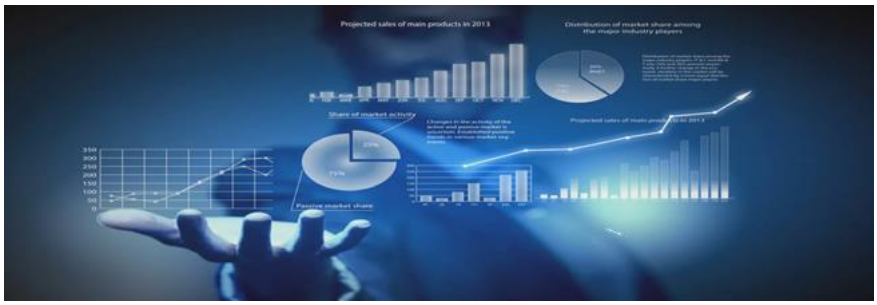
Project Proposal

HR Analytics: Job Change of Data Scientists



Content

A company that is active in Big Data and Data Science wants to hire data scientists among people who successfully pass some courses that conduct by the company.

Many people signup for their training. Company wants to know which of these candidates are really wants to work for the company after training or looking for a new employment because it helps to reduce the cost and time as well as the quality of training or planning the courses and categorization of candidates.

The study aims to find out what are affective factors on employee decision, predict the probability of a candidate work for the company.

I am personally interested in this dataset, I am looking forward to interpreting affected factors on employee decision.

Dataset

This dataset can be found at Kaggle in this link: *HR Analytics: Job Change of Data Scientists | Kaggle* (2020-12-07).

This dataset includes 19158 candidates with attributes such as enrollee id, gender, relevant experience, training hours etc.

Column name	Column description
`enrollee_id`	Unique ID for enrollee
`city`	City code
`city_development_index`	Developement index of the city (scaled)
`gender`	Gender of enrolee
`relevent_experience`	Relevant experience of enrolee
`enrolled_university`	Type of University course enrolled if any
`education_level`	Education level of enrolee

Column name	Column description
major_discipline	Education major discipline of enrolee
experience	Enrolee total experience in years
company_size	No of employees in current employer's company
company_type	Type of current employer
last_new_job	Difference in years between previous job and current job
training_hours	training hours completed
target	0 – Not looking for job change. 1 – Looking for a job change

The dataset is available as the .csv file. a sample of data is shown in the following table:

	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-

Model:

- Predict the probability of a candidate to look for a new job.

Tools:

There are tools that will be used to achieve the goal of this study, such as: numpy, pandas, matplotlib and seaborn for discovering and showing the data and train a model. The work will be done through Jupyter notebook using python.

TO DO:

- Explore the data and come up with EDA phases then use a model to fit the data.
- **NOTE:** the used features may be increased or changed and the model as well.