Final Report

Done By:

Lo Tai Yi Linus (3A3/19) [Leader]

Jonathan Ho Cheng En (3A3/12)

Liu Yip Kiu Ricky (3A3/18)

Ma Yi Fan (3A3/21)

Hwa Chong Institution (High School)

2021

Content Overview

1. Abstract

Project Athena aims to research the use of machine learning to help tackle a pressing societal issue – fake news. With misinformation becoming more and more ubiquitous, we view it as a pertinent issue. Artificial Intelligence (AI) or Machine Learning (ML) is a promising solution to this problem due to its ability to trawl through copious amounts of information and analyse them without fail. Project Athena also looks into the possible implementation of these ML models in a convenient user-friendly platform such as a website of Google Chrome extensions. However, due to Google's imminent clamp down on third party servers through their new Manifest V3,we have decided to scrap the idea of an extension and simply make a website and the model instead. The project was able to develop two models with the first having an accuracy of 96% while the second with a decent 89% on two and four classes respectively. Furthermore, web hosting such a ML model is possible using applications Google Cloud or WordPress. The project has used XAMPP as a show of concept to how it works. The use of ML and the convenience of web hosting has led us to conclude that this area can be further developed to a mature system that is able to quickly detect misinformation online.

2. Introduction

2.1. Fake news in the 21st Century

With the advent of the digital age, we are seeing more and more users getting news and information from the web. While such a trend brings about greater convenience, there is a flip side, that is the spread of misinformation or "fake news". In fact, a survey done by NCID shows that 6 in 10 Singaporeans have received fake news via social media regarding the coronavirus, and an unofficial survey conducted by journolink showed that 45% of Britians believe they encounter fake news everyday (The number is likely even higher since those who believe in fake news will not think that the misinformation they read is "fake"). With a click of a button, these pieces of fake news can spread to millions, thus the need to keep them in check. While many reputable news outlets still publish real and verified news, there is a huge increase in the number of unreliable news sites that have popped up in recent years, as evidenced above. How can we combat this problem?

What fake news is

Fake news, in dictionary definition, are: "Sources that entirely fabricate information, disseminate disinformation and deceptive content, or grossly distort actual news reports." This covers conspiracy theories and misinformation. We also plan to be able to detect political and agenda driven news, all of which are not entirely fake or misinformation but should be approached with caution.

2.2. Rationale & Objectives

The aims of this research are to **(a)** curb the spread of fake news, conspiracy theories and misinformation online through a google plugin or a website, **(b)** inform the general public of the reliability of contents within each site, and **(c)** provide alternative, reliable and fact-checked information from other news sites.

2.3. Target Audience

Our Google Chrome extension will be available to anyone who uses the chrome browser and hence is targeted at any individual, regardless of age, gender, or race. As a result, the extension will be designed such that it caters to everyone, especially the older generation who may find it difficult to use. However, due to technical and time issues, we were unable to finish the extension. Instead, our project will now be available as a ML model that can be run on a python script.

2.4. Scope

Fake news is a very wide topic to cover. Thus, this project will only focus on a few key aspects of fake news and how to combat it, such as from**(a)** online news articles and, **(b)** conspiracy theories spreading online. Fake images and videos using Deep Fake as well as news on social media will not be covered due to it being out of the scope of our project.

2.5. Timeline

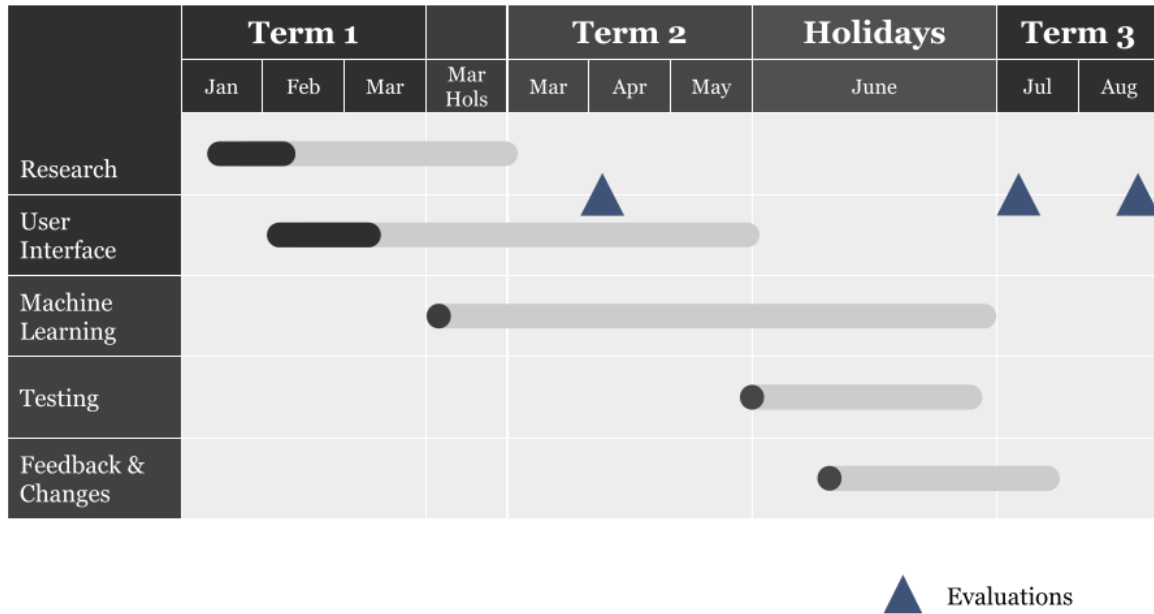Attached below is our timeline set at the beginning of the year.

Fig.1. Timeline for Project Athena.

3. Literature Review & Case Studies

3.1. Case Study 1: Donald Trump

President Donald J. Trump, the 45th president of the United States of America, has made over 15 000 lies during his four years in office (Glenn Kessler, 2020). His actions and false remarks have resulted in violence and unrest around the country such as the recent Capitol riot which killed a police officer and other members of the public (Ali & Zengerle, 2021). The power of celebrities and influencers in projecting fake news is a worrying trend that is expected to continue as technology continues to develop in the 21st Century.

3.2. Case Study 2: Arab Spring

The Arab Spring demonstrates the power and influence social media and online news sites have over people around the world. The Arab Spring in 2010 consisted of multiple movements and protests for freedom in the Middle Eastern world (History.com Editors, 2018). Social media was one of the key factors of the Arab Spring with information being spread quickly through the

Internet. A study argued that the media coupled with the right political environment could spark protests and movements (Wolfsfeld, Segev, & Sheafer, 2013). The current coronavirus pandemic and the rise in nationalism may lead to the proliferation in political stability and misinformation, leading to undesirable impacts for society.

3.3. Case Study 3: RGS Girl Develops AI-based Fake News Detector

An RGS girl has recently won a gold award at the Singapore Science and Engineering Fair for her fake news detection machine learning model, which has achieved accuracies of more >90%, which is rather remarkable (Kwang, 2019). However, we believe that we can further improve on her methodology, as well as put our model to practical use instead of it only being able to work as a script. (Our project was not inspired by her, rather we found out about it after registration).

3.4. Use of artificial intelligence

Artificial Intelligence (AI) and Machine Learning (ML) is becoming very popular. Over the recent years, research in the field of fake news detection has been intense. Current machine learning algorithms(both deep learning and classic) are able to achieve results near 95% accuracy when categorizing news in the categories ture of false. One example of the use of the modern deep learning algorithms in this field is Fabula.ai. Recently acquired by twitter, it takes into account both the content of the news and the features extracted from the social networks, achieving results of 0.93 AUC. However, using only the content of the news , a convolutional neural network based technique is proposed to detect fake news using the titles and the heading image, obtaining results of 92% accuracy.

3.5. Current fake news detection extensions

To have a better understanding of current efforts to combat fake news, a grading was conducted on Google Chrome extensions currently available. The extensions are graded based on their AIs accuracy in tone, facts as well as overall user experience (UX). A '✓' is given when the extension has a given feature, and will then be further graded with a maximum number of ticks 5 ticks to denote how good the feature is. A '×', on the other hand, means that the extension does not have such a feature. A '$' denotes that the extension requires a monthl subscription. Table.1. shows the collated findings.

Other Extensions	Accuracy of Tone	Accuracy of Facts	User Experience
FakerFact	✓✓✓✓	×	✓✓✓
Fake News Detector	✓	×	✓
Centr: Political Bias Detector	✓	×	✓✓✓
Fake news detector	×	×	×
Fake news debunker	×	×	×
Surfsafe	×	×	✓✓✓
NewsGuard	$	$	$

Table.1. Collated ratings for other Google Chrome extensions that target fake news.

As seen in Table.1., many of the fake news detection extensions available do not have good user interface (UI) or UX. Furthermore, it should be noted that a majority of these extensions are unable to determine whether a news article is fake or not and can only detect its tone. These extensions assume that if an article is agenda driven and non-journalistic, it means that the article is fake. This works decently well, but could be improved in combination with a real or fake news detector and external factors such as site reliability and providing similar news articles since large and reputable news outlets typically appear as top search results. It is to note that FakerFact is also no longer available as of July for unknown reasons, showing the scarcity of good online fact checkers.

4. Fake news detector

4.1 Proposed method

The usage of ML in natural language processing is not unprecedented and has achieved remarkable results. Due to the complexity of Natural Language Processing (NLP), Artificial Neural Networks (ANN) are often used to extract the high- level features from text data, and thus allow one to obtain useful information from the text. In this project, we will be using Long Short-Term Memory (LSTMs) to detect fake news. The model receives the title and the content of the article as the input for the model to make its predictions. In addition, this initial model will be supplemented with another model, going into more details about what type of fake news it is.

4.2 Data preprocessing and cleaning

We used two datasets to train our models. For the initial training, we used the ISOT fake news dataset. Using this dataset, models were very accurate in predicting true articles. However, we found out that it had trouble discerning opinion articles from fake news articles. Thus we fine tuned the model on a small part of the Fake News Corpus dataset of over 30 million pieces of dataset. This is to make sure that our model did not overfit on that particular clean dataset.

We then trained another, separate model on the fake news corpus. We reduced the categories into 4 more distinct categories- real, fake, unreliable and political. Due to the unnecessary size of the dataset, we only used 50000 pieces of news from each category to train our model. The dataset is not as clean as the first one but it is good enough for practical purposes.

The article is received as object type data. Firstly, we would remove stopwords. This means that we remove unnecessary words such as "the, are, " et cetera. All the words are then appended to a single list containing all the words in the dataset.The model is still unable to read the words as individual strings. Thus, these words need to be converted into numbers. A tokenizer represents these words as numbers and is first fitted onto the list of the words. It is then applied to the title and the content of the articles. Lastly, we would ensure that all of the data have the same length. We decided to have a maximum length of 700 for the content, and 15 for the article. We decided on these values as over 75% of our articles lie with this range of words. The articles with fewer

than 700 words are then padded to have the same length as the other articles (adding zeros at the back).

4.3 Architecture

As mentioned previously, we have two separate models with similar architectures, one of which is more accurate but can only tell between real and fake news whereas the other model can classify the news into two more categories.
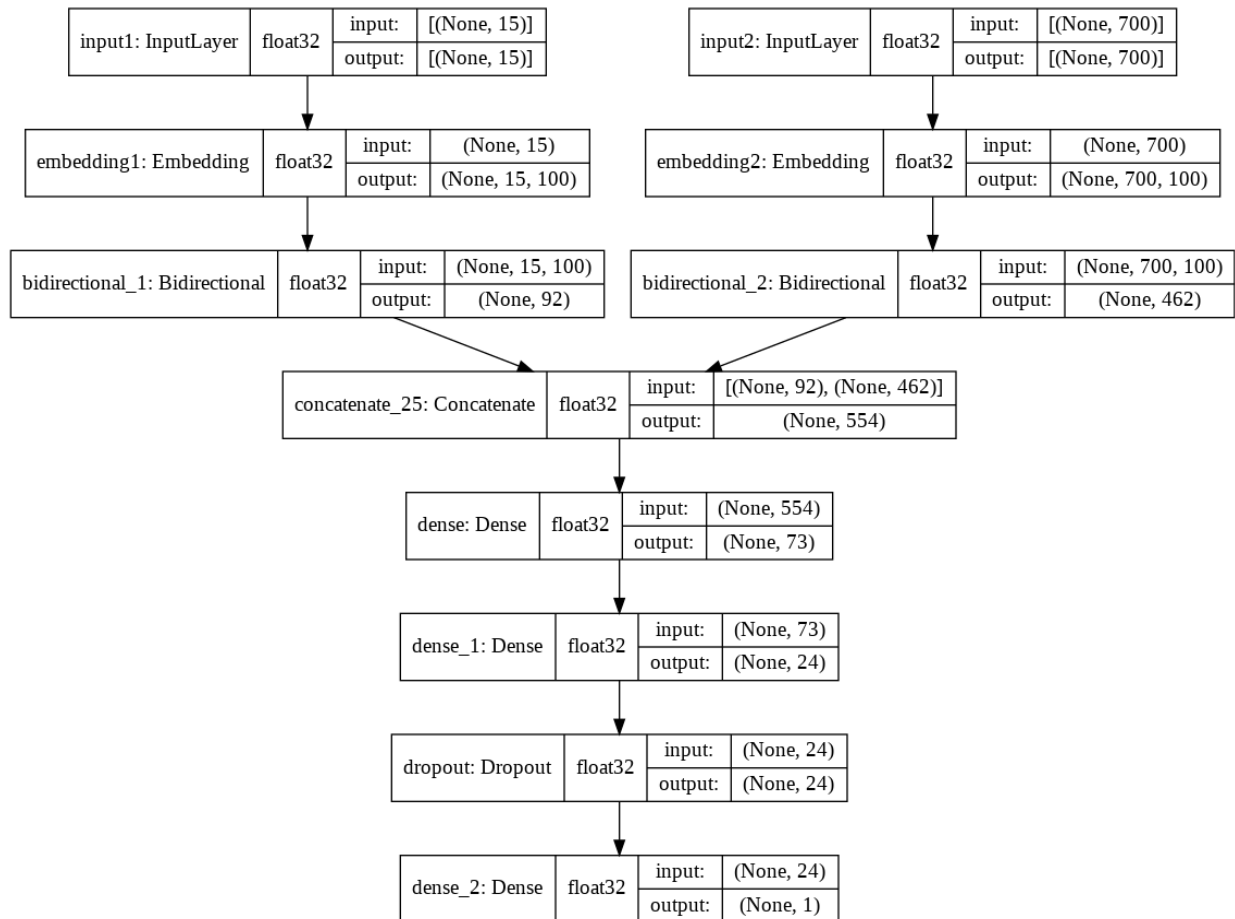

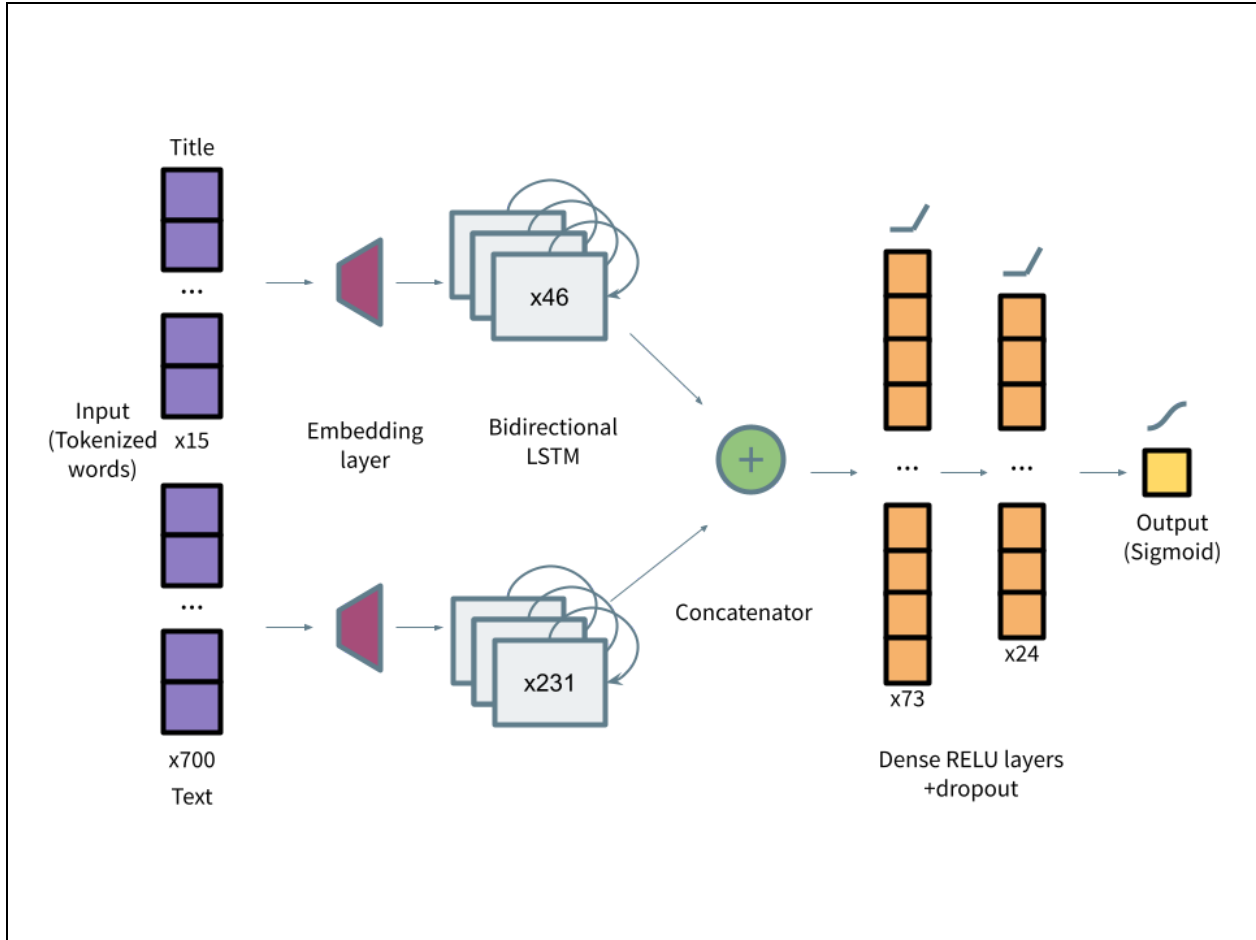**Fig.2. Diagram showing the architecture.**

10

Fig.3. Simplified version of architecture in Fig.2.

The detailed structure of our first model is shown in figure 1. The fake news detector is based on the structure proposed by Rodríguez, Á., & Iglesias, L. (2021). The architecture receives two inputs, the title and the content. In these individual branches the title and the content will be embedded and then run through a bidirectional LSTM layer. The output of the two branches will then be concatenated before going through 2 dense layers. Afterwards, dropout will be added before the final output layer. The final output layer has the sigmoid activation function to give us a value between one and zero to tell us if it is fake or real.
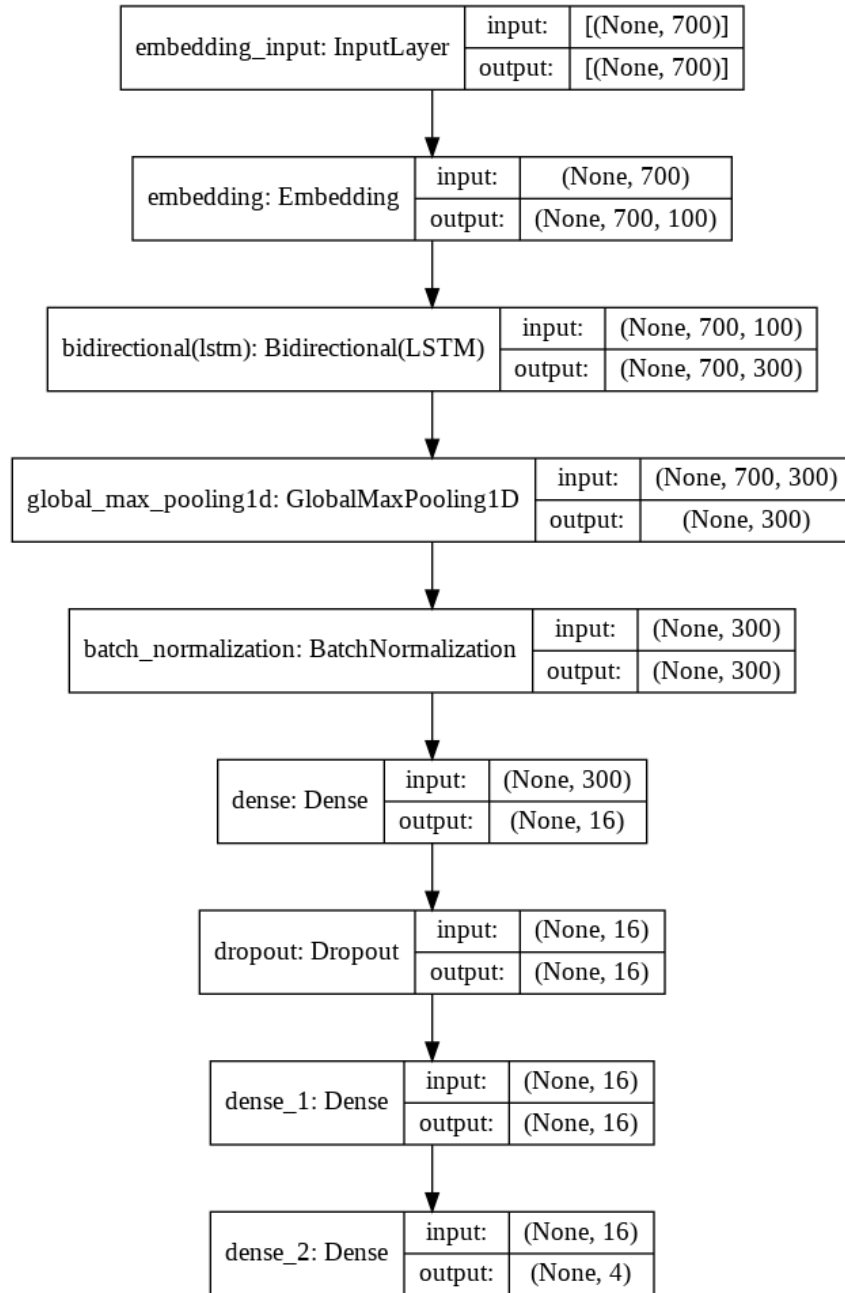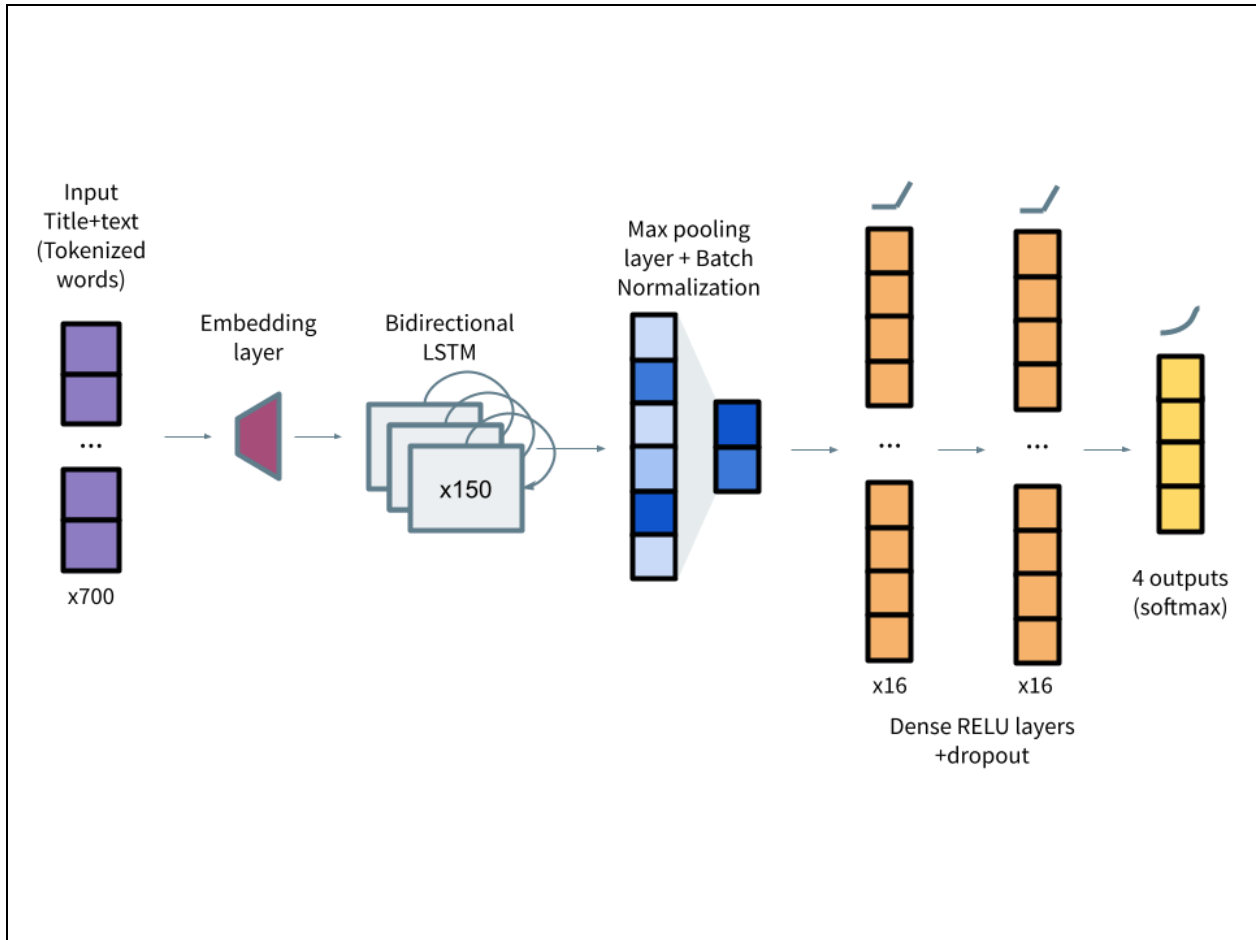
Fig.4. Diagram showing the architecture.

Fig.5. Simplified version of architecture in Fig.4.

The detailed structure of our second model is shown in figure 2. For this task, we tried our own architecture which turned out to be quite successful. The text and the title are first concatenated, and the first 700 words of this are taken and fed into the model. The tokenized words are then embedded using an embedding layer, which turns it into a 100 dimensional representation of the words. This is then fed into a bidirectional LSTM, which transforms a regular LSTM by doubling it and passes both the sequence and reverse of the sequence to ensure that it can catch any word order, no matter what the sequence or order it is in. The LSTM returns a 300 by 700 matrix, which is fed into a max pool layer that highlights important information the LSTM has picked out. After the result is batch normalized, it is passed into two fully connected layers of 16 RELU cells with dropout, and outputs a probability of each class using the softmax function.

4.4 Results

The first model achieved a validation accuracy of approximately 96% while the second had a decent accuracy of 89%. The confusion matrix of both the first and second second model are shown here respectively.
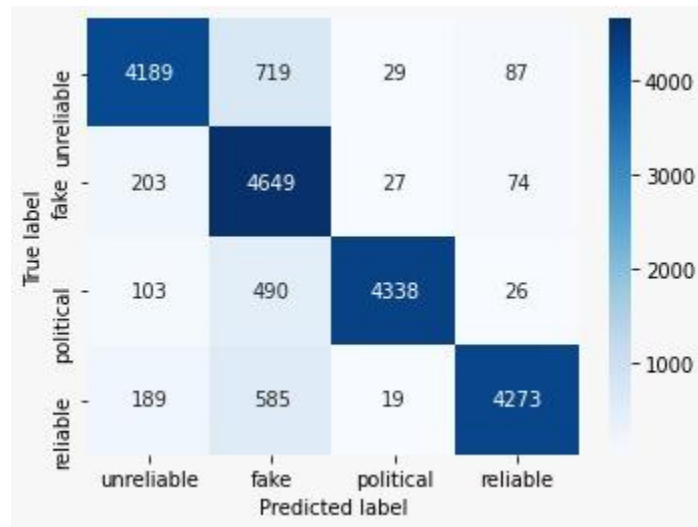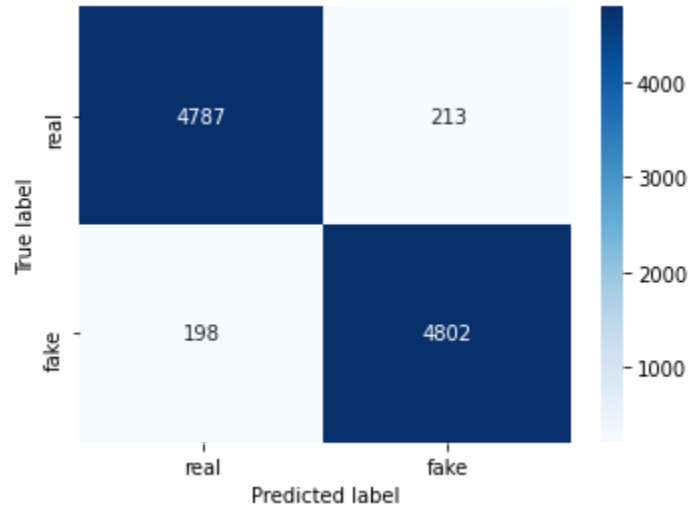

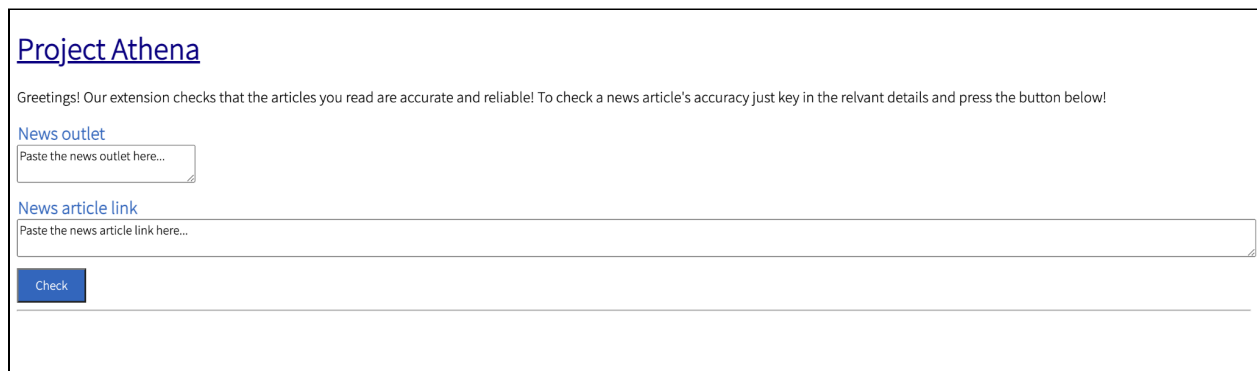**Fig.6. & Fig.7. Confusion matrices.**

5. Website

Although stated earlier in the year that this project wishes to implement its ML model into a Google Chrome extension to provide ease and convenience to the user, this has become difficult due to Google's introduction of Manifest V3 which prevents remotely hosted code (Chrome Developers, n.d.). This posed a problem as the ML models mentioned above were coded using Python and not TensorFlow JS. So, we decided that a website could provide similar functionality.

5.1. Creating the website

The website was created from scratch using basic HTML, CSS, and JavaScript. The Google Sites below shows the progress in the development of this website. View the Google Sites at https://sites.google.com/student.hci.edu.sg/creating-project-athena. The full HTML code can be found in the annexes.

5.2. Features of the website

The website features an easy to use fill-in page where users are required to fill in the news outlet and add the link of the news article they are reading.

Project Athena

Greetings! Our extension checks that the articles you read are accurate and reliable! To check a news article's accuracy just key in the relvant details and press the button below!

News outlet

Paste the news outlet here...

News article link

Paste the news article link here...

Check

Fig.8. Shows the fill-inpage.

Next, they can press the "Click" button. This will reveal Athena's verdict where the results of the ML model will be displayed in an easy-to-read manner.

15

Fig.9. and Fig.10. Shows what the user can expect after he or she presses the "Click" button.

6. Setup of server

However, just creating the website is not sufficient. Our project requires a place to store not only the ML model but also the contents of the news article to be processed. Thus, this requires the setting up of a server.

6.1. Introduction to XAMPP

XAMPP is an easy to use Apache distribution that contains MariaDB, PHP, and Perl. We installed XAMPP in our computers and were aiming to use our computers to run the servers (Apache Friends, n.d.)

6.2. PHP cURL

As part of our target to make our website as user friendly as possible, we decided to simplify the user experience by requiring them to key in only the news outlet from which their news article came from and the news article's link. Thus, it would require running a PHP file with the PHP cURL to extract the contents for our ML model to process. An example code can be found in the annex. After extracting the source of the website and further extracting on its contents, usually the <p> elements, the news article content can then be stored in a .txt file for the ML model to use.

6.3. Demonstration of concept

Another benefit of using a server is that it enables us to hold and run different files (.html, .php, .py, .txt, etc) all in one place. Below is how we can host a website using XAMPP.
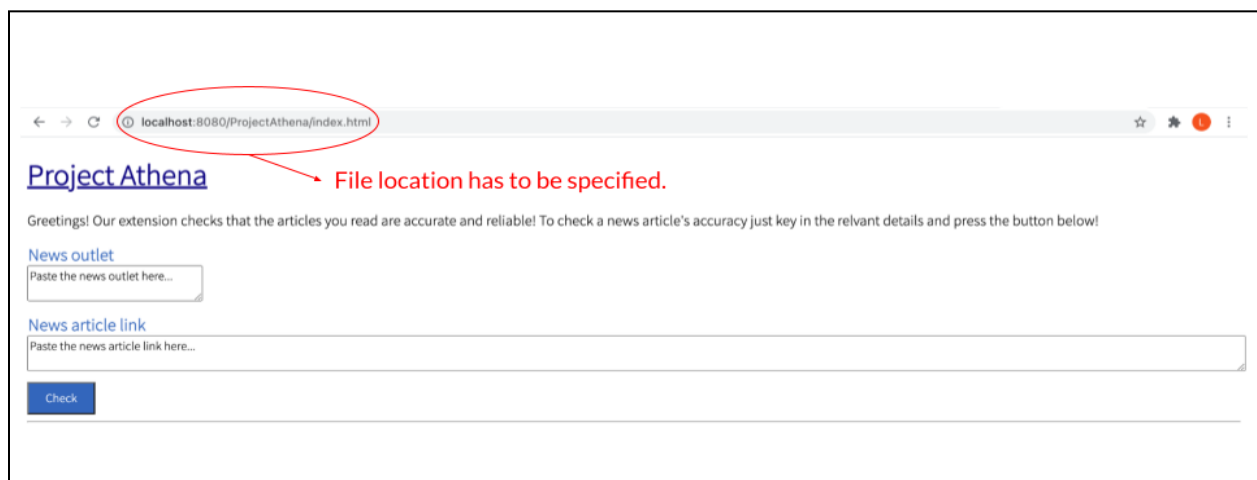
Fig.11. Shows the website being hosted using XAMPP. Notice the file location.

After setting up the website we could then add the other necessary files to be run. This has to be done correctly by naming the correct location of the files in the XAMPP application. For example, to run our ML model we have to call it from /ProjectAthena/MLmodel/model.py. Due to time constraints and unforeseen circumstances, we were unable to fully set up this server. However, we could show the HTML file to the public by using our IP address and allowing the computer to run. We wanted to use Google Sites as it is free to use and easy to navigate.

7. Conclusion, Areas of Improvement & Future Recommendations

As described above, the models produced to detect fact news were both efficient and accurate. Besides that, it has an added bonus of categorising different types of news into multiple categories. This shows that it is highly plausible for tech giants such as Google, Facebook, and Twitter to implement such models to help combat false information. In fact, some of these tech giants have already begun testing and trying out machine learning to help tackle the rampant spread of online fake news. They include Google's newly unveiled About this Result search feature which allows users to compare a site with what other sites say about it. Furthermore, government and non-governmental organisations (NGOs) could also provide similar if not better use of machine learning models. This could be in the form of policing sensitive societal issues in their country or supplementing similar use by tech giants to providing free and impartial services on an array of topics in the case of NGOs. We also believe that there is room for improvement for Project Athena in particular its UX could be further improved with the use of more Bootstrap.

8. Reflection

Jonathan Ho Cheng En

This project was an enlightening experience as it was my first infocomm project. At first, it was extremely difficult as I had to grasp many new concepts such as machine learning. Most of my time during this project was spent learning all these new concepts and programming libraries. Even though it was tough, in the end I still enjoyed the whole process of learning these new skills. It was this project which sparked my interest in the field of A.I. Outside of this project, I went on to learn other fields of A.I. such as computer vision, generative adversarial networks and even reinforcement learning. Overall, this project was an eye opening experience to the world of A.I. and the possibilities of future infocomm projects. While we may not have got the full implementation down, I look forward to carrying on with a more complex infocomm project next year with my new found knowledge.

Liu Yip Kiu Ricky

This project was a very enriching experience for me as this was one of my first serious programming projects. The steep learning curve when it came to learning the language (python) and the required libraries needed to do the Machine Learning portion was not an easy task to do. This project has not only increased my technical knowledge, but has also taught me on how to find resources online to teach myself and further increase my knowledge, and has given me a profound interest in AI and machine learning, which has encouraged me to learn other ML architectures outside of the scope of our project. This project has also greatly increased my knowledge of python and programming skills, which can be further used in future projects and other jobs.

Lo Tai Yi Linus

I found this project work highly enriching as it enabled me to hone my web programming skills and put it to practical use. I now realise that creating a functional website is much harder than it seems as there are nuances such as the appropriate use of classes and identity. Furthermore, I have learnt how to set up a server and its various components such MySQL, and how to navigate it. However, I realise that the time issue is something that I could improve on as I have to balance my project work with numerous other responsibilities. Through the project work, I have truly understood the importance of time management and have been able to pick up habits that have saved a lot of time. Overall, I have learnt a lot during this project.

Ma Yi Fan

This project was an enriching and interesting experience for me as it gave me an opportunity to explore new things in programming and being able to put them into practical use. I spend a lot of time on learning programming languages and various new concepts such as the required libraries required to do Machine Learning and AI programs. It was a fun experience and this project has not only expanded my technical knowledge on programming but also sparked my interest in AI and machine learning, which has encouraged me to explore further outside of the scope of our project, such as learning to develop a web page and server functions. This project has also taught me the importance of maintaining good communication between group members, time management and having the responsibility to finish your task on time. I look forward to working on more infocomm projects in the future with my increased knowledge on programming. I believe the skills I have picked up will also be very useful for future jobs in our advancing society.

9. Bibliography & Citations

Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake news detection using machine learning ensemble methods. *Complexity*, *2020*, 1–11.

Ali, I., & Zengerle, P. (2021, April 02). Police officer killed in Vehicle attack on U.S. Capitol. Retrieved April 06, 2021, from https://www.reuters.com/article/us-usa-capitol-security-idUSKBN2BP1BD

Apache friends. Apache Friends RSS. (n.d.). https://www.apachefriends.org/index.html

Bhatia, R., Bhatia, R., & Bhatia, R. (2021). Why Convolutional Neural Networks Are The Go-To Models In DL. Retrieved 5 April 2021, from https://analyticsindiamag.com/why-convolutional-neural-networks-are-the-go-to-models-in-deep-learning/#:~:text=The%20main%20advantage%20of%20CNN,for%20each%20class%20by%20itself.

FreeCodeCamp.org. (2018, January 21). How to create and publish a Chrome extension in 20 minutes. Retrieved April 02, 2021, from https://www.freecodecamp.org/news/how-to-create-and-publish-a-chrome-extension-in-20-minutes-6dc8395d7153/

Glenn Kessler, S. (2020, January 17). Analysis | President Trump has MADE 15,413 false or misleading claims over 1,055 DAYS. Retrieved April 04, 2021, from https://www.washingtonpost.com/politics/2019/12/16/president-trump-has-made-false-or-misleading-claims-over-days/

Google Chrome. (n.d.). Extensions. Retrieved April 02, 2021, from https://developer.chrome.com/docs/extensions/

History.com Editors. (2018, January 10). Arab spring. Retrieved April 06, 2021, from https://www.history.com/topics/middle-east/arab-spring

Journolink.com. 2021. *Fake News Statistics - How Big is the Problem? - JournoLink*. [online] Available at: https://journolink.com/resource/319-fake-news-statistics-2019-uk-worldwide-data [Accessed 28 March 2021].

Kwang, K. (2019, April 18). Liu Haohui's fake news detector prototype won her the Gold award at the Singapore Science and Engineering Fair 2019, and the right to represent Singapore at the Intel International Science and Engineering Fair next month. Retrieved April 5, 2021, from CNA website: https://www.channelnewsasia.com/news/technology/rgs-student-develops-award-winning-ai-fake-news-detector-11458402

Liu, S. (2021, January 20). *Internet users worldwide by Browser 2014-2020*. https://www.statista.com/statistics/543218/worldwide-internet-users-by-browser/#:~:text=In%202020%2C%20there%20were%20an,Safari%20with%20446%20million%20users

Mahabub, A. (2020). A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers. *SN Applied Sciences*, *2*(4). doi: 10.1007/s42452-020-2326-y

Migrating to manifest v3. Chrome Developers. (n.d.). https://developer.chrome.com/docs/extensions/mv3/intro/mv3-migration/

Min, C. (2020, May 21). 6 in 10 people in Singapore have received FAKE COVID-19 news, likely on social media: Survey. Retrieved March 28, 2021, from https://www.channelnewsasia.com/news/singapore/fake-covid-19-news-study-ncid-messaging-platforms-whatsapp-12756084

Natural Language Toolkit — NLTK 3.5 documentation. (n.d.). Retrieved April 5, 2021, from Nltk.org website: https://www.nltk.org/

Rodríguez, Á., & Iglesias, L. (2021). Fake news detection using Deep Learning. Retrieved 31 March 2021, from https://arxiv.org/abs/1910.03496

Wolfsfeld, G., Segev, E., & Sheafer, T. (2013). Social media and the Arab Spring. *The International Journal of Press/Politics, 18*(2), 115-137. doi:10.1177/1940161212471716

Fake-News Cleaning+Word2Vec+LSTM (99% Accuracy). (2021). Retrieved 10 August 2021, from https://www.kaggle.com/atishadhikari/fake-news-cleaning-word2vec-lstm-99-accuracy/notebook

- END -

10. Annexes

10.1. Source Codes

📄 HTML Source Code

📄 PHP cURL