|  |  |
| --- | --- |
| Student Name: Max MacDonaldMobile Number: 0861991455 | Student Number: C15740661 **Supervisor: Cindy Liu** |
| Project TitleThe Truth about Twitter | |
| Summary The goal of this project is to create an artificial intelligence, within a client application, that can decipher, based on its model, whether a twitter account is fake or real.  The user will be able to input the desired twitter account they wish to verify into the client application which will call on the Twitter API for data, this data will be formatted and passed to the artificial intelligence in the backend which will then decide on the veracity of the account.  A dataset will be acquired from one of several sources online. This will be carefully selected based on the data used within then possibly cleaned and reformatted to ensure it is in the desired format.  The model, that the artificial intelligence will use, will be built up using this dataset until it is effective at deciding whether an account is real or fake. I will then attempt to refine the model to get to as high a level of accuracy as I can.  A simple client application will be constructed alongside this that can call on the Twitter API to gather data related to a specified twitter handle. This will then have the artificial intelligence integrated into it to produce the final product. | |
| **Background**  As someone who has grown severely frustrated with the amount of fake social media accounts that have sprung up in the last number of years, whether they are Twitter accounts pushing political propaganda or Instagram users who have purchased bot accounts to like their posts to boost their accounts, this project gives me a chance to sink my teeth into this issue in a manageable way.  While I haven’t had much exposure to the areas of data analytics and machine learning, I am quite familiar with some of the technologies behind them. While on work placement I learned Python through self-learning and employed it heavily in my work. I will be researching heavily into R and all the main python modules that will be needed for this in the first few weeks.  My hope is that through this project I can continue to polish my current skillset while adding to it as well as learning new technologies, methodologies and techniques that I can take with me into my career post-college.  **Possible Datasets**  <https://botometer.iuni.iu.edu/bot-repository/datasets.html>  <https://github.com/jubins/MachineLearning-Detecting-Twitter-Bots>  **References**  Efthimion, Phillip George; Payne, Scott; and Proferes, Nicholas (2018) "*Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter Bots*", SMU Data Science Review: Vol. 1 : No. 2 , Article 5. | |
| Kelleher, John; Mac Namee, Brian; D'Arcy, Aoife (2015) “*Fundamentals of machine learning for predictive data analysis*”, Cambridge Massachusetts: MIT Press;.  Raschka, Sebastian (2015) *“Python Machine Learning”*, Birmingham, UK: Packt Publishing Ltd;  Richert, Willi; Coelho, Luis Pedro (2013) *“Building Machine Learning Systems with Python”*, Birmingham, UK: Packt Publishing Ltd;  Morstatter, Fred; Liu, Huan; Kumar, Shamanth (2013) *“Twitter Data Analytics”*, Springer New York  **Web References**  <https://www.toptal.com/python/twitter-data-mining-using-python>  <https://www.rstudio.com/online-learning/>  <https://stackabuse.com/accessing-the-twitter-api-with-python/> | |
| Proposed Approach I propose to use the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology to control the lifecycle of the A.I. being used in this project. The diagram below lays out the steps involved in this methodology. The Client application is discussed at the end separately. | |
| **Research**  I plan to spend a lot of time on this section, and it will continue throughout the entire project, as I have a lot to learn in the areas of Machine Learning, Data Mining and A.I. before I can proceed with understanding the data or implementation of the model. I will also be researching other projects that are similar and trying to learn what they did right and avoid what they did wrong.  **Data Understanding and Data Preparation**  Once I have acquired a dataset for use, a significant portion of the project will be assigned to understanding the data and preparing it for use in the model. I will have to investigate thoroughly the correlations between each of the data features to ensure as accurate a model as possible.  **Modelling & Evaluation**  I have investigated various ways of implementing the model such as with R or Python and am heavily leaning towards Python, mainly due to personal preference due to using it heavily while on work placement, although I am open to using other tools. This will be decided upon during the initial research phase while learning more about the core topics of this project.  The model will use supervised learning as the datasets I have found all include labelled outcomes. An evaluation method will also be chosen during the research stage although I have already started looking at some methods including Confusion Matrix & k-Fold Cross-Validation. Evaluation will cycle back around to Business Understanding multiple times as the accuracy of the model is fine tuned.  **Client Application**  I propose to follow the Waterfall Model for the client application as it is quiet straightforward outside of the A.I. which will be incorporated near the end of the project. Its main job will be to call on the Twitter API for data, passing it onto the A.I and then displaying the results back. | |
| Deliverables  * Interim Report * Project dissertation * Client application, with built-in Artificial Intelligence, for demonstration | |
| Technical Requirements  * Laptop * Python 3, R | |

## Project Reviews – Please include reviews of two of LAST 2 years projects from either DT228, DT282 or DT211C.

|  |
| --- |
| **Project 1**  **Title:**  A Computer Application for Teaching Children Computer Skills  **Student:**  Jennifer Seery  **Description**  The objective of this project was to create a customisable and inclusive learning application in the computing field such that those with Autism Spectrum Disorder can use it alongside their peers without ASD.  **What is complex in this project**  Enabling the application to be tailored for any child with ASD as it is such a broad disorder.  Getting feedback from target userbase as children with ASD cannot always convey their thoughts into words and coherent sentences well.  **What technical architecture was used**  3 tier architecture:   * Electron framework, for the front end, which allows the use of HTML, CSS and JavaScript * Django, a Python framework, for the middle tier * Heroku as the back-end server used to host the application   **Explain key strengths and weaknesses of this project, as you see it**  Strengths  Everything is thoroughly researched and well thought through.  Good implementation of numerous new technologies and Jennifer wasn’t afraid to change to a better fitting technology if she found it.  Weaknesses  The application is only developed to a certain extent, as there is a lot that can be done with it. which Jennifer herself notes, stating more filters could be added alongside new tutorials to further develop the project if it was taken on after final year. |
| **Project 2**  **Title:**  Driving Instructor Assistor  **Student:**  Baolach Morrison  **Description**  The objective of this project was to create an Android application which assists driving instructors while on a driving lesson, keeping track of all necessary business information and displays example locations for practicing specific manoeuvres.  **What is complex in this project**  The use of geographic information system technology to add locations on a map through the app.  Connecting the front end to the selected back end. While a solution was found, choosing a different back end could possibly have made this easier.  **What technical architecture was used**  Model View Template (MVT) architecture: Java through Android Studio for the front end  * Android API and Django REST framework as the middle tier connecting front to back * Django, a Python framework, as the back end   **Explain key strengths and weaknesses of this project, as you see it**  Strengths  Fills a market gap by aiming to help the driving instructor instead of the client as most apps in this area are aimed at the latter.  Nice simple layout with good colour contrast and several useful functions.  Weaknesses  Not accessible offline and the login component is not that secure. Both are acknowledged by Baolach at the end, stating with more time he would have liked to address these. |

