Final Project Report: Deepfake Detection for Videos

### Team: Starks

### Date: February 7th, 2024

**Problem Statement:**

1. **Deepfake Menace**: In an era defined by technological advancement, the rise of deepfakes, powered by AI, poses a significant threat. Reports, like the World Economic Forum's Global Risks Report, emphasize AI-generated misinformation as a major global risk, surpassing concerns about climate change.
2. **Political Integrity at Stake**: Deepfake instances, including AI-crafted speeches by political leaders and manipulated videos impacting election integrity, highlight the profound impact on political narratives. With crucial global elections approaching, there's a pressing need to combat deepfakes' potential to manipulate public opinion.
3. 3. **Urgent Call for Deepfake Detection**: The escalating threat calls for urgent action in developing advanced deepfake detection mechanisms. Generative AI tools like ChatGPT raise concerns about the ease of manipulation, underscoring the necessity to safeguard information authenticity, protect democratic processes, and ensure trust and truth prevail.

**Technical Approach:**

We developed a website for Deepfake detection using deep learning techniques to accurately identify deepfakes in the uploaded videos.

Our system employs:

* **Model Training:** Built and trained a neural network for aggregating the probability from individual model to form a single probability score for the video being deepfake. The aggregate model was prepared on historic data. The results show an accuracy of more than 75% on the validation set. Currently experimenting with hyperparameter tuning to improve accuracy and explore other potential model architectures.
* **Visualization and Actionable Insights:** Developed interactive graphs and table displaying deepfake detection based on analysis.
* **Data Collection:**  Successfully acquired a Deepfake dataset containing adequate real and fake videos

**Results:**

* On paper, a commendable accuracy of 79% has been achieved, exceeding the project's initial aim of 75%.
* Users have the ability to upload videos and receive outputs in the form of easily interpretable graphs.
* The website currently functions with an intuitive user interface and user experience.

**Challenges and Future Work:**

* Clashes: Dependency clash, initially the aggregate function was build using TensorFlow, but it had clashes with the media pipe library with was used for media pre-processing.
* Dataset: Acquiring the well-balanced dataset. Had to come up with the augmentation techniques to solve the issue.
* New detectors: Creating a new detector without the help of open source and using them for detection.
* Accuracy: improving the accuracy with the help of more advanced detectors and methods.

**Conclusion:**

The project has attained a commendable accuracy of 79%, surpassing the initial target of 75%. Users are afforded the capability to upload videos and obtain outputs presented through easily interpretable graphs. The website presently operates with an intuitive user interface and experience.