

Facet Machine Learning Coding Challenge

Problem:

This challenge is designed for you to demonstrate your programming prowess and ability to implement a machine learning algorithm using Python and Scikit-Learn. The data that you will be analyzing contains an assortment of features, which represent prospective applicants applying to the position Chief Troublemaker. Out of the 35,644 applicants 1,075 have had positive interviews and 2,143 have had negative interviews. We want to know which of the remaining 32,426 applicants are the most likely to have positive interviews.

Instructions:

- Use the data contained in the csv that was attached with these instructions. The csv contains both the labeled and unlabeled data. You need to apply labels to the unlabeled data.
- Use either a Support Vector Classifier or Random Forest Classifier to solve the problem. You can use any tools at your disposal to tweak the hyperparameters.
 - <http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>
 - <http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- Write a short summary about why you chose that classifier and add it to the ReadMe for your project. The summary should contain some information about the precision and recall.
- The results of your solution should be an ordered list of ids with a label applied representing a positive or negative interview outcome.
- If you feel like there are any improvements that could be made to your solution, but you are unable to implement them during the allotted time, then note them in your ReadMe.
- Finally, include some metrics and a confusion matrix in your ReadMe.

Requirements:

- The solution must be written in Python.
- The solution must use the Scikit-Learn library.
- The code must be committed to a GitHub repository.
- Your solution should output a list of labeled ids in a csv.

Submission:

- You should submit your ordered list in the format of a csv to gorkem@facetwealth.com and benjamin@facetwealth.com. Do not include this in your repo.
- Your submission should also contain a link to your GitHub repository with the solution containing the code and ReadMe for this challenge.

Email either of the two emails listed in the submission instructions above if you have questions. You should spend 4-5 hours on this project. It is ok to not have the best solution to this problem. Our primary concern is your ability to implement this solution logically and thoughtfully. Good luck!