

|  |  |  |
| --- | --- | --- |
| The Truth about Twitter | | |
| **Student Name:**  **Max MacDonald** | **Student ID:**  **C15740661** | **Supervisor Name:**  **Cindy Liu** |
| **Link to Software Repository:** |  | |

**Table of Contents**

1. Project Statement 3

2. Research 3

Background Research 3

Alternative Existing Solutions to Your Problem 3

Technologies Researched 3

Other Relevant Research Done 3

Resultant Findings and Requirements 3

Bibliography 3

3. Approach and Methodology 3

4. Design 4

a. Technical Architecture Diagram 4

b. Other Design Documents 4

5. Prototyping and Development 4

6. Testing 4

7. Issues and Risks 4

8. Plan and Future Work 4

# Project Statement

Social media sites, such as Facebook, Instagram and Twitter, have been flooded with numerous types of fake accounts in the past few years and each of these sites have an urgent need to be able to tell the difference between a real user and a bot user.

The goal of this project is to create a web application that users can use to find out the likelihood that a Twitter account is real or a type of bot account. Users of this application will be able to input a Twitter handle that will be used to hit the Twitter API, retrieving specific information about the account. This data will be passed into a machine-learning model which has been trained on a dataset of real and bot accounts. The model will then decide whether the account is real or not, displaying it’s answer to the web front with relevant statistics.

# Research

What research has been done and what are the outputs? This concerns background research on the topic itself, as well as any research on the data analytics that you plan to perform and the technologies that you wish to use. Use the following headlines to fill in your information. Use one of the standard citation styles consistently throughout this section. Discuss the following:

## Background Research

About twitter, social media and fake accounts,

### Types of Bots Accounts

benign vs malignant. Traditional vs newer social bots

### Importance of Identifying Bots Accounts

why fake accounts can be so damaging,

### Characteristics of Bot Accounts

important characteristics of fake accounts,

## Alternative Existing Solutions to Your Problem

Twitters own attempts and studies done on this

### Twitter

### BotRNot

### 2.2.3 Academic Study

## Technologies Researched

Details of techs researched.

### 2.3.1 Technologies for Data Mining & A.I. Model

Python + Django framework, Scikit, Pandas, Numpy, Mathlib

### 2.3.2 Technologies for Web Application

Twython, Twitter API, R, Java, Apache vs Heroku vs Amazon, MySQL vs Hadoop vs ElasticSearch vs Postgres

### 2.3.3 Technologies for Version Control

GIT in PyCharm

## Other Relevant Research Done

Approaches, Methodolgies and Technical architectures dealt with in their own section.

### Data Science

#### Big Data

#### Data Mining

#### Machine Learning

## Resultant Findings and Requirements

Chosen techs.

### Chosen Technologies

### Chosen Data Sets

## Bibliography

Your researched resources. Use Vancouver citation style.

# Approach and Methodology

What is your approach to this project? Are you using any particular software methodology? Eg. Are you delivering design/ code in phases, or are you completing all design up front, followed by all coding? Have you some sections lower priority if time runs short?

Talk about all theses in terms of the project then sections explaining them

Cross Industry Standard Process for Data Mining (CRISP-DM), SEMMA, KDD

Agile/ Scrums / Sprints initially Waterfall

## Agile & Scrum/Sprint

## Data Mining Project Management Models

#### KDD

#### SEMMA

#### CRISP-DM

#### Differences in Models

#### Conclusion



# Design

## Technical Architectures

Model view template which is a Django’s variation on the MVC

## Technical Architecture Diagram

Insert the architecture for your solution

## Other Design Documents

Insert other design artefacts that explain your system: e.g. Use cases/ ERDs/ Class diagrams

# Prototyping and Development

Explain exactly what prototyping and development you have completed.

## Vertical Prototype

Web app which can hit the twitter API and basic model

## Development

# Testing

Explain your planned testing approach: For example: who will be involved, what test scripts are planned, how will the testing be executed.

Random samples from datasets as well as researching and selecting several accounts to test against.

# Issues and Risks

Explain the main issues / challenges that are unresolved on your project. – and your suggested approach to solving them. This is a critical part of your report to show that you understand what is required to complete the project.

Proper understanding of the data and connections between them.

Creating an accurate model, choosing classifiers and algorithms.

# Plan and Future Work

What are the key deliverables and date for the remainder of the project? Include a GANTT chart or another means of illustrating your project plan.

Key deliverables: Dissertation

Integration of accurate model into web application ie combining both parts of the vertical prototype.

