1. **Work overview:**

I am aiming to narrow down the factors that affect voter turnout specifically through social media sentiment as a step to the creation of a model to forecast the percentage of voter turnout in each constituency. I hope to provide an accurate model of the voter turnout, along with specific factors that could be used to increase turnout in future elections. This could also lead to the determination of more localised factors (both geographic and demographic) for statistics in aid of targeted advertising towards increasing voter turnout.

To achieve this, I will use a combination of web scraping, natural language processing and a classification style neural network. Alongside which a series of more standard statistical and coding techniques will be used.

1. **Technologies and Materials:**

I intend on primarily using python to develop the model as it is versatile and should provide all of the features (along with a wide array of useful libraries) required to carry out this project. There are several well developed and maintained python libraries for both natural language processing and neural network development, and I can envisage using the Natural Language Toolkit and TensorFlow respectively. This is due to their superior userbase and online resources. I also intend on using the Python library Tweepy as a means of accessing the Twitter API. There should be no extra equipment or physical materials required for this project.

1. Python –

<https://docs.python.org/3/>

1. Natural Language Toolkit –

<https://www.nltk.org/>

1. TensorFlow –

<https://www.tensorflow.org/api_docs/python>

1. Tweepy –

<https://tweepy.readthedocs.io/en/latest/>

1. **Motivation research:**

A general interest in politics along with using computers to answer questions is guiding my proposed project. This along with the gap in voter turnout between ages, particularly in recent years, leads to questions of why, which I feel can be answered using similar techniques as people have previously used to forecast election results.

The Prediction of Voter Turnout – By Richard F. Yalch - <https://www.acrwebsite.org/search/view-conference-proceedings.aspx?Id=12062>

Social media popularity and election results: A study of the 2016 Taiwanese general election – By Xiaodong Zhang - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6261632/>

1. **Existing knowledge:**

I will be utilising knowledge gained from my previous two years, in particular the Problem Solving for Computer Science module which provided me with a wealth of Python skills. From the modules I am currently taking, all of Neural Networks, Machine Learning and Data Mining should provide me with transferable skills to be used in this project.

1. **New knowledge:**

I will need to learn a set of new skills surrounding web scraping, particularly taking and formatting data from social media sites (largely Twitter) and find out how to use a variety of tools offered by the Twitter API. In addition, I require knowledge surrounding the possibilities/ limitations surrounding natural language processing and how it can be implemented with classifying machine learning model.

The modules I am taking this year combined with online resources should provide the knowledge required for the machine learning aspects of this project. I will be acquiring all other necessary skills from online resources/ documentation.

1. <https://www.promptcloud.com/blog/scrape-twitter-data-using-python-r/>
2. <https://towardsdatascience.com/gentle-start-to-natural-language-processing-using-python-6e46c07addf3>
3. <https://www.youtube.com/watch?v=FLZvOKSCkxY>
4. <https://www.youtube.com/watch?v=wlnx-7cm4Gg>
5. <https://www.youtube.com/watch?v=1gQ6uG5Ujiw>
6. <https://www.tensorflow.org/tutorials/keras/classification>
7. <https://www.youtube.com/watch?v=wQ8BIBpya2k>
8. **Timeline and milestones:**

Include as much detail as possible.

Be specific about what you will research when and what parts of your code and piece you will build when

Include contingency planning - what is your minimum viable product ?

The minimum viable product for this project would be a working model with a accuracy validation benchmark against previous votes dating back as far as accurate Twitter data can be obtained.

Project Repository - <https://gitlab.doc.gold.ac.uk/jashw002/final-project>