**DC DOES**

4058 Minnesota Ave NE Washington, DC 20019 (202) 724-7000

Success Classifier Proposal

**5th June 2021**

Project Team:

|  |  |  |
| --- | --- | --- |
| Name | Role | Office |
| Dhruv Singh | Data Scientist | OLMRP |
| Amjani Jujjavarapu | Analyst | OIT |
| Katrina Mathews | Analyst | OLMRP |

**OVERVIEW**

The Office of Labor Market Research and Performance is responsible for analyzing performance on DC DOES’ workforce training programs. This research project seeks to push such evaluation just a step further, by incorporating predictive analytics into the process.

1. The greatest aim of interest to management is to find out which programs work in getting people employed. This means more funding dollars for programs that are working, with justifiable cause to back it up. This will be the descriptive part of the project.
2. Additionally, an interesting area to explore using predictive analytics is to forecast a new enrollee’s likelihood of success in obtaining employment

The first part of this report will be conducted using regression analysis and will be supplemented with robust exploratory data analysis.

The proposed method for predicting candidate analysis is a survival analysis. However, we may use more traditional machine learning methods as well.

The central data source for this report will be The Locally Developed Maryland Custom Report, on DC Networks. Here we’ve chosen to restrict the data to 2011 to 2021, for purpose of building the model.

Some useful variables in conducting our analysis include program, gender, race, education, CASAS scores.

**PROBLEM STATEMENT**

OLMRP has access to a wealth of demographic information that is not being used to its full benefit. Even the program success that is being tracked using ratios, can be evaluated by comparing programs side-by-side in a regression model to gauge which programs are having the greatest impact.

It is therefore being proposed that we use supervised machine learning models to analyze our data more effectively and achieve better outcomes.

**PROPOSED SOLUTION**

The best solution is to create a sample and demo it before office management. To present results in a short presentation to upper management in order to create buy in for future work in this area. By whetting their appetite for predictive analytics, and giving them a readymade solution, they can reverse engineer solutions, and hire talent to do it better. This aligns with your personal goal of reducing the unemployment rate.

**GOALS**

1. Robust Exploratory Data Analysis – to uncover previously unknown insights – and guide future modeling
2. Implement Regression and Classification Models
3. Document and create presentation with visualizations

**SPECIFICATIONS**

The Locally Developed Maryland Custom Report is a rich source of data on DC Networks that includes information on programs, participants and outcomes. We are therefore seeking to leverage this information to feed it back into our work as public servants. By rigorously evaluating the data we collect, we can improve our delivery of services to the public and help them achieve their goals.

**BUDGET**

1 FTE analyst (OLMRP): 8 hours per week

1 FTE analyst (OIT): 8 hours per week

1 FTE management analyst (OLMRP): 8 hours per week

At an average hourly rate of $55 hourly, this would cost:

$55/hr \* 8 hours \* 3 FTEs \* 4 weeks \* 4 months = $21,210 total.

Since all tools used are open source, and the analysts are equipped with the necessary knowledge, there are no license costs or knowledge acquisition costs to be considered. Equipment used will be standard issue equipment from the office.

**MILESTONES**

**Robust Exploratory Data Analysis**

By September 7th, we would like to have the exploratory data analysis portion of the project automated running at the click of a button. This includes tabulation, summarization, missing data, data transformations, correlations, and visualizations.

In this phase we will also binarize all categorical variables, post EDA., and replace any missing values, drop unnecessary columns and create useful ones.

**Model Building**

The second phase of this project goes from September 7th to 14th, and seeks to fit our explanatory and predictive models. By this time, all binarized columns, and subsetted data should be ready to be fed into our model.

We will have created a binary variable for all those who obtained employment. The last part of this phase, the latter half, should be used to interpret and finetune the model.

**Presentation**

Combining both EDA and insights, we now have a 15 slide presentation for upper management.

Causing them to rethink some of their problems

And to propose new solutions moving forward.

Note: Best Practice is to have a Presentation ready for each stage of the analysis. And to continually refine it so you have something to report on.