**Task:**

Our problem for project 5 is about working for a non-profit that is trying to recruit student volunteers to help with Alzheimer’s patients. Our job is to predict how suitable a person is for this volunteering job. We are using the young people survey dataset from Kaggle.com provided in the HW5 write up.

**Preprocessing step:**

What we did for the preprocessing was convert all the categorical data into integers. The reason behind this was that we planned on running different models that may have not been able to process categorical data such as K Nearest Neighbors. The categorical data was converted to integers ranging from 1-6.

The first thing we did was to convert all the answers that are strings into either 1 to 2 or 1 to 5 integers depending on how many possible answers the person could answer. The reason why we did this is because you can’t have integers and strings in your data. It makes sense to convert the strings to integers for this problem. We also added a response to the empty answers because it makes more sense to add a random number instead of adding a zero or dropping the rows that don’t have an answer. After we are done processing the data then we split the data into 70% training, 10% development, and 20% test. We save these into their own separate files.

The solutions we chose for this project were binary tree classifier, SVM “poly”, and most frequent classifier, and perceptron because we feel confrontable working with these classifiers and they would work with binary classification.

The technologies we used for this project are Python3, Jupyter Notebook, Pandas, Random and Scikit for perceptron, SVM, BT classifier, and we coded our own most frequent classifier similar to the homework 2.

**Sources**

**Scikit-learn -** <https://scikit-learn.org/stable/>

**Scikit-learn documentation**

<https://scikit-learn.org/stable/documentation.html>

**Pandas** - <https://pandas.pydata.org/>

**Pandas documentation**

<http://pandas.pydata.org/pandas-docs/stable/>

**Numpy and Scipy** - <https://docs.scipy.org/doc/>

**Numpy and Scipy Documentation**

<https://docs.scipy.org/doc/numpy/>

**Young people survey csv files**

<https://www.kaggle.com/miroslavsabo/young-people-survey/>

**Pickle** - https://docs.python.org/3.1/library/pickle.html