MITS HACKATHON

God's Eye

Team Name: **Shield**
Problem Statement : **Microsoft**

Cyber Crime Prevention (MC01)

BY:

Surya Narayanan CS
Monish SR
Janani Sri R

PROBLEM STATEMENT

- Today we live in a “post-truth” era, where a piece of information or misinformation is utilized by malevolent actors to manipulate public opinion.

- DeepFakes involves videos, often obscene, in which a face can be swapped with someone else's using neural networks. **DeepFakes are a general public concern, thus it's important to develop methods to detect them**.

Idea Introduction

God's Eye

Tech Stack

1. GAN
2. AI/ML
3. Python

Resources Used

1. Deep Learning
2. Tensor Flow

Third party API used:

Deepware

Deepfake Detection-Outline

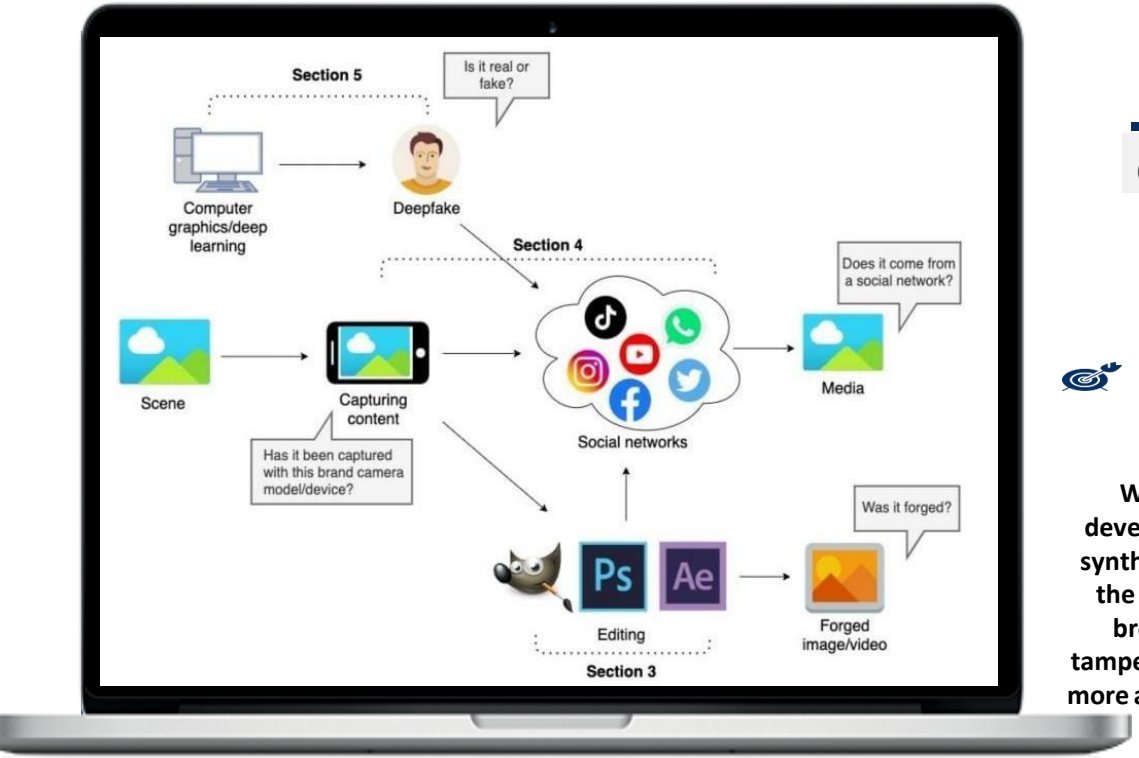
Problem Statement:

IDENTIFICATION OF ARTIFICIALLY GENERATED DEEPFAKE CONTENT.

Our Objective:

To build a solution that can automatically intimate the user if the probability of the content being deepfake is high.

Our Approach Towards Idea



Our Approach and our Innovation

Approach

With the rapid development of face synthesis technology, the security threat brought by face tampering is becoming more and more serious.

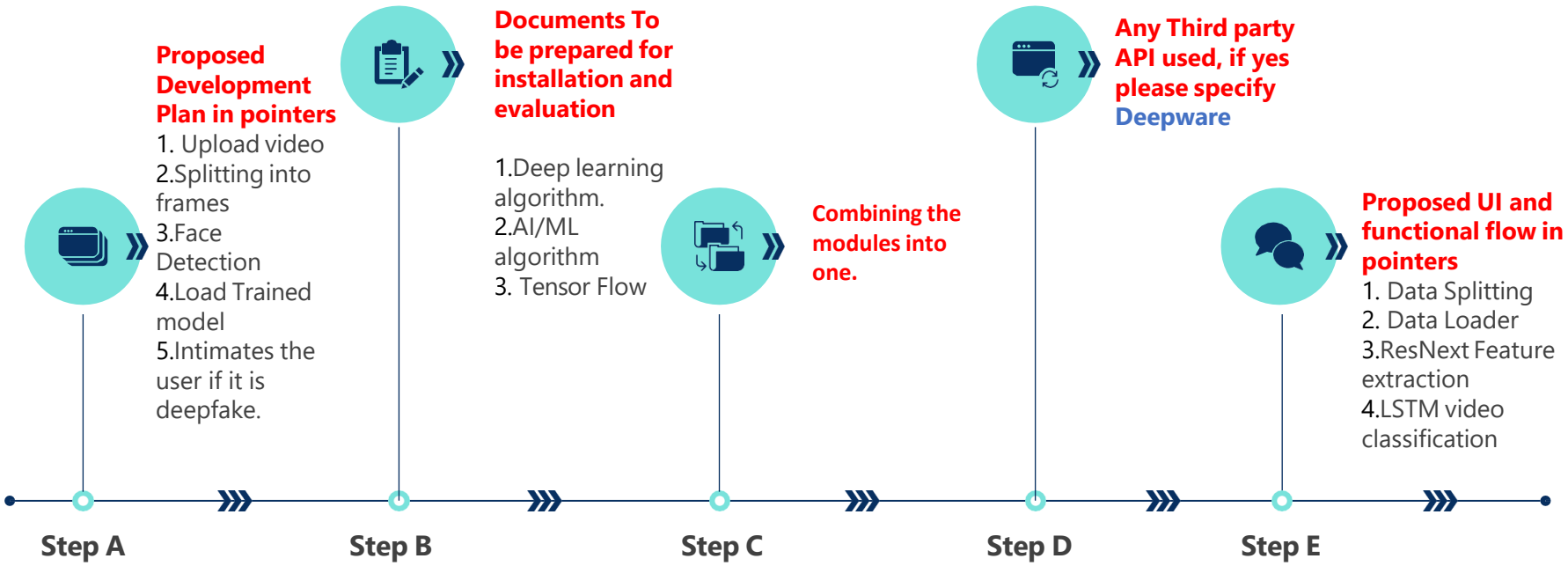
Our Innovation

Our fake detection algorithm will automatically intimate the users if the video is artificially generated.

Requirements

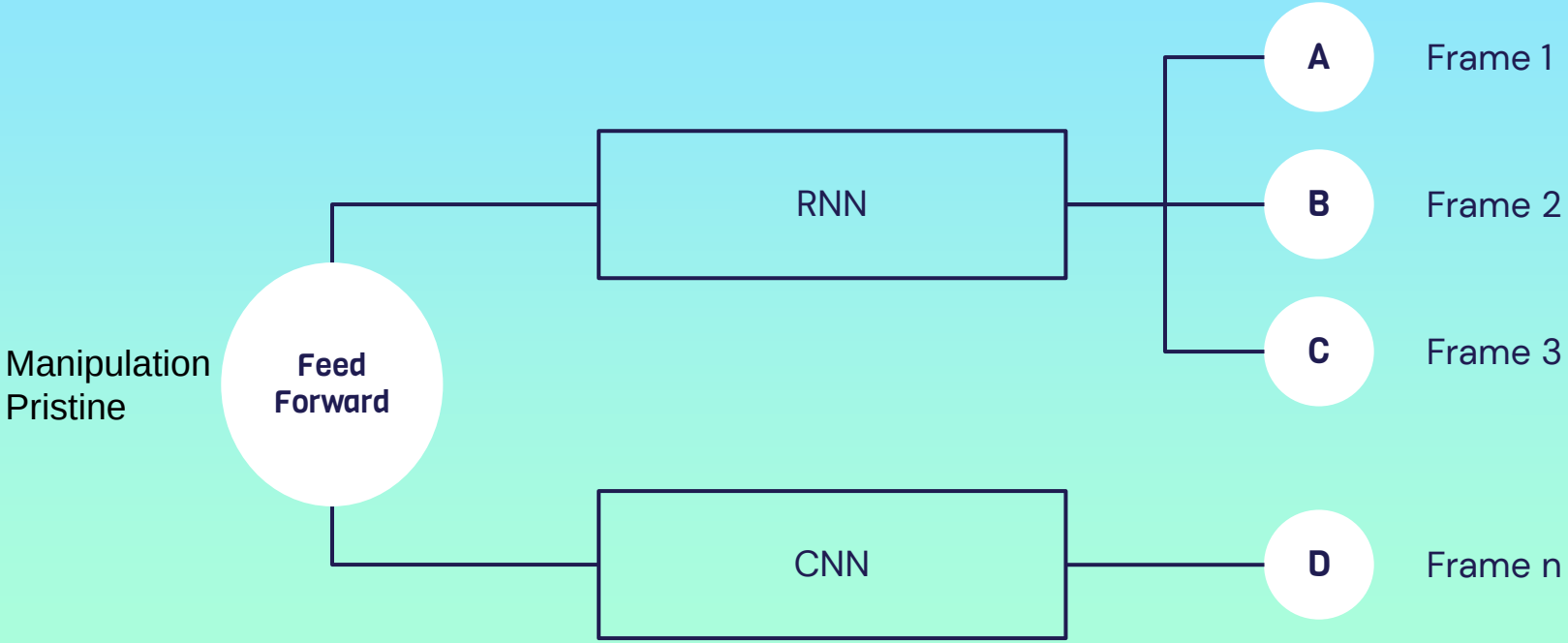
1. Tensor Flow
2. Deep Learning & AI/ML Algorithms
3. GAN
4. Deepware

Development Pipeline



ORGANIZATIONAL CHART

Exploiting the temporal dimension using recurrent neural networks

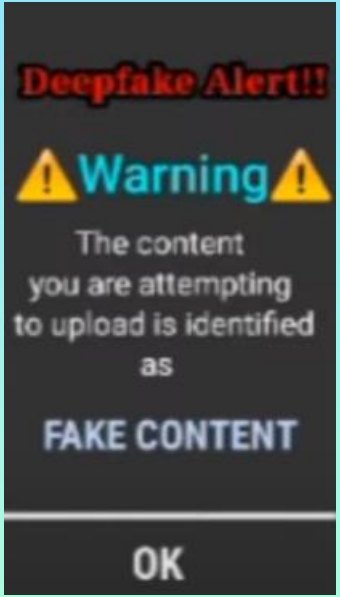


OUR PROCESS

- Once when the user uploads the content to any social media, our algorithm checks for any deepfake or hate content that we have given as a dataset.

- Once any deepfake content found, it sends a alert message to the user. It also sends a notification to the person who is being deepfaked in the content.

- The end user has the access to report the content and the user.



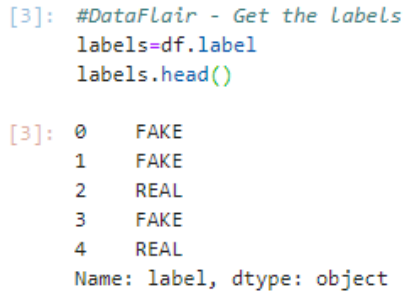
Detecting fake news with python

Using sklearn, we build a TfidfVectorizer on our dataset. Then, we initialize a Passive Aggressive Classifier and fit the model. In the end, the accuracy score and the confusion matrix tell us how well our model fares.

The necessary imports are:

```
1. import numpy as np
2. import pandas as pd
3. import itertools
4. from sklearn.model_selection import train_test_split
5. from sklearn.feature_extraction.text import TfidfVectorizer
6. from sklearn.linear_model import PassiveAggressiveClassifier
7. from sklearn.metrics import accuracy_score, confusion_matrix
```

Output screenshot:



```
[3]: #DataFlair - Get the labels
     labels=df.label
     labels.head()

[3]: 0    FAKE
     1    FAKE
     2    REAL
     3    FAKE
     4    REAL
     Name: label, dtype: object
```

We have included the code in github

Vision of Innovation/Idea/Solution

Today we live in a “post-truth” era, where a piece of information or disinformation is utilized by malevolent actors to manipulate public opinion. Disinformation is an active measure that has the potential to cause severe damage. Misinformation is defined as false or inaccurate information that is communicated, regardless of an intention to deceive. This makes us to develop this idea.

Innovation

Apart from detecting whether the content is fake or real, our detection algorithm provides a probability whether it is deep fake or not. If the probability of deepfake is high, it automatically intimates the user about it and restricts the user to upload the content.

Time Duration

Since we are using Deep Learning algorithm to detect the Deepfake , it doesn’t require lot of time to complete. It would approximately require a month to give the final product.

Initial Innovation

An innovation process always starts with the search for and finding innovative potentials and the derivation of ideas. First we thought of detecting the deepfake, while the development was going on we came to an innovative idea of intimating the user about their content using our algorithm which will be much more feasible for this problem statement.

Github link:

https://github.com/SURYASTIC/MITS-HACKATHON

We have uploaded the code in github. We have completed the project separately. Now we are in process to complete it. Soon we will complete.

Conclusion

We are looking forward to present this project in this hackathon.

Our instant process makes the user happy and improves satisfaction.