JanataHack: HR Analytics

A training institute which conducts training for analytics/ data science wants to expand their business to manpower recruitment (data science only) as well.

Company gets large number of signups for their trainings. Now, company wants to connect these enrollees with their clients who are looking to hire employees working in the same domain. Before that, it is important to know which of these candidates are really looking for a new employment. They have student information related to demographics, education, experience and features related to training as well.

To understand the factors that lead a person to look for a job change, **the agency wants you to design a model that uses the current credentials/demographics/experience to predict the probability of an enrollee to look for a new job.**

**Data Dictionary**

|  |  |
| --- | --- |
| **Variable** | **Description** |
| enrollee\_id | Unique ID for enrollee |
| city | City code |
| city\_development\_index | Developement index of the city (scaled) |
| gender | Gender |
| relevent\_experience | Relevent experience |
| enrolled\_university | Type of University course enrolled if any |
| education\_level | Education level |
| major\_discipline | Major discipline |
| experience | 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 |

**Note that the dataset has a number of missing values and must be treated accordingly before fitting any model. You can refer to this**[**article**](https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/)**for a primer on how to deal with missing values.**

**sample\_submission.csv**

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| enrollee\_id | Unique ID for enrollee |
| target | probability of an enrollee looking for a job change |

**Evaluation Metric**

The evaluation metric for this competition is [AUC-ROC score](https://www.analyticsvidhya.com/blog/2016/02/7-important-model-evaluation-error-metrics/).

**Public and Private split**

Note that the test data is further randomly divided into Public (40%) and Private (60%) data. Your initial responses will be checked and scored on the Public data.

The final rankings would be based on your private score which will be published once the competition is over.