Project Documentation

This document is meant to merely show what I learn from the project from the following source:  
<https://thecleverprogrammer.com/2020/12/06/resume-screening-with-python/>

# Prior Knowledge

I have done machine learning projects a long time before. Although I originally got confused with AI and machine learning, I soon learned the difference. I did some simple graphing work and categorization with housing. Also, I did the classic iris-classification project.

However, it has been a long time since I’ve worked on those projects and I would need to reference the libraries typically used and probably have to go through the entire workflow again in order to redo them.  
  
As for NLP, I have no knowledge of how it works. I personally have only used machine learning when it’s based off of a rigid dataset. So, any sort of analyzing from organic speech would be very new to me.

# On the Go Learning Notes

Refamiliarized myself with pandas. Learned it stands for Python Data Analysis Library

At first I was confused what the xticks in line 28 was doing. I realize now that it’s what is making the graph go to the side instead of up.

A step not mentioned in the project that I troubleshooted was why the graph wasn’t displaying after line 30. I realized that I needed the plt.show() command.

As the code went on with little description of exactly what it was doing, it became much harder to follow. Eventually I turned to copy and pasting the code and analyzing it later than understanding it as I was writing it.

# Conclusive Notes

I refreshed my memory on a few things. Using matplotlib to graph data, the KNeighborsClassifier to predict data.

There was a lot of new things I learned though.

* How to make a wordcloud display in matplotlib, and how disabling the axis prevents it from being a normalized graph
* Finding use count and using it to determine which words are necessary to determine which job the resume applies to
* Formatting floats was new to me in Python. It took me a second to understand what the lines such as %1.1f%% on line 47 were actually doing
* I’ve seen Regex before, but this was both the first time I’ve actually seen it applied in code and used for a purpose in that code.
* Word tokenizers and stop words to split apart long strings and exclude unnecessary words was new
* For a long time I didn’t understand what the Resume Screener was actually doing. After reviewing the input text and the target, including the test input, it was what made me realize that the machine is learning to guess the job from the text in the resume. I would assume this would be used in order to see if the resumes of applicants meet the criteria for the posted job.

All in all, it was a great example of machine learning to review what I’ve learned in the past and how NLP, with word tokenizers and stop words, can be used when applied to them. I hope to dive deeper into NLP and see if I can make any personal projects with it.