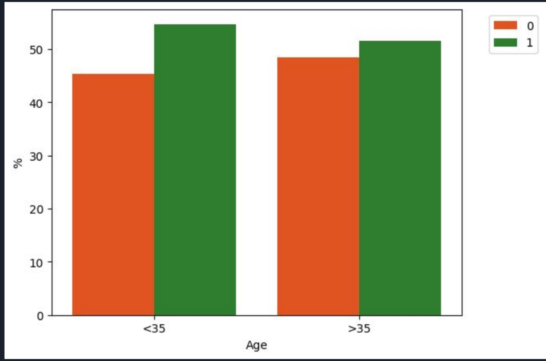


When we compare the % of hired job applicants we notice a 10 % difference between women and the rest of the population.

Potential risk of model bias

Exploratory Analysis





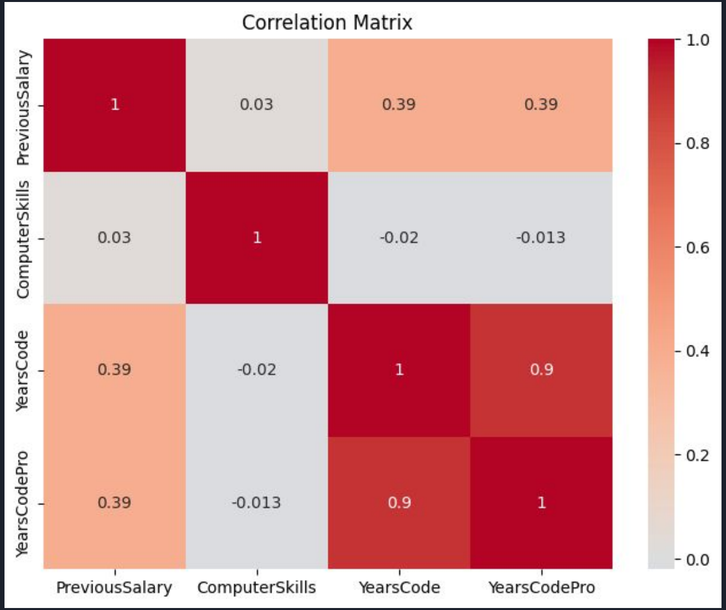
Age	Gender	Employed	proportion
<35	Man	1	55.2
		0	44.8
	NonBinary	1	56.0
		0	44.0
	Woman	0	53.7
		1	46.3
>35	Man	1	52.0
		0	48.0
	NonBinary	0	51.8
		1	48.2
	Woman	0	59.0
		1	41.0

Number of <35 applicants is double that of > 35

Potential risk of model bias

There is a small difference in the chance of employment between the two age groups (5%), but when we couple it with gender we notice a gap that disfavours women and non-binary.

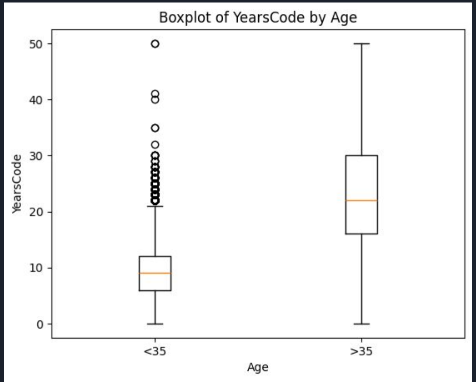
Exploratory Analysis

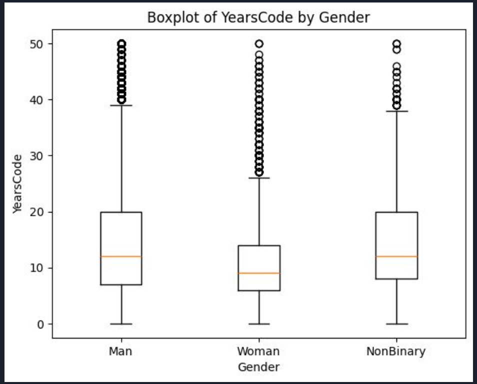


People with more total coding experience also tend to have more professional coding experience

Surprisingly, there is no correlation between the number of computer skills and experience (YearsCode), nor with previous salary.

Exploratory Analysis





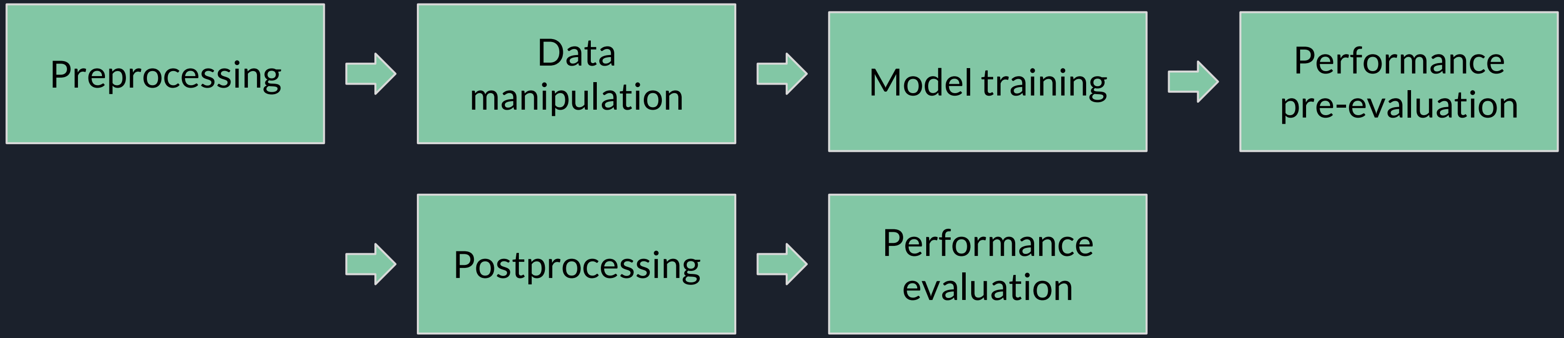
There is no relationship between Number of computer skills and age, nor gender .

As expected, the older age category has more coding experience.

Coding experience for women in this dataset is slightly lower and less variable.

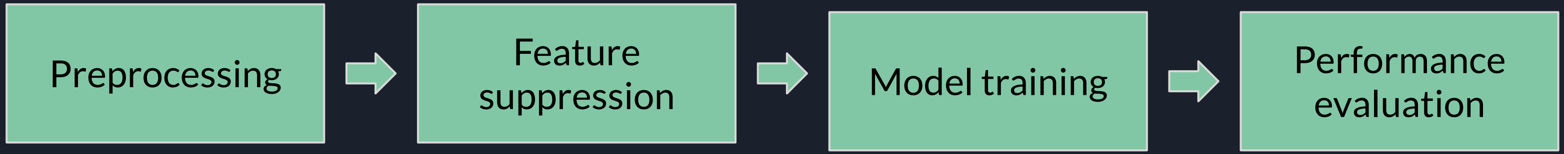
Approach explanation

- **General approach :**



Approach explanation

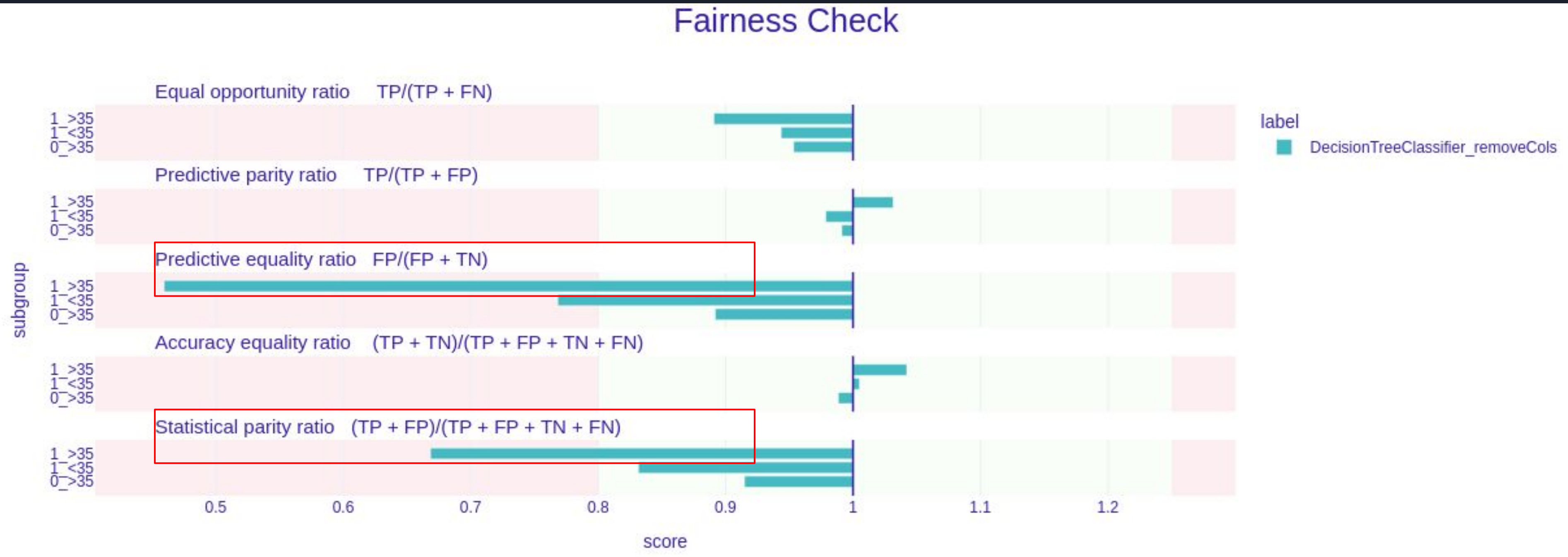
- **Approach 1: DecisionTree with data manipulation**



- **Dropped features : [Age, Gender]**

Results

	recall	precision	f1	accuracy	auc
DecisionTreeClassifier	0.800288	0.790104	0.795163	0.779391	0.860834



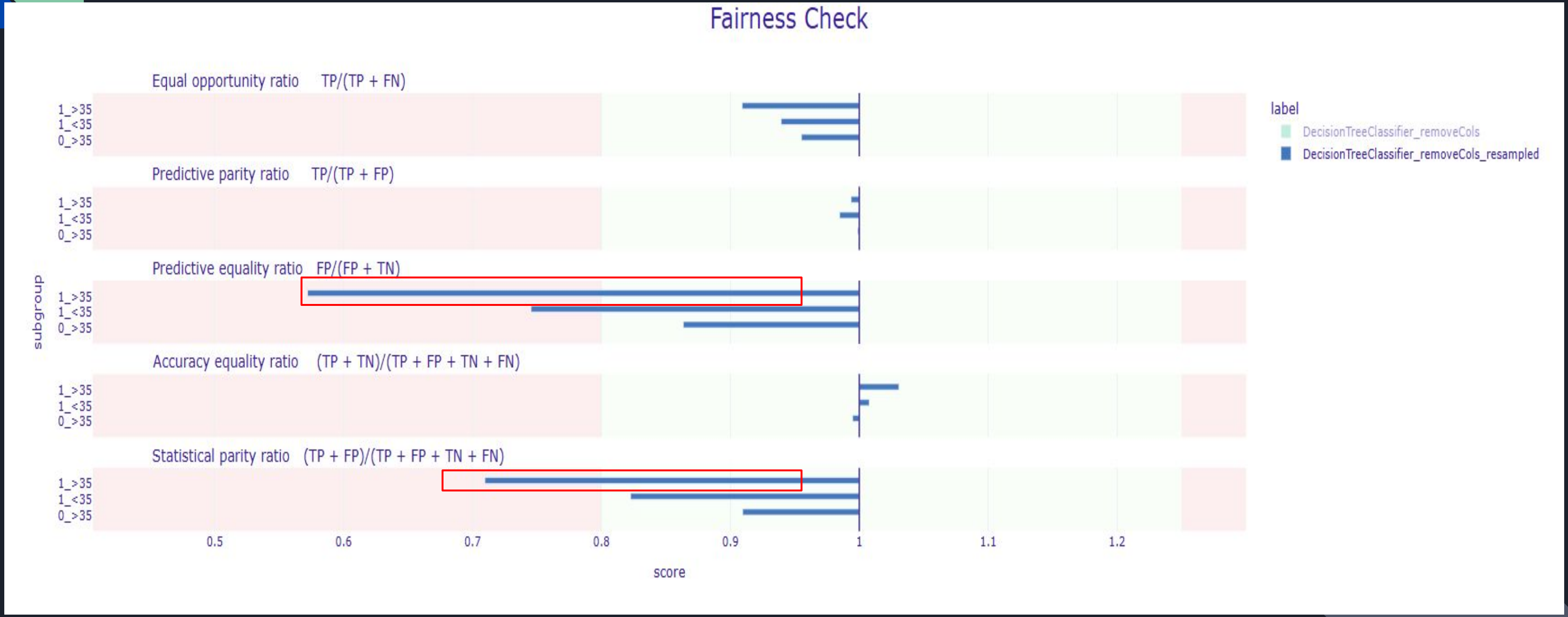
Approach explanation

- **Approach 2: DecisionTree with data manipulation and resampling**

Preprocessing → Feature suppression → Resampling → Model training → Performance evaluation

- **Dropped features : [Age , Gender]**

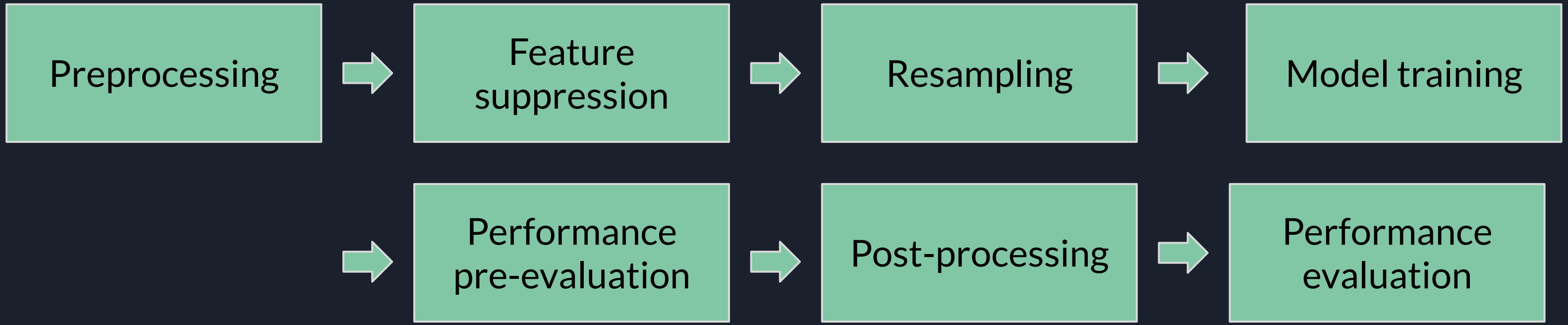
Results



	recall	precision	f1	accuracy	auc
DecisionTreeClassifier	0.810465	0.786002	0.798046	0.780525	0.860048

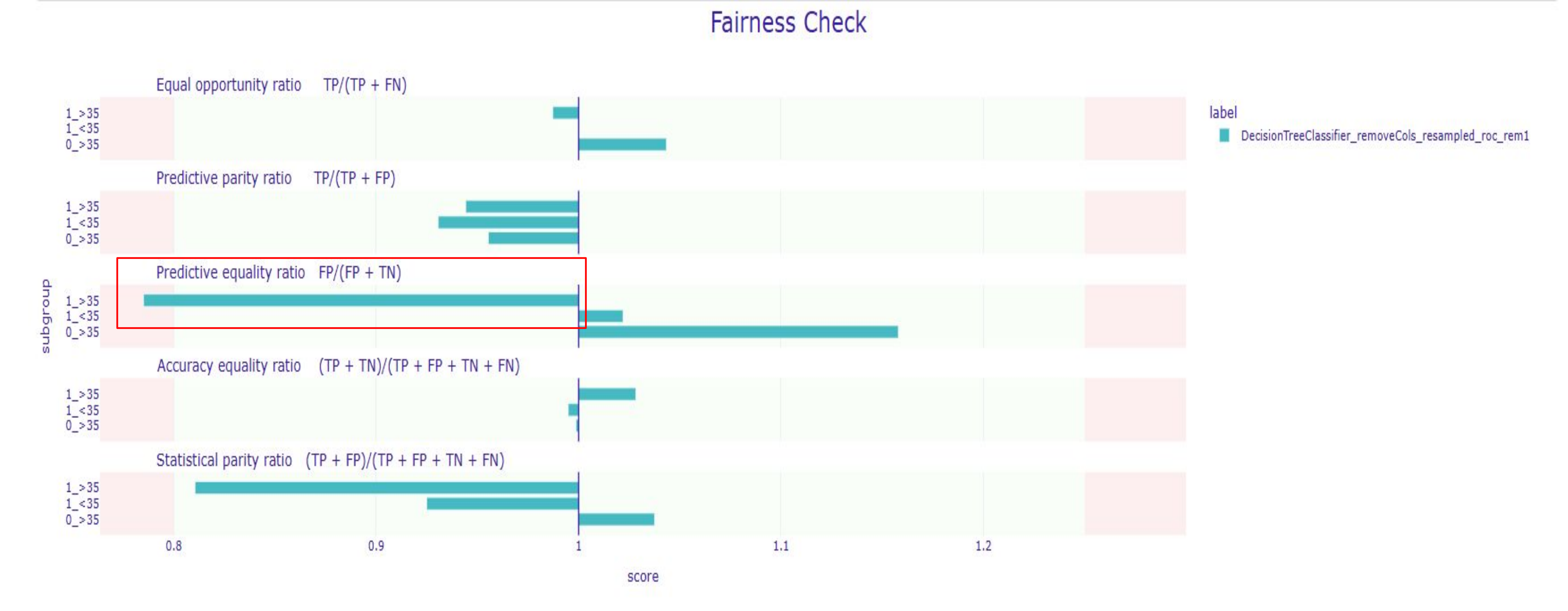
Approach explanation

- **Approach 3: DecisionTree with data manipulation , resampling and post-processing**



- **Dropped features : [Age, Gender, Years Code, EdLevel, PreviousSalary]**
- **Theta value : 0,033**

Results



	recall	precision	f1	accuracy	auc
DecisionTreeClassifier	0.7952	0.792378	0.793787	0.778937	0.864833

Result comparison

- Without losing much in efficacy !
  (measured by F1 score : [0.79, 0.8])

Thank you for your attention