Project Specifications Report

VERITAS

Team Members:

- Hanzallah Azim Burney - 21701829 - CS
- Abdul Hamid Dabboussi - 21701076 - CS
- Mohamad Fakhouri - 21701546 - CS
- Sayed Abdullah Qutb - 21701024 - CS
- Hassan Raza - 21701811 - CS

Supervisor: Ercument Cicek

Innovation Expert:

1.0 Introduction

1.1 Description

Social media platforms such as Twitter, Facebook, and YouTube are seeing a rise in the spread of fake news and misinformation on their platforms. According to research, Facebook users engage with misinformation 70 million times per month on average [1]. The rise of fake news is problematic for societies. It has the potential to sway public opinion, promote conspiracy theories, and instill fear, thereby eroding confidence in public institutions, and in democracy.

Therefore, our project aims to combat the spread of political lies and misinformation in order to better inform the public about what politicians are saying. To this end, we will build either a YouTube wrapper website that would perform fake news labelling and fact checking on live-streamed political speeches. The main targets of this project are political speeches and presidential debates from the United States of America. In order to accomplish this task, multiple ML/NLP papers on fake news will be consulted. A fake news/fact checking classification model will be built and trained on an appropriate dataset. Once a good enough classifier is achieved based on the metrics deemed ideal for such a problem, a platform would be built to use this classifier on political speeches streamed on YouTube. The model would then inform the viewer of any wrong claims being said by the speaker through various tags and captions.

1.2 Constraints

1.2.1 Economic Constraints

- The libraries and APIs used will be open-source and therefore, free.
- Model training will be done on Google Colab, which provides 12 hour segments of free training on Google GPUs and TPUs [2].
- Heroku would be used to host the backend and is also free for our use case [3].

- Our machine learning model would take in text and not speech, as such, we would need to transcribe videos first. A pipeline with AWS Transcribe could be built. AWS Transcribe offers free 60 minute transcription and costs 0.006 USD per 10 seconds [4].
- The datasets used to train the model are provided by researchers and available for free.

1.2.2 Implementation Constraints

- The machine learning backend will be built using the Python language.
- The wrapper website will be built using JavaScript.
- In order to speed up development times, open-source libraries for Python and JavaScript (React JS) will be used.
- The labelling of fake-news will be done through transcribed captions for every video.

1.2.3 Social and Political Constraints

- Veritas will only be trained and tested on political speeches from the United States of America focusing on President Donald J. Trump (POTUS) specifically.

1.2.4 Language Constraints

- The machine learning models will be tested on political speeches in English only.
- The datasets used for training will be filtered for English only text.
- The captions will also be displayed only in English.

1.2.5 Data Constraints

- The political speeches will only be extracted from YouTube.
- Different datasets would have a different number of features and would need to be cleaned and unified.
- Politifact [5] is the primary source used for collecting the datasets.

1.2.6 Ethical Constraints

- Regardless of how high prediction accuracy is, Veritas would never claim to be able to 100% point out misinformation and should only be used as a guideline to point out possible misinformation that the user should fact-check again.
- Veritas would never label any speaker as a liar and only aims to be used as a tool to alert users to possible misinformation.

1.2.7 Time Constraints

- The website would play-back live streamed videos in *near* real-time. A few second delay would be utilized for processing the speech and displaying the classified captions.
- The reports and presentations within the scope of CS-491 will be delivered under the deadlines stated on the course web-page.
- The project will be conducted in iterative stages and will be ready with the promised features in late Spring 2020.

1.3 Professional and Ethical Issues

After conducting some research regarding the ethical issues of our platform, we came up with a few points noted below:

- Regarding the privacy of the videos being considered and tested under our platform, there are no issues because we are using public political speeches' videos which are uploaded in youtube and are open to the public. These videos will be embedded in our platform and will not be stored locally, in case a video is deleted by the uploader at some point. The transcription of the videos, which is done by an automated software, will be stored locally and accuracy tests are being done based upon that. We do not see any issues regarding the privacy of the video or its transcription.
- Another ethical issue which comes to mind is the result of the accuracy test. Once a speech is transcribed from the video and its factual test results are out,

this might influence the voter and might make him biased towards the speaker of the speech.

- Legal issues regarding availability of a video might arise, for instance some videos are not locally available in some countries and we are considering omitting those videos from our training set so that no legal issues arise from it.
- There is an issue relating to accountability of the platform. This issue occurs due to the accuracy of the transcription of the video. Since we are fact checking a video's transcribed speech, and if the transcription is not 100% accurate based on the speech of the video, miss-classification might occur and our used algorithms might get misguided and the results would be false and incorrect.
- Many different machine learning techniques and algorithms will be used in this platform and we will do our best to refer to them according to IEEE Code of Ethics [6]. In addition to the techniques and algorithms, we will utilize some of the ready-made libraries from different sources to help us not write some functions from scratch and these libraries will be credited in accordance to IEEE Code of Ethics rules and regulations [6].

2.0 Requirements

2.1 Functional Requirements

- The user would be able to copy the link of a political speech video or livestream on YouTube and paste it into our website to watch the video.
- Generate captions in real-time by transcribing the speaker's speech.
- Get feedback about possible misinformation in real-time.
- Have the ability to verify claims the website classifies as a lie with a click of a button that takes the user to a reliable site with the information.

2.2 Non-Functional Requirements

2.2.1 Accessibility

- The website and its constituent services should be accessible by anyone on the internet. The website will be in the English language.

2.2.2 Usability

- The website should have a simple and intuitive interface with a minimalistic design.
- The website should be stable without interruptions and responsive enough to support *near* real-time feedback (5 minutes).

2.2.3 Compatibility

- The website should be compatible with different browsers and should work on both mobile and computer.

2.2.4 Scalability

- The website should be able to scale and handle upwards of 1000 users at a given time.
- The backend should be scalable in the sense that it should be able to run prediction models on 50 videos at a given time.

2.2.5 Performance

- Performance should be good and results should be obtained within 5 minutes.
- Overall response time of a website must also be instantaneous (less than 1 second).

2.2.6 Extendability

- The website should be extendable so that in future it could accept videos from sources other than YouTube.

2.2.7 Accuracy

- The machine learning classifier's "accuracy" would be measured using Macro-F1, Micro-F1 and Accuracy metrics to ensure the reliability of the model.

2.2.8 Reliability

- The website should be reliable to users and by being available all the time especially during live-streamed presidential speeches where traffic would increase substantially.

2.2.9 Testing and Error-Handling

- It should be easily testable for any errors and bugs that might occur. After that writing of code should be such that any bugs found should be easily traceable and handled.

3.0 References

[1] "Fake News, Misinformation, & Fact-Checking | Ohio University MPA | Ohio University", Ohio University, 2020. [Online]. Available: https://onlinemasters.ohio.edu/masters-public-administration/guide-to-misinformation-and-fact-checking/. [Accessed: 07- Oct- 2020].

[2] Google, “Colaboratory,” Google. [Online]. Available: https://research.google.com/colaboratory/faq.html. [Accessed: 11-Oct-2020].

[3] Heroku, "Pricing," [Online]. Available: https://www.heroku.com/pricing. [Accessed 12 10 2020].

[4] Amazon Web Services, "Amazon Transcribe Pricing," [Online]. Available: https://aws.amazon.com/transcribe/pricing/?nc=sn&loc=3. [Accessed 12 10 2020].

[5] The Poynter Institute, "Politifact," [Online]. Available: https://www.politifact.com/. [Accessed 12 10 2020].

[6] IEEE, “IEEE Code of Ethics”. [Online]. Available: https://www.ieee.org/about/corporate/governance/p7-8.html. [Accessed: 11-Oct-2020]