HR Analytics

Stavroula Panaretou
Evaggelia Fostiropoulou

Big Blue Data Academy

HR Company

- HR Company wants to hire data scientists among people who successfully pass some courses.
- HR Company wants to know which of these candidates are really wants to work for the company after training or looking for a new employment?
- So we have to answer this question because it will help to reduce the cost and time as well as the quality of training or planning the courses and categorization of candidates.

STEPS

01

Data preprocessing

- Data Cleaning
- Nan values
- Type Conversions

02

Data ‘Insights’

- Imbalanced or not?
- Outliers

03

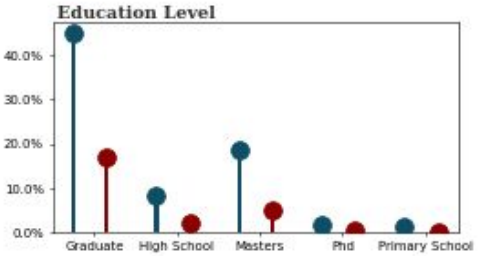
Machine Learning

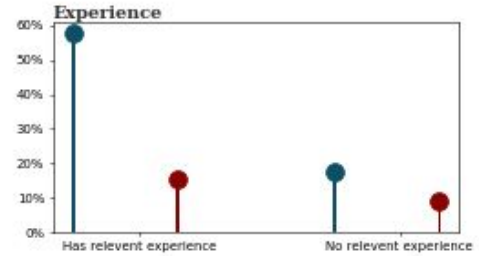
- Logistic Regression
- KNN
- Random Forest
- Ada Boost
- Gradient Boosting

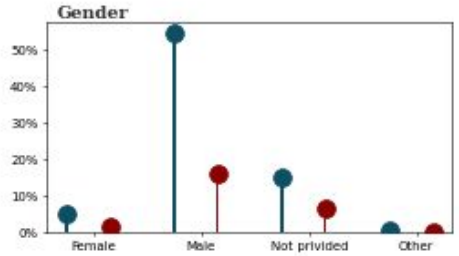
Let's see if our dataset is imbalanced...

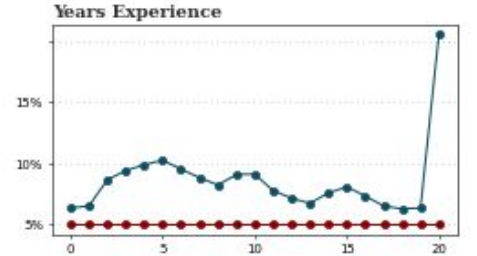
Obviously we have to deal with an imbalanced dataset

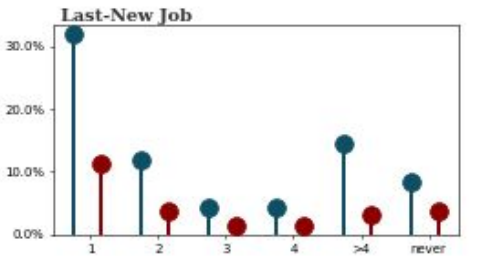
How much our Job & Non Job Seekers are imbalanced?

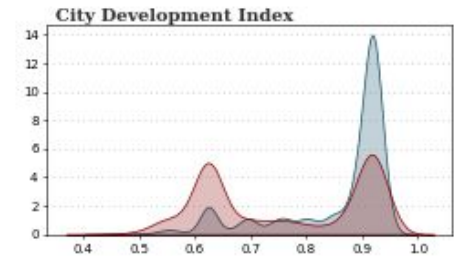


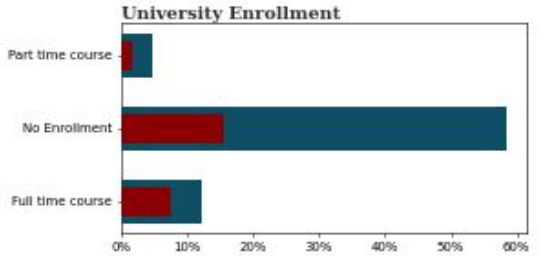


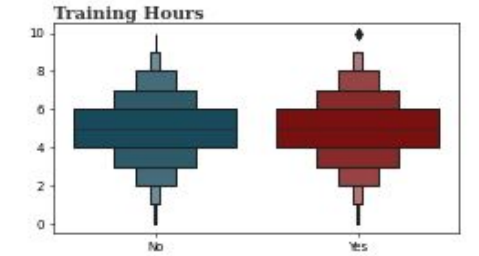


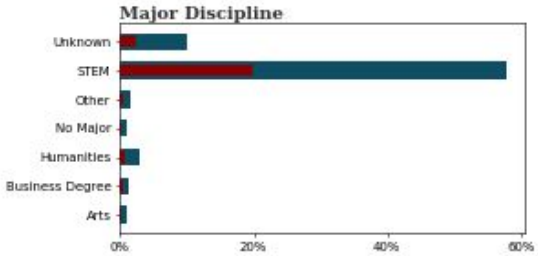












First attempt...Vanilla

Algorithm	Score	Precision	Recall	F1-Score
Logistic Regression	76,3%	0,55	0,23	0,32
Random Forest	77,66%	0,56	0,45	0,50
KNN	74,74%	0,49	0,35	0,41
ADABoost	78,16%	0,61	0,33	0,43
Gradient Boosting	80,1%	0,61	0,56	0,58

Let’s deep inside...

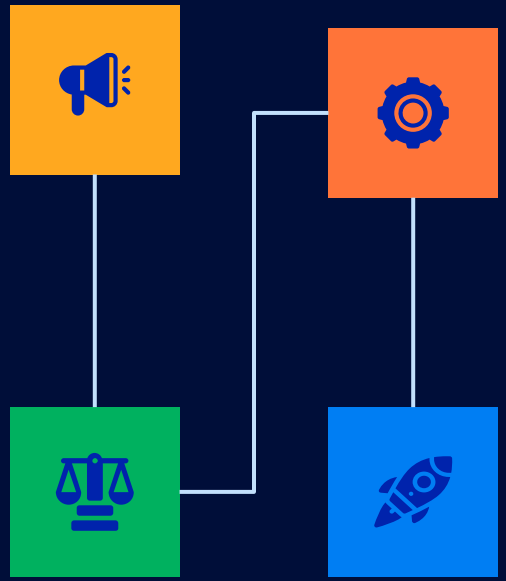
Techniques

Isolation Forest

- Find Outliers

Smote

- Oversample Technique



Stratified KFolds

- Trying to boost the training set

Randomized Search CV

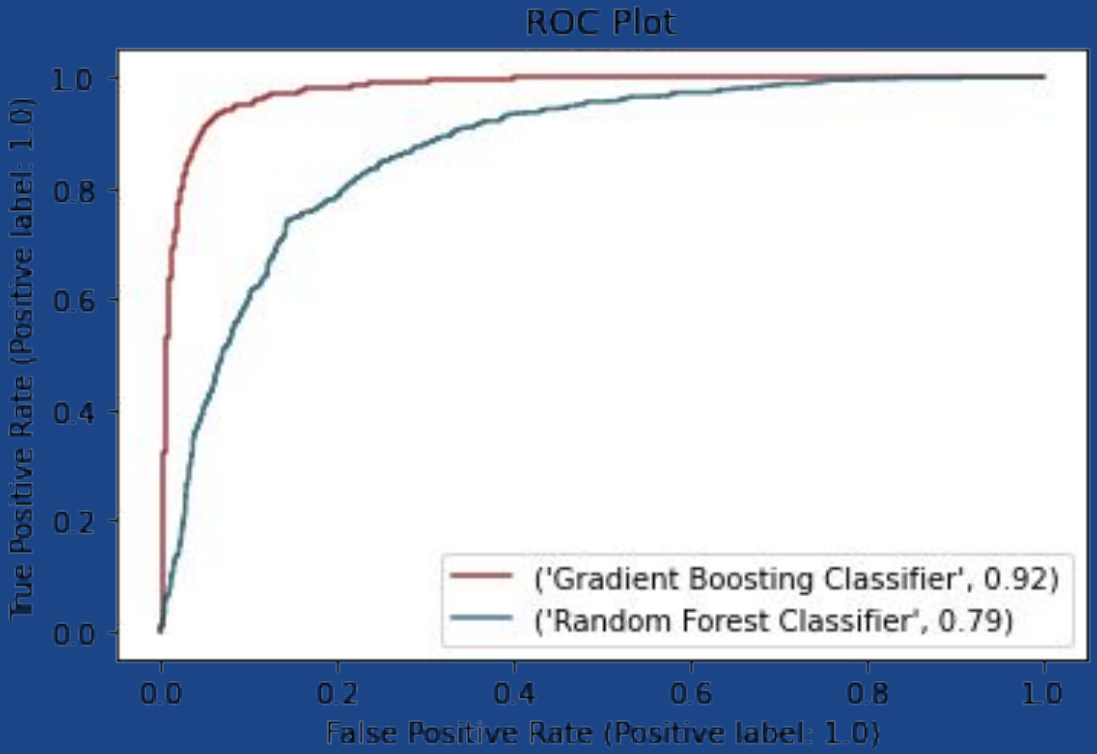
- Finding the best parameters

Final Metrics...

Algorithm	Score	Precision	Recall	F1-Score
Random Forest	81,21%	0,58	0,76	0,66
Gradient Boosting	94,18%	0,87	0,89	0,88

Check for Overfitting

Set	Accuracy Score
Train	0,97
Test	0,94


ROC Plot
True Positive Rate (Positive label: 1.0)
False Positive Rate (Positive label: 1.0)
0.0
0.2
0.4
0.6
0.8
1.0
('Gradient Boosting Classifier', 0.92)
('Random Forest Classifier', 0.79)

Conclusion

- Our prediction score(f1 score) for minority class is 88% and for the majority class is 96%.

- So we are confidence to say that our model gives to HR Company the ability to win in time and cost!!!