An Examination of Bias and Fairness in Data Science Job Change Predictor

Claire Saint-Donat, Xiangyue Wang

Center for Data Science, New York University

1 Background

In recent years, the growing demands for jobs in data science have driven the creation of a myriad of degree programs, massive online open courses, online certification programs, and boot camps. While the abundance of those learning resources has made data science accessible to a wide range of people looking to enter the field, the variety of credentials has also made it difficult for companies to evaluate the qualification of job seekers. The process of hiring data scientists is often both expensive and time-consuming. Therefore, companies have the incentives to hire qualified data scientists and keep them from leaving.

The dataset, HR Analytics - Job Change of Data Scientists, contains data from a company which wants to hire data scientists among people who successfully pass some courses that the company offers. Based on data about the trainees, the company wishes to predict which of these candidates wants to work for the company after training. The company also wishes to gain insights about what makes a data scientist seek new jobs.

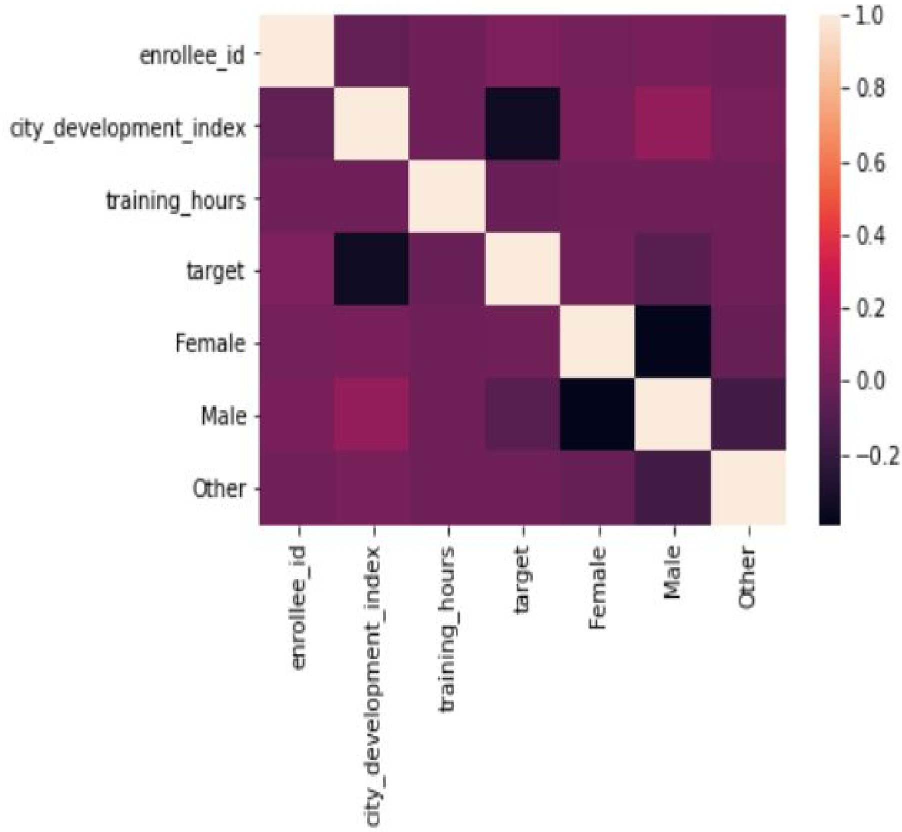
We will examine an Automated Decision System (ADS), developed by the Kaggle user Josh using the dataset, that predicts whether a given data scientist is seeking jobs outside of the company. Using a set of features about a given data scientist, the ADS categorizes them as either a "job seeker" or "non-job seeker". The algorithm is designed to understand the factors that lead a candidate to leave a current job for the purpose of keeping data scientists from leaving a company as well as potentially recruiting data scientists from other companies. We will examine the data, the cleaning process, and the system to evaluate whether the ADS discriminates on the basis of gender or socioeconomic status. If present, such discrimination could result in certain groups of employees receiving less frequent training, promotions and other benefits. Our aim is to design a nutritional label for the ADS that examines the bias in the data, the processing, and one ML model used in this ADS.

2 Input and output

The input of the ADS consists of information related to demographics, education, and experience of the candidates in the training course. The features, shown below, are collected during the training sign-up and enrollment.

Feature	Description
enrollee id	Unique ID for candidate
city	city code
city development index	Development index of the city (scaled)
gender	Gender of candidate
relevant experience	Relevant experience of candidate
enrolled university	Type of University course enrolled if any
education level	Education level of candidate
major discipline	Education major discipline of candidate
experience	Candidate total experience in years
company size	Number 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

Most of the features are categorical, with the only numeric exceptions being "experience" and "company size". The outputs of ADS are predictions of the feature "target", which is a binary label indicating whether the person is looking for job change or not. The ADS also aims to show what factors are the most significant in a person's decision. There is an evident gender imbalance in the data, with almost 70% being male, only about 21% being female. There is also an evident imbalance in people seeking jobs: 75% of the surveyed people are non-job seekers. The correlation between features are shown in the diagram below.



3 Implementation and validation

Our plan for implementation and validation is to first examine the false positive rate and false negative rate based on gender, experience (which correlates with age) and education level (which correlates with socio-economic status) for the random forest algorithm. We will also examine the accuracy and disparate impact based on those groups. We will also measure the impact of how ADS preprocess the data on those statistics.

4 Outcome