HR Analytics:
Job Change of Data Scientists

By: Banna AL-Otaibi

- In this project we will build a model based on our dataset “ Candidates for the position of data scientist“ to predict the probability of a candidate to look for a new job after training or not ,as well as interpreting affected factors on employee decision.

helps to reduce the cost and time as well as the quality of training or planning the courses and categorization of candidates.

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_discipline	experience	company_si
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	STEM	>20	Na
1	29725	city_40	0.776	Male	No relevent experience	no_enrollment	Graduate	STEM	15	50-
2	11561	city_21	0.624	NaN	No relevent experience	Full time course	Graduate	STEM	5	Na
3	33241	city_115	0.789	NaN	No relevent experience	NaN	Graduate	Business Degree	<1	Na
4	666	city_162	0.767	Male	Has relevent experience	no_enrollment	Masters	STEM	>20	50-

- you can find it at Kaggle in this link:
  HR Analytics: Job Change of Data Scientists | Kaggle
- This dataset contains :
  19,158 Candidates,and 14 features
- The dataset is imbalanced.
- Most features are categorical (Nominal, Ordinal, Binary), some with high cardinality.

Keep Learning

- Handle Missing
- Treat imbalanced data.
- Feature Importance Methods

Feature Selection

Train & Test

Data split train & validation & test

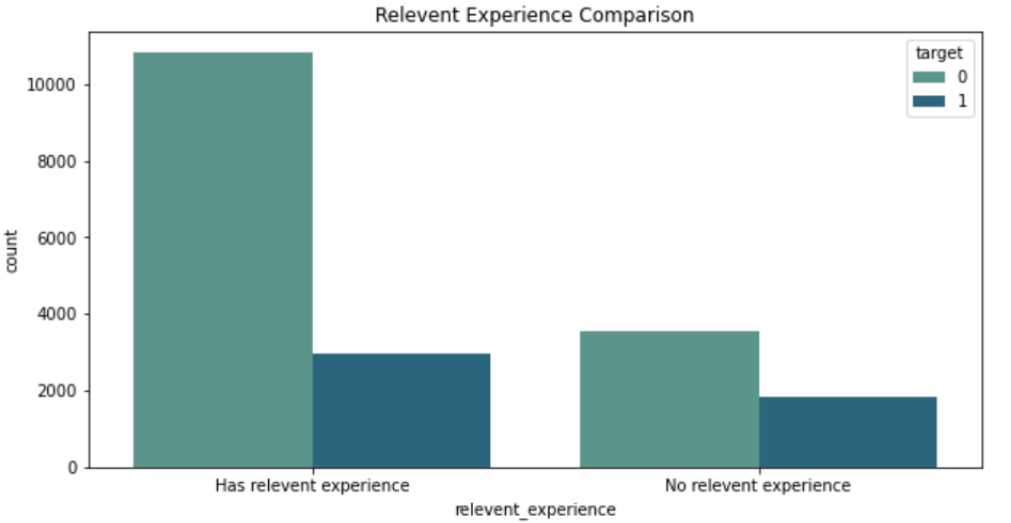
- Cross Validation
- Different Models

Model Tuning

Evaluate

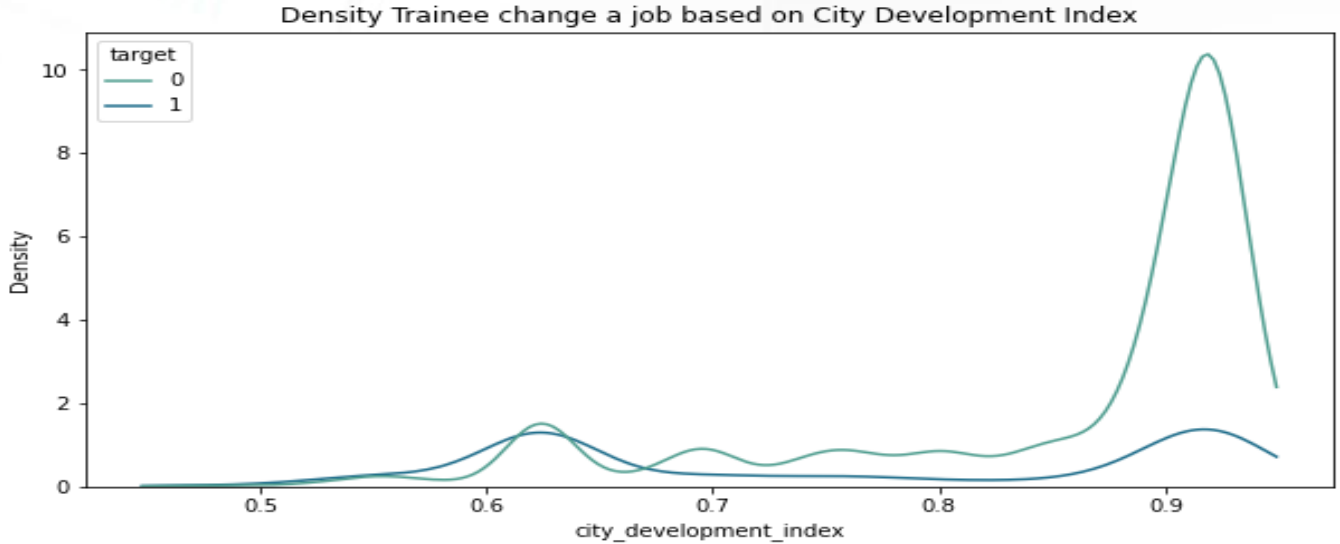
- Precision
- Recall
- F1

Which type of Candidates are the most change their job (based on relevant experience)?



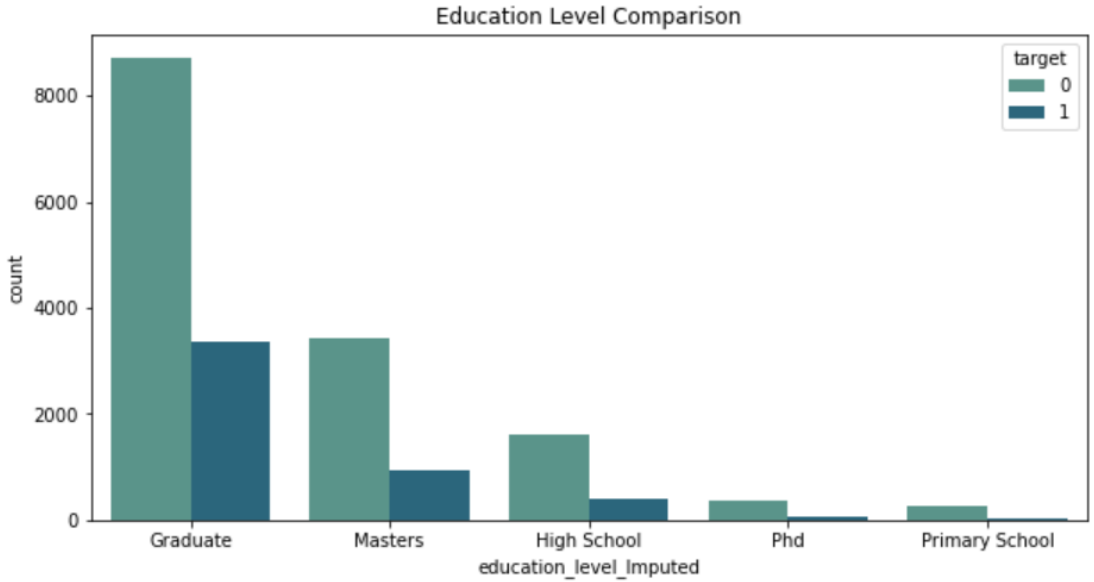
- Candidates with relevant experience is the ones who change their job.

Which type of Candidates are the most Change their job (based on city development index)



City with higher development index is less likely have trainee who wants to change their job.

Which type of Candidates are the most Change their job (based on education level)?



Graduate are the most ones who change their job.

Model

Testing Results:

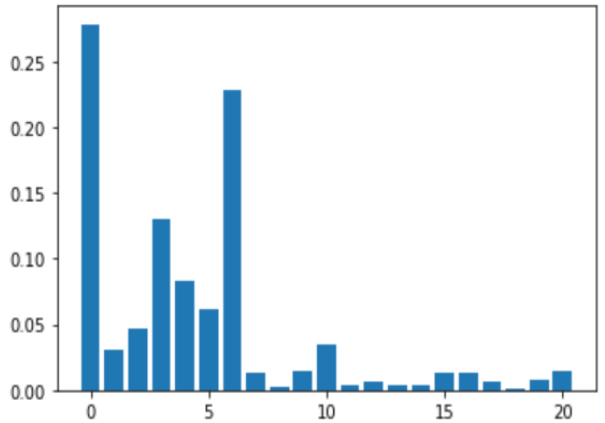
	precision	recall	f1-score	support
0	0.81	0.80	0.81	2877
1	0.80	0.81	0.81	2876
accuracy			0.81	5753
macro avg	0.81	0.81	0.81	5753
weighted avg	0.81	0.81	0.81	5753

RandomForestClassifier

Testing Results:

	precision	recall	f1-score	support
0	0.73	0.74	0.73	2877
1	0.73	0.73	0.73	2876
accuracy			0.73	5753
macro avg	0.73	0.73	0.73	5753
weighted avg	0.73	0.73	0.73	5753

Feature importance



Model

Model	precision	recall	f1-score
Random Forest Classifier*	0.80	0.81	0.81
Decision Tree Classifier	0.77	0.77	0.77
Random Forest Classifier	0.73	0.73	0.73
Support Vector Classification	0.71	0.58	0.64
Logistic Regression	0.70	0.59	0.64
KNeighborsClassifier	0.66	0.66	0.66

CONCLUSION

- By performing different ML models, we aim to get a better result
- From the previous results we can conclude that Decision Tree Classifier is the best model.

Thank you
I hope you enjoyed the presentation

By: Banna AL-Otaibi