1. Introduction

In recent years, an increase in unreliable information, often accessed via social media sites, has emerged as a vital and ongoing crisis in many modern communities. Social media sites provide access to quick, low cost and easily accessible news. However, these same qualities have allowed social media to be largely impacted by an influx of fake news.

Fake news is commonly understood to represent information presented in a misleading manner, this information is then shared by members of the public, whether deliberately or accidentally. In recent history, fake news is demonstrated to have impacted notable events, for example it can be argued that the 2016 US election was largely impacted by fake news. [1]. With a fear that the 2020 election was also impacted.

During the COVID-19 pandemic fake news has greatly impacted public understanding regarding key issues. An example of this is that fake news in relation to the effectiveness of chloroquine led to an increased number of overdose cases involving chloroquine [2]. Public misinformation regarding COVID19 is also thought to have caused the suicide of a man who, after watching a number of online videos, convinced himself that he had contracted COVID19, and took his own life to avoid spreading this to his family [3].

Public consumption of fake news regarding COVID-19 has also led to a spike in racism, particularly towards Chinese communities, who have faced a number of issues perpetrated by both individuals and businesses, an example of this is that several businesses have refused to provide service to Chinese communities [4]. Additionally, it can be evidenced that there has been an increase in anti-Asian hate crime, including verbal and physical attacks [5]. For these reasons, it is essential to detect both misinformation and people who spread misinformation.

The problem of detecting fake news has received a considerable amount of research attention. Content-based approaches are commonly used in to solve this classification problem. In content-based approaches, visual or textual features that are indictive of fake news are identified and used to make predictions about the validity of news articles. For example, [6] incorporated text features such as number of pronouns and swear words into a neural network to make judgments about the truthfulness of news articles. Other researchers have extracted emotions from text and have demonstrated that emotional features can be utilized to detect fake news [7]. One of the main problems with content-based approaches is that fake news articles are often written in a similar fashion to real news articles, which can make it difficult for machine learning algorithms to discriminate between the two. [8] To avoid this issue, some researchers have attempted tackle the problem of fake news by identifying its disseminators, rather than the news itself.

The body of literature regarding the detection of people who spread misinformation grew rapidly in 2020 as a direct consequence of the 2020 PAN at CLEF competition. The organisers of the competition hypothesised that the characteristics of social media profiles which spread fake news differ to that of profiles which do not. To test their hypothesis, they challenged participants to develop algorithms that can identify these differences and use them to detect fake news spreaders on twitter. Most participants developed their models using Support Vector Machines (SVM), Logistic Regression, or a combination of both. A large variety of different features were incorporated into these models. For example, Manna et al. [9] used stylometric features (e.g. punctuation marks, capital latter’s) emojis (classified by emotions) and lexical features (E.G expressions related to clickbait headlines). Russo [10] used stylometric features and a range of text-based features indicative of emotions (E.G sum of all lemmas indicative of sadness). The most accurate model for detecting fake news was proposed by Buda and Bolonyai [11]. They used 5 different machine learning models combined with n-grams and other stylometric features such as average tweet length or lexical diversity of tweets. The model was able to detect fake news with an accuracy rate or 77.75% percent.

NEED CONCLUSION/HYPOTHESIS - Discuss with Supervisor

2. Description

## 2.1 Aims of the project

This project has as its main aim the creation of classification models of fake news spreaders in a social media context. The creation of these models is motivated by the existence of models [12-15] proposed in the last four years for the identification of fake news. The team intends on improving what these models do by classifying not which news are fake or not but who spreads them and, in this way, find a way of preventing these fake news from spreading at all. This classification of who is a fake news spreader or not will consider factors such as a user’s posts, likes, shares, and so on.

The second aim of our project is the testing of our research hypothesis which is represented by the following statement: “An attention model can improve the accuracy of the classification of fake news and by consequence the identification of fake news spreaders”

## 2.2 Objectives

In line with the aims of the project, three main objectives were defined:

1. Determine if the work previously done by others can be used and/or further improved upon or if the project requires doing the work from scratch.
2. Depending on the previous objective the next step is the design or improvement of classification models which are suited for the problem that the project’s team is proposing to solve.
3. Related to objective (2), we have the implementation of proofs of concept for the models in order to test and prove their feasibility and accuracy.

## 2.3 Milestones

The milestones defined for the project are as follows:

1. Pre-process the data from the dataset(s) to be suitable for use.
2. Design the algorithm/model to be used.
3. Implement the decided model/algorithm in code.
4. Finish testing the code and record the final results.
5. Finish writing a paper stating the development of the project and its results.

## 2.4 Expected outcomes

The main outcome that is expected of this project is the successful creation of a classification algorithm which presents an accuracy that appears satisfiable to the team. The number of models/algorithms designed can range from one to several depending on the team’s efficiency during the project. Considering that part of the team does not come from a Computer Science background and those that do have a more basic knowledge of machine learning, the team will likely create one or two models at most, due to the requirement of learning certain aspects of the knowledge that is needed for the project during the project.

We also make a distinction between essential and desirable outcomes:

|  |  |
| --- | --- |
| Essential | Desirable |
| * Creation of one classification model * Achieve an accuracy in line with previous existent work | * Create additional models of different natures * Achieve an improvement of 3-4% over the top results of other pieces of work and research |

3. Plan

3.1 Work breakdown and timescales

* Data finding and processing (6-8 days)
  + A. Search across Twitter, Facebook, GitHub, Kaggle, etc for good datasets (3-5 days)
  + B. Check ethics (3-5 days/concurrent with above)
  + C. Data pre-processing (3 days)
* Model design (9 days)
  + D. Research previous existing models’ implementation/code (3 days)
  + E. Decide model to implement/approach to follow through with (3 days)
  + F. Determine changes needed to do from previous models (3 days)
* Development/Coding (15-21 days)
  + G. Implement the decided model/algorithm (7 days)
  + H. Design attention layer for neural network (7 days)
  + I. Train model (1-7 days)
* Testing and fine-tuning (3-9 days)
  + J. Run evaluation data through the model and do hyperparameter tuning (1 day)
  + K. Run test data to analyse accuracy of model (1 day)
  + L. Change model and go back to training if needed (1-7 days)
* Dissertation (14 days)
  + M. Write paper with conclusions from results and deciding if research hypothesis was validated or not (7 days)
  + N. Review paper and do error checking, so on (7 days)

3.2 Activity sequencing

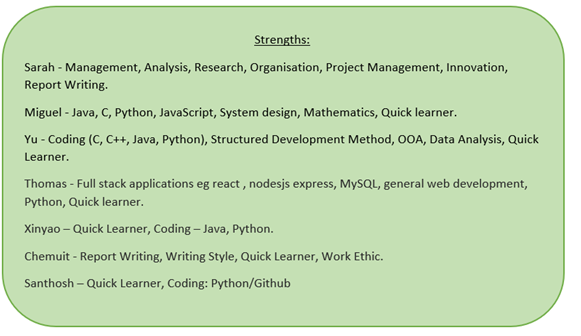
图示

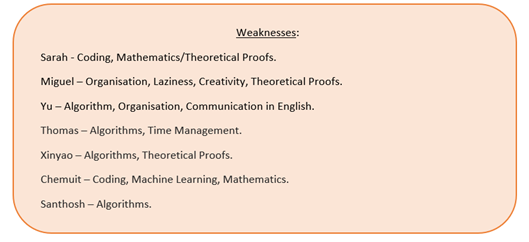
描述已自动生成

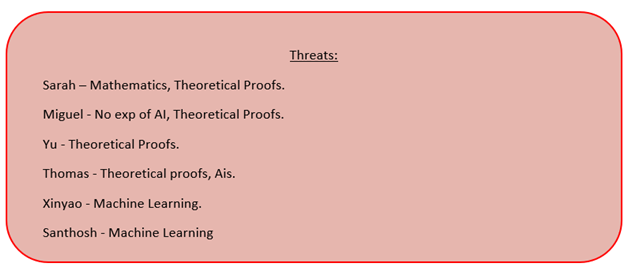
Critical Path: 5+3+3+3+7+7+7+7+7=49 days

3.3 Scheduling and Allocation of Team Members

At the beginning of the project Group 4 completed a SWOT analysis to discover areas of Strength, Weakness, Opportunity and Threat. Group 4 allocated tasks throughout the project based upon this analysis, this will ensure that all team members have the opportunity to showcase their strengths, while working in a supporting role to develop further skills in areas of interest.







3.4 Risk Management

4. Resources and Data

4.1 Data sources

This project’s main goal is to detect and predict fake news super spreaders in social media. Our data sources will mainly come from social media, such as tweets and online news, etc. We have also considered downloading datasets from some websites that provide this kind of dataset. Additionally, we can also collect data present in previous research if it is available and allowed for use.

The LIAR dataset contains brief news statements and user profile information such as subject, state, text/location, political party, speaker, and previous history [16]. The examples in the data set come from different sources, including interviews, political debates, tweets, press releases, and Facebook posts.

BuzzFeedNews dataset contains complete samples of news published by 9 news agencies on Facebook within a week of the U.S. election from September 19 to 23 and September 26 to 27, and each post and linked article has been verified. Buzzfeed News has two data sets, one is a fake news data set, and the other is a real news data set in the form of a csv file. Each dataset has 91 observations and 12 feature variables.

* Self-collected data: Twitter, Reddit, Facebook
* Dataset repositories: GitHub, Kaggle
* Public dataset: LIAR, BuzzFeedNews

4.2 Software configuration

Data collection can be done through the API interface of the relevant platform or GUI. Middleware configuration may be required here, such as Storm. We can also use Python web crawlers/scrapers (like Scrapy) to collect data from the previously mentioned websites.

The research model will be built in Python. Depending on the size of the dataset and nature of the classification model, the HPC facilities provided by the university might be what will be used for training.

4.3 Ethical issues

The data used in this project will be either from publicly available repositories such as GitHub and Kaggle or it will be extracted in real-time from websites such as Twitter or Facebook. All these sources and the information contained in them are of public domain and therefore do not require ethical approval. Some of them however, namely Twitter, require the hashing/masking of user ids for the purposes of using their data and doing research on it. These and any other kind of Terms of Service conditions will be complied with.

5. References

[1] Allcott, H. and Gentzkow, M., 2017. Social media and fake news in the 2016 election. Journal of economic perspectives, 31(2), pp.211-36.

[2] Busari S, Adebayo B. Nigeria records chloroquine poisoning after Trump endorses it for coronavirus treatment. CNN; 2020 Mar 23. Available from: <https://edition.cnn.com/2020/03/23/africa/chloroquine-trump-nigeria-intl/index.html>.

[3] Wallen J. Coronavirus: Indian man ‘died by suicide’ after becoming convinced he was infected. Telegraph. Telegraph; 2020 Feb 12. Available from: <https://www.telegraph.co.uk/global-health/science-and-disease/coronavirus-indian-man-died-suicide-becoming-convinced-infected/>

[4] Rich, M., 2020. As coronavirus spreads, so does anti-Chinese sentiment. The New York Times, 30.

[6] Rashkin, H., Choi, E., Jang, J.Y., Volkova, S., Choi, Y.: Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 2931–2937. EMNLP ’17 (2017)

[7] Ghanem, B., Rosso, P., Rangel, F.: An Emotional Analysis of False Information in Social Media and News Articles. ACM Transactions on Internet Technology (TOIT) 20(2), 1–18 (2020)

Detection of Online Fake News Using N-Gram Analysis and Machine Learning Technique

[8] <https://arxiv.org/pdf/1712.07709.pdf>

[9] Manna, R., Pascucci, A., Monti, J.: Profiling Fake News Spreaders through Stylom- etry and Lexical Features. UniOR NLP @PAN2020. In: Cappellato, L., Eickhoff, C., Ferro, N., N ́ev ́eol, A. (eds.) CLEF 2020 Labs and Workshops, Notebook Papers. CEUR-WS.org (Sep 2020)

[10] Russo, I.: Sadness and Fear: Classification of Fake News Spreaders Content on Twitter. In: Cappellato, L., Eickhoff, C., Ferro, N., N ́ev ́eol, A. (eds.) CLEF 2020 Labs and Workshops, Notebook Papers. CEUR-WS.org (Sep 2020)

[11] Buda,J.,Bolonyai,F.:An Ensemble Model Using N-grams and Statistical Features to Identify Fake News Spreaders on Twitter. In: Cappellato, L., Eickhoff, C., Ferro, N., N ́ev ́eol, A. (eds.) CLEF 2020 Labs and Workshops, Notebook Papers. CEUR- WS.org (Sep 2020

[12] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. SIGKDD Explor. Newsl. 19, 1 (June 2017), 22–36. DOI:https://doi.org/10.1145/3137597.3137600

[13] Yang, S., Shu, K., Wang, S., Gu, R., Wu, F. and Liu, H. 2019. Unsupervised Fake News Detection on Social Media: A Generative Approach. Proceedings of the AAAI Conference on Artificial Intelligence. 33, 01 (Jul. 2019), 5644-5651. DOI:https://doi.org/10.1609/aaai.v33i01.33015644.

[14] Sebastian Tschiatschek, Adish Singla, Manuel Gomez Rodriguez, Arpit Merchant, and Andreas Krause. 2018. Fake News Detection in Social Networks via Crowd Signals. In Companion Proceedings of the The Web Conference 2018 (WWW '18). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 517–524. DOI:https://doi.org/10.1145/3184558.3188722

[15] https://arxiv.org/abs/1704.07506

[16] W. Y. Wang, Liar liar pants on fire: A new benchmark dataset for fake news detection (2017)