Catalyte - 2024

Predicting Job Readiness Using Artificial Intelligence, How Effective Can We Be?

Background

There is an ongoing significant talent shortage in the U.S., particularly in jobs in the tech sector. Many people want to switch over from other sectors to tech as a better career choice in many dimensions, but not everyone has the credentials, relevant work experiences, or even the know-how to make the move successfully. Catalyte.io took the challenge and created a web-based screening application, where we gather a comprehensive set of data from applicants, regardless of degrees, work experiences, or demographics. We built an A.I. powered prediction engine based on actual data gathered throughout the years so that we can identify people with a high probability of success in going through our proprietary training program, becoming apprentices working with us on real projects in real teams inside of actual companies after the training, and eventually getting hired into full-time roles in the tech sector.

Project Scope

The goal of this project is to evaluate and implement new models to build Catalyte's prediction engine using a hybrid of classic and modern artificial intelligence technologies. The project will be carried out in three phases:

1. Understand Catalyte's existing data infrastructure and data lakes, document existing data gathered, and make recommendations on other data/signals that could be used to improve the model without introducing biases.
2. Explore AI models that take all input data and predict candidates likelihood of success, generate a framework to compare different models, make recommendations on a new model
3. Build and implement the AI model with available data, ideally if time allows, build model-refresh mechanisms with new data.

Deliverables

- Project Presentation – A presentation to be given at the end of the semester detailing the findings and suggested next steps.
- Project Documentation – A detailed technical documentation that reviews the dataset available, proposed new signals (if not implemented), a data flow diagram, feature engineering methodology, models evaluated, rationale for the proposed model, and model performances

Project Code – Well-documented code that was used on the project.