

HR Analytics

Stavroula Panaretou
Evaggelia Fostiropoulou

Problem: Probability of someone search for a new job

Step One: Data “Manipulation”

- Do we need all this info ?
- Dealing with missing values
- Data type conversions

Step Two: Data “Insights”

- Imbalance
- Outliers

Step Three: Machine Learning Techniques

- Random Forest ★★
- Logistic Regression ★
- K Nearest Neighbour ★
- AdaBoostClassifier ★
- GradientBoostingClassifier ★★★

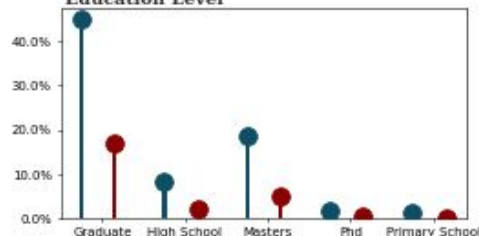
Let's see if our dataset is imbalanced...

Obviously we have to deal with an imbalanced dataset

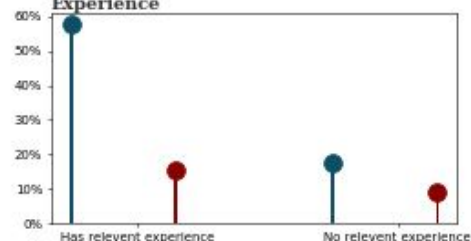


How much our **Job** & **Non Job Seekers** are imbalanced?

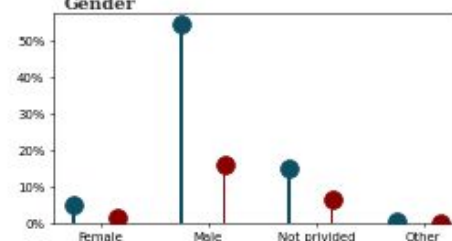
Education Level



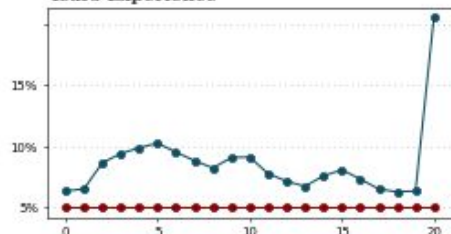
Experience



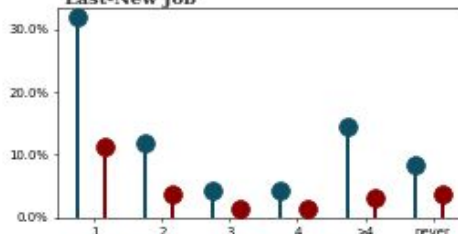
Gender



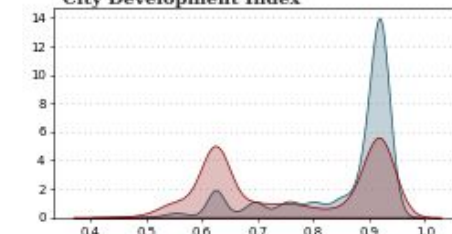
Years Experience



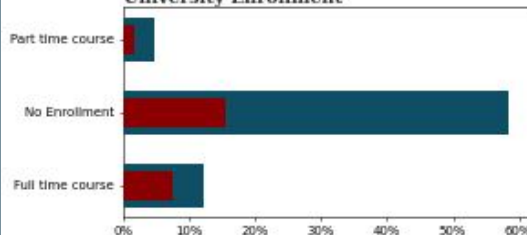
Last-New Job



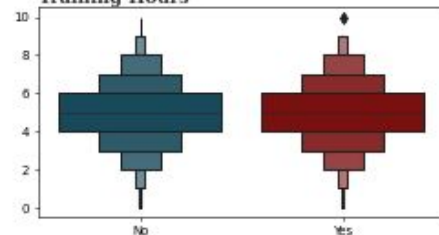
City Development Index



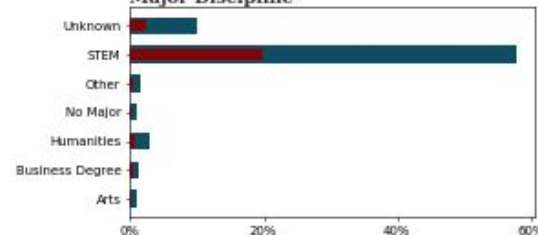
University Enrollment



Training Hours



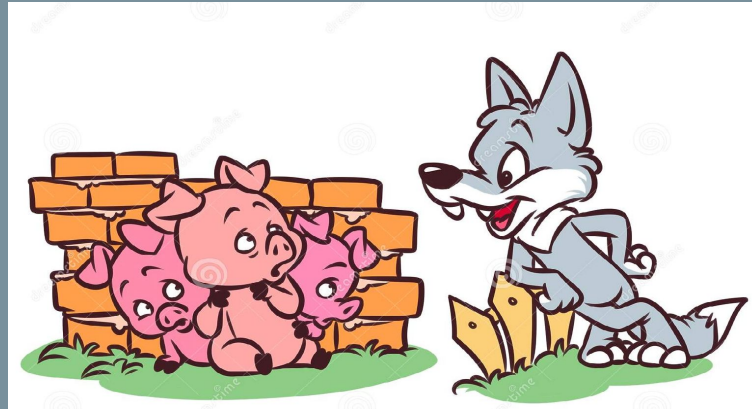
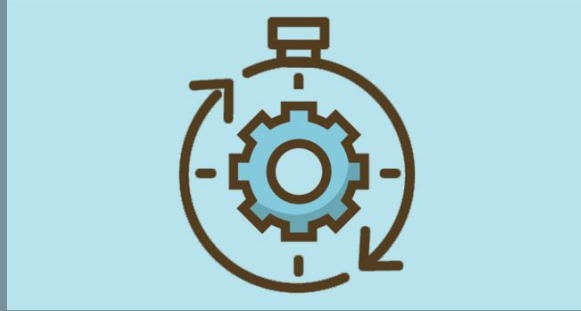
Major Discipline



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

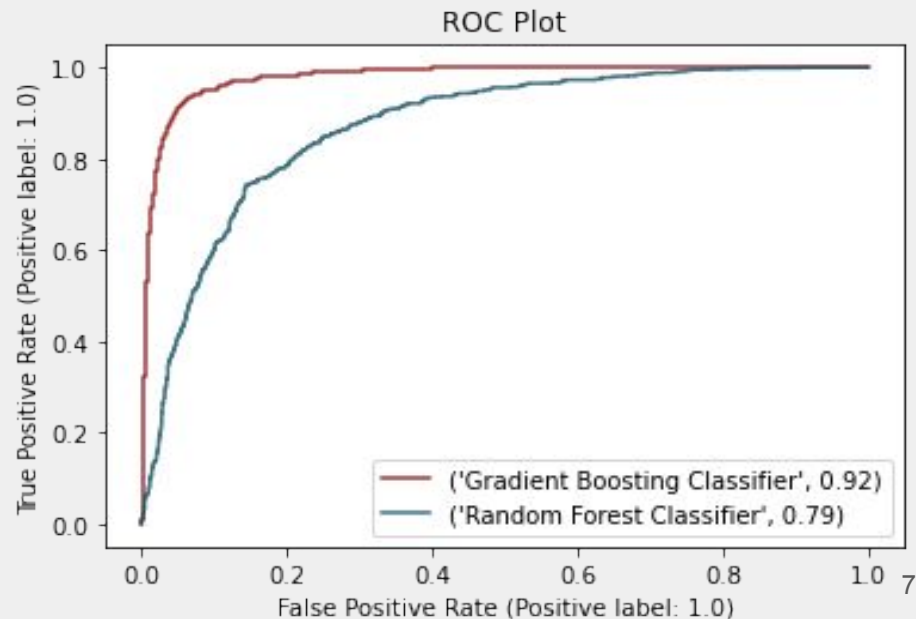
Serious Techniques



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 for Gradient Boosting

Set	Accuracy Score
Train	0,97
Test	0,94



Conclusion

- Our prediction score(f1 score) for minority class is 88% and for the majority class is 96%.
- So we are confident to say that our model is successfully operating in spite of the small dataset volume.
- Work in progress : Applying new techniques on our model as well as testing new parameters.