School of Electrical and Information Engineering
University of the Witwatersrand, Johannesburg
ELEN4002/4012: EIE Investigation (Capstone Project) 2022

Tender Bid for Project Number: 22p65

Project Title: Personality prediction using online social platforms.
Group Number: 22G51

Project Overview: *(give a brief outline of how you will approach the investigation.)*

The aim of this project is to predict an individual's personality type from their social media account. The personality type will fall into 1 of 16 types outlined by the Myer Briggs personality test. This personality type is based off of the individual introversion, intuition, perceiving and thinking capabilities. To make this prediction a Natural Language Processing (NLP) tool and various sentiment analysis techniques will be implemented on the social media posts. This will create a data set which will be analysed by multiple machine learning algorithms, such as the Naïve Bayes Algorithm, to make a classification of the individual's personality type. This will then be compared to the individual's actual personality type to determine the accuracy of the algorithm to make a correct prediction. This process will be done for multiple algorithms to determine the most effective machine learning algorithm for this scenario by weighing performance vs complexity.

Preliminary Budget & Resources:

Budget:
This project has no need for a budget as there is nothing that needs to be purchased to perform the steps outlined in the project overview. The project also has the benefit of being able to be completed remotely.

Resources:
Python and matlab
Social media posts/data (already acquired)

Weekly Milestones: *(give specific deliverables for each week.)*

1. **Week 1: Planning and setup**
   The project specifications need to be clearly defined and understood. This step will also include choosing an effective NPL tool and outlining some algorithms to be used.
2. **Week 2: Data collection and processing**
   Once the NPL tool is set up, the social media posts need to be automatically extracted and processed into a format/training set for the algorithms. The quality of the information needs to be evaluated to identify any shortcoming or bias which may be present.
3. **Week 3: Algorithms and setup**
   A script will be created to process the information. This will be a general model which will work in conjunction with the various algorithms. The algorithms to be used will also be chosen.
4. **Week 4: Algorithm refinement and application**
   The final solutions should be developed, all bugs need to have been removed and testing will begin.
5. **Week 5: Processing and evaluation.**
   Comparisons of different algorithms need to be conducted so that conclusions can be drawn. This will include an error analysis to determine the accuracy of the developed systems.

Risks and Mitigation:

1. **Time management:**
   The project members currently have not completed a machine learning project before however to mitigate this completely the members chose the course Software Engineering which has a focus on machine learning and AI. This will allow smooth and timeous operations of the project.
2. **Project setbacks:**
   Unseen requirements may appear which may create setbacks. To mitigate this, there will be effective communication with the supervisor, Dr. Yuval Genga, to ensure that these are identified as soon as possible so that they may be dealt with.
3. **Bias:**
   Bias is a serious problem which can cause extreme inaccuracies. To mitigate this all forms of bias, need to be understood in the early stages of the project so that it can be avoided when choosing algorithms and selecting applicable data.
4. **Data quantity and quality:**
   Low data quantity and quality will lead to inaccuracy. To mitigate this an understanding of how much data is needed for accurate results will be obtained. If this is not possible a full breakdown as to why the results were skewed, will be given.

- Once complete, save as pdf before submitting – no other format will be accepted.
*- Any submission longer than **1 page** will be recorded as a non-submission and an SP warning will be issued.*