HACKERU

Data Science & Analytics

Mini Project

Introduction

This project allows you to show your ability to develop a data science/machine learning analysis. It should include data wrangling, exploratory data analysis, data visualization, building a classification model, and an evaluation of the model. We leave the choices of your approach to you. Whatever you choose, you should be able to explain and justify the choices you are making for this analysis. You should use either the R or Python programming language and their respective libraries to carry out your goal. This will help you show your thought processes and technical abilities in data science/machine learning.

Once you create your analysis, you will be presenting it to the hiring team. People participating in the interview process may ask questions to you during the presentation to ask you clarification questions and give you an opportunity to demonstrate your understanding of analysis and analytics. You will send your complete analysis for review prior to attending to present your analysis to the team.

We expect this project to take no more than four (4) hours of your time. Please complete it and return it to us as soon as possible. Typically, we'd like your project completed within 3-4 days, sooner if you are able. The team will review the project and invite you to present it. We also expect this to be only YOUR work with no outside aid.

Requirements for your final report

- Demonstrate your ability to cleanse data, find outliers, and engineer new features
- Demonstrate techniques of exploratory data analysis and exploratory data visualizations
- Demonstrate the ability to build a predictive model and evaluate the results of the model

The Data Set

http://archive.ics.uci.edu/ml/datasets/Adult

Use the Adult dataset to make predictions on whether income exceeds $50K/year based on census data. Be able to address each of the following questions

The **data cleansing part** of your analysis should incorporate (programmatically, in code):

- if outliers exist
- if there is any bogus data
- drop rows that have NULL values in any column
- drop the column named "native-country"

HACKERU

Data Science & Analytics

Mini Project

- engineer a new feature, using 1 if the person makes greater than 50k/year, otherwise 0.
- any necessary data transformations for downstream analysis
- the number of rows and columns in your cleansed data after all of the previous steps are completed

The **exploratory data analysis** part of your analysis should incorporate:

- data visualizations using the visualization library(ies) of your choice
- box plots, histograms, and bar charts as appropriate (you need not create visualizations of every feature within the data set)
- insights gained from some of your visualizations

The **predictive model** part of your analysis should incorporate:

- At least one classifier that you are familiar with and can speak about in detail
- Why you chose the classifier you did. You may want to make comparisons across 2-3 classifiers
- evaluation metrics of the classifier and its ability to make predictions
- strengths and weaknesses of this classifier

Your entire analysis should be completed within three-four hours, and you will have up to 45 minutes to present it to the interviewing team via Zoom. Be prepared to engage in a discussion about your analysis. You may be much stronger in some areas than others, and we expect that. We do not expect this analysis to be perfect – but we want you to show what you can do with data science and analytics with either R or Python. It is up to you. You will have a great opportunity to discuss the mini-project with the team. Keep in mind that the team will want you to spend approximately 1/3 of your time on each of the sections, saving the final 10% of the discussion to go over the strengths and weaknesses of your analysis. The interviewing team may ask directed or clarifying questions during your presentation.

When you have finished your project, please make a .ZIP archive of your files. Name the zip file with your last name. Once you submit your project, your technical interview presentation will be scheduled.

HACKERU

Data Science & Analytics

Mini Project