# Digital Authenticity Verification

An AI-Powered Solution for Deepfake Misinformation Identification.


## Balaji Kartheek


**Date : 28-01-2023**

## Abstract:

As humans, we pride ourselves on our intelligence and ability to understand and navigate the world around us. But In today's **digital age**, where we consume **vast** amounts of information **online**, As the use of **AI technology** continues to advance, the ability to create deepfake videos has become more accessible, which leads to an **increase** in the **spread of deepfake videos**. The proposed project will play a **vital** role in ensuring the authenticity of the **information** we consume on daily basis and will make an important contribution to the field of **deepfake detection**.

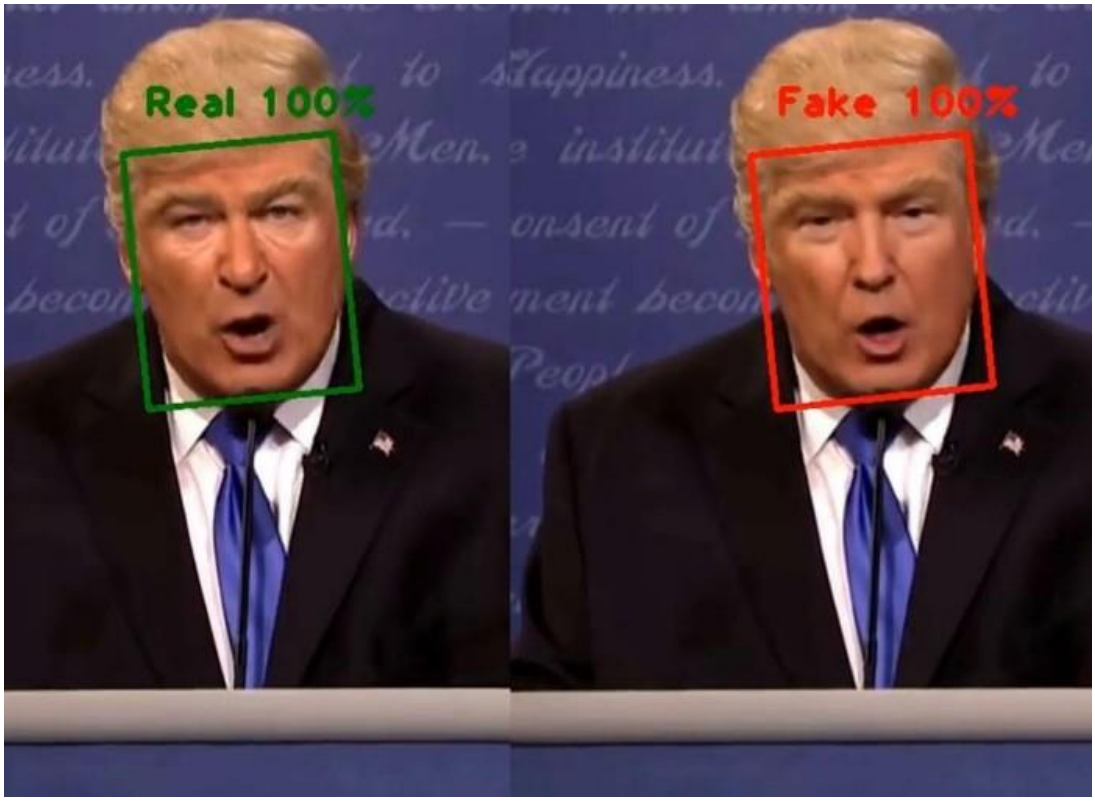
Deepfake videos, created using artificial intelligence and machine learning, have the **power to manipulate reality** and spread misinformation. With the ability to convincingly mimic real people and events, deepfake videos have the potential to cheat the public and disrupt the trust in our **digital communications**.

The solution will use a combination of computer vision, machine learning, and deep learning algorithms to detect and identify deepfake videos. It will also **analyze the motion of the videos** to detect any **inconsistencies** that may indicate that the video is not authentic. The solution will be trained on a large dataset of deepfake and real videos to ensure that it can accurately detect and identify deepfake videos. it will be implemented as a separate tool that can be integrated into various platforms, such as **social media, video conferencing,** and **news websites**, to ensure that the public is not **misled** by deepfake videos. It will also be made available to media companies, journalists, and other content creators to verify the **authenticity of videos and image files.**

Overall, the proposed model will make an important contribution to the field of deepfake detection by **developing an AI-powered solution** for deepfake misinformation **identification** that can be integrated into various platforms to **beat the spread** of deepfake videos.

## 1. Problem Statement:

The growing **prevalence** of deepfake technology has raised serious concerns about the authenticity and
**integrity** of digital information. With the ability to **manipulate** videos and images, deepfakes can be used to
spread misinformation and deceive the public. This can make people **not trust** digital communication and also have serious problems in various fields such as finance, security, and law enforcement. The problem is to find a solution to detect and **identify deepfake videos before they cause harm**, and **protect** the public from the spread of false information. The proposed project aims to develop an AI-powered deepfake detection system that can **analyze the visual characteristics** and **motion** of the videos to detect any inconsistencies that may indicate that the video is not authentic. This will help to **prevent the spread** of false information and ensure the integrity of the information in digital communication.



## 2. Market/Customer/Business Need Assessment:

### a. In the Media Industry:

In the media industry, deepfake detection technology can play a critical role in ensuring the **authenticity** of the information that is **being published**. It can help news organizations, media companies, and journalists verify the authenticity of videos and image files **before** they are published. This can **prevent the spread of false information and misinformation**, which can have serious consequences for the public and the media industry. It can also help them identify and **track the source of deepfakes**, which is important for investigative journalism. Additionally, the technology can be used to monitor and **track deepfake videos** on social media platforms. This can help media companies identify and report on deepfake videos that are being used to spread false information and misinformation. Overall, technology can help media companies improve the **accuracy and credibility** of their reporting, and help them maintain the **trust** of their audience.

### b. In Political Sphere:

In the political sphere, deepfake detection technology can play a critical role in ensuring the **integrity of democracy** by preventing the spread of **false information** and misinformation during elections and other political events. It can help political campaigns, political parties and election commissions detect and **prevent** deepfake videos and images that are being used to **manipulate public opinion**. The technology can also be used to track and trace the origin of deepfake videos and images, which can help authorities identify and hold accountable those who are **responsible** for creating and spreading them. Additionally, the technology can be used to monitor and track deepfake videos on social media platforms. This can help political campaigns and parties identify and **counter false information** and misinformation that is being **spread about their candidates** and policies. Overall, the technology can help political campaigns, parties, and election commissions protect the integrity of democracy by preventing the manipulation of public opinion and ensuring **fair** and **transparent** elections**.**

### c. In Social Media Platforms:

Deepfake detection technology is a powerful tool for social media platforms to ensure the authenticity of the information shared on their platforms. By implementing this technology, social media companies can **automatically detect and remove** deepfake videos and images which can be misleading and harm their

reputation. This technology can also be used to detect deepfake contents that are being used to harass or bully users, creating a safer environment for the users. By monitoring and tracking deepfake videos on social media platforms, social media companies can identify and remove deepfake videos that are being used to spread false information and misinformation. Additionally, this technology can also help social media companies to track and trace the origin of deepfake videos, which can help them identify and hold accountable those who are responsible for creating and spreading them. In summary, deepfake detection technology is a valuable tool for social media platforms to improve the accuracy and credibility of the information shared on their platforms and maintain the trust of their users.

### d. In the Security Industry:

The use of deepfake detection technology can be **vital** for the security industry to identify and prevent potential security threats. By implementing this technology, security agencies can automatically detect and remove deepfake videos and images that can be used to mislead and manipulate public opinion, particularly **during high-security** events or terrorist activities. This technology can also help security agencies track the origin of deepfake videos, enabling them to hold those responsible accountable. Additionally, by **monitoring** deepfake videos on social media platforms, security agencies can identify and remove deepfake videos that are spreading false information and **misinformation**. This technology can also be used by **border security agencies** to detect deepfake documents, which can be used by criminals or terrorists to cross borders. Intelligence agencies can also use this technology to detect and prevent the use of deepfake voices or videos to impersonate officials or dignitaries. By using this technology, security agencies can **improve** their surveillance and identification capabilities and thus prevent potential threats and maintain safety and security.

## 3. Target Specifications and Characterization

Our target customers are highly **concerned** with maintaining the integrity of the information they provide to their audience, and are willing to invest in technology that can help them achieve this goal. They have a **high level** of technical expertise and are comfortable incorporating new technologies into their **workflow**.

Our target market includes any organization or individual that **consumes** and/or creates **digital content,** such as social media platforms, news websites, video conferencing platforms, media companies, journalists, and content creators.

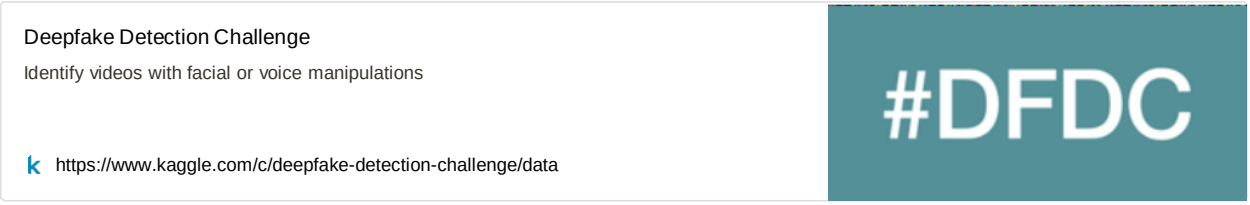
These organizations and individuals have a need to ensure the **authenticity of the digital content** they consume and/or create, in order to prevent the spread of false information and maintain the trust of their audience.

Our AI-powered deepfake detection solution will offer small businesses and organizations a **cost-effective** and easy-to-use tool to detect and identify deepfake videos and image files. This will help them **protect** their **reputation and credibility**, as well as **ensure the integrity of the information** they consume and/or create. Additionally, the solution will be able to **integrate** into various platforms and can be used by media companies, journalists, and other content creators to verify the authenticity of videos and image files.

# 4. External Search

**a. Dataset**

**Link:**




The DFDC (Deepfake Detection Challenge) is a dataset for Deep-Fake detection consisting of more than 100,000 videos.

The DFDC dataset consists of two versions:

- Preview dataset, with 5k videos. Featuring two facial
- modification algorithms. The full dataset, with 124k videos.

  Featuring eight facial modification algorithms

**A Sample of this Dataset is used for our model.**

## b. References/ Sources of Information

- InceptionV3 model Image Based on a Coupled Network

  InceptionV3.pdf
- EfficientNet Model

  efficientnet.pdf
- [Types of deepfake anti-spoofing software — Antispoofing Wiki] (https://antispoofing.org/Deepfake_Detection_Software:_Types_and_Practical_Application#:~:text=Deepfake detection technology allows prompt recognition of a,it helps to tackle the so-called liar's dividend.)

- Deepfakes Detection Techniques Using Deep Learning: A Survey (scirp.org)

## 5. Bench marking alternate products

There are a number of products and software currently in use for deepfake detection. Some examples include:

- Deepfake Detection Tool from Facebook: This tool uses machine learning to analyze videos and detect signs of manipulation. It is currently being used by Facebook to identify and remove deepfake videos from the platform.
- DALL-E from OpenAI: DALL-E is a deep learning-based image manipulation detection tool that uses a combination of computer vision and machine learning techniques to detect manipulated images.
- DeepTrace from Deeptrace: This software uses a combination of machine learning and computer vision techniques to detect deepfake videos in real-time.
- Deepfake Detection API from Truepic: This API uses computer vision and machine learning to detect deepfake videos, images and audio files.
- Deepfake Detection System from Deep Vision Technologies: This system uses deep learning and computer vision to detect deepfake videos in real-time.

## 6. Applicable Patents

1. Method and system for detecting manipulated video content - this patent covers a method and system for detecting and identifying manipulated video content using a combination of computer vision, machine learning, and deep learning techniques.
2. Artificial intelligence-based deepfake detection - this patent covers the use of AI and machine learning algorithms for deepfake detection, including image and video analysis, motion detection, and facial recognition.
3. Automated deepfake detection in real-time - this patent covers a system and method for detecting deepfake videos in real-time using computer vision and machine learning techniques.

## 7. Applicable Regulations

1. Data protection and privacy regulations: In order to comply with data protection and privacy regulations, the deepfake detection system must ensure that personal data is collected, stored, and **processed in accordance with applicable laws.** This may include obtaining consent from users before collecting their data, implementing **robust security** measures to **protect** personal data, and providing **clear and transparent** information about how data is collected, stored, and used.
2. Cybersecurity regulations: The deepfake detection system must comply with cybersecurity regulations to ensure that it is **protected against unauthorized access**, use, and disclosure of sensitive data. This may include implementing **strong authentication** and access controls, regularly monitoring for security threats, and **reporting any security incidents to the relevant authorities.**
3. Technical standards: The deepfake detection system must comply with relevant technical standards to ensure that it is **reliable, accurate, and secure**. This may include adhering to standards for data storage, data processing, and software development, as well as testing the system to ensure that it meets these standards.
4. Environmental regulations: The deepfake detection system must comply with relevant environmental regulations to ensure that it does not have a negative impact on the environment. This may include ensuring that the system is **energy-efficient** and that any waste generated during its operation is **disposed of in an environmentally-friendly manner.**

## 8. Applicable Constraints

1. **Space**: The Model will require a dedicated space for the development and testing of the deepfake detection system. This may include a **computer lab or data center** where the necessary hardware and software can be **installed**.

2. **Budget**: The Model will require a significant budget for the development and testing of the deepfake detection system. This may include costs for the hardware, software, and personnel required to develop and test the system.
3. **Expertise**: The Model will require a team of experts with knowledge and **experience in the field of computer vision, machine learning, and deep learning.** This may include engineers, data scientists, and computer vision experts who can **develop and test** the deepfake detection system.
4. **Data Collection and annotation**: The Model will require a **large dataset** of deepfake and real videos. Also it will require proper annotation of the data to **train the model**.
5. **Legal and Ethical Constraints**: The Model will need to comply with all relevant laws and regulations related to **data privacy and security**. Also it will require to consider ethical concerns related to the use of deepfake technology.

## 9. Business Model

1. **Subscription-based model:** Offer a monthly or yearly subscription for individuals and organizations to access the deepfake detection tool. This can include different levels of access and services **based on the subscription plan.**
2. **Partnership model:** Partner with other organizations, such as **fact-checking agencies** or news organizations, to offer the deepfake detection tool **as an added service**.
3. **Advertising model:** Integrate targeted advertising into the deepfake detection tool, where businesses can pay to advertise to users who are utilizing the tool.
4. **Consulting model:** Offer consulting services to organizations, such as media companies and government agencies, to assist them in identifying and **addressing deepfake-related issues**.
5. **Data-selling model:** Collect and analyze data on deepfake videos and images and sell insights to businesses and organizations for **market research** or other purposes.
6. **AI-as-a-Service:** Offer the deepfake detection tool as a service for other businesses and organizations to run on their own systems.

## 11. Concept Generation:

The proposed AI model for deepfake detection is a response to the growing concern about the authenticity and integrity of digital information. With the **increasing prevalence** of deepfake technology, it has become more important than ever to find a solution to **detect and identify** deepfake videos **before** they cause harm and spread false information. The concept of this model is to develop an AI-powered deepfake detection system that can analyze the **visual characteristics** and motion of videos to detect any inconsistencies that may indicate that the video is not authentic. This will help to prevent the spread of false information and ensure the integrity of the information in digital communication.

The concept of this model is to use the latest advancements in AI, machine learning, and computer vision to create a powerful deepfake detection system. This system will be trained on a large dataset of deepfake and real videos to ensure that it can accurately detect and identify deepfake videos. It will be **implemented** as a separate tool that can be integrated into various platforms, such as social media, video conferencing, and news websites, to ensure that the public is not misled by deepfake videos. Additionally, it will be made available to media companies, journalists, and other content creators to verify the authenticity of videos and images files.

Overall, the concept of this model is to use the power of AI and machine learning to fight against the spread of deepfake videos and misinformation, and ensure the integrity of digital communication. This model will have a strong focus on simplicity and ease of use, making it accessible to small and medium enterprises, as well as large enterprises.

## 12. Concept Development

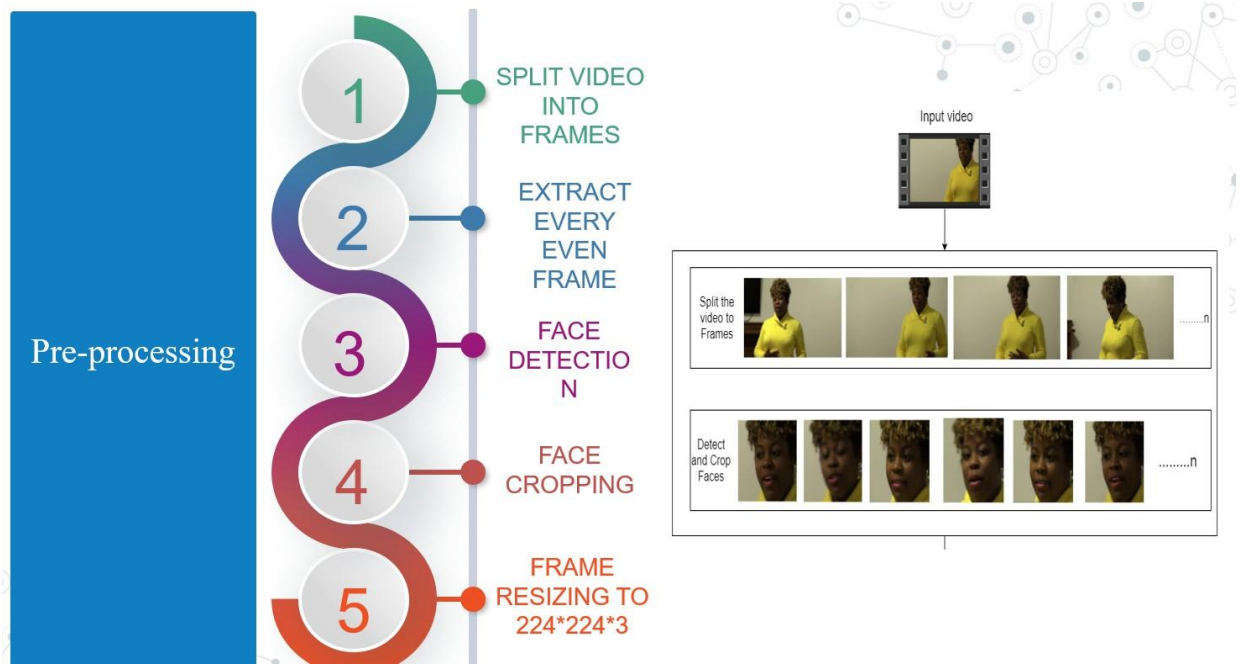
Once the dataset has been gathered, we will use computer vision and machine learning techniques to analyze the visual characteristics and motion of the videos. This will be done to detect any inconsistencies that may indicate that the video is not authentic. We will also use deep learning algorithms to detect and identify deepfake videos.

## 13. Final Product Prototype (abstract) with Schematic Diagram

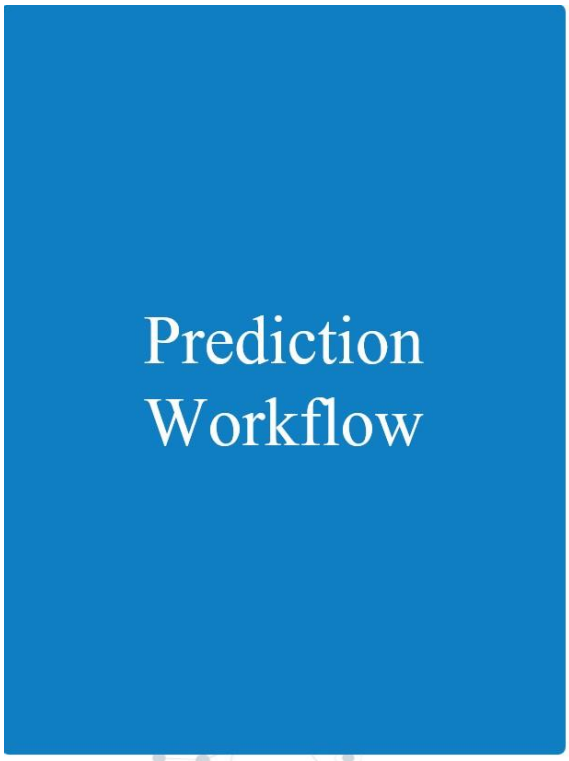
### Pipeline

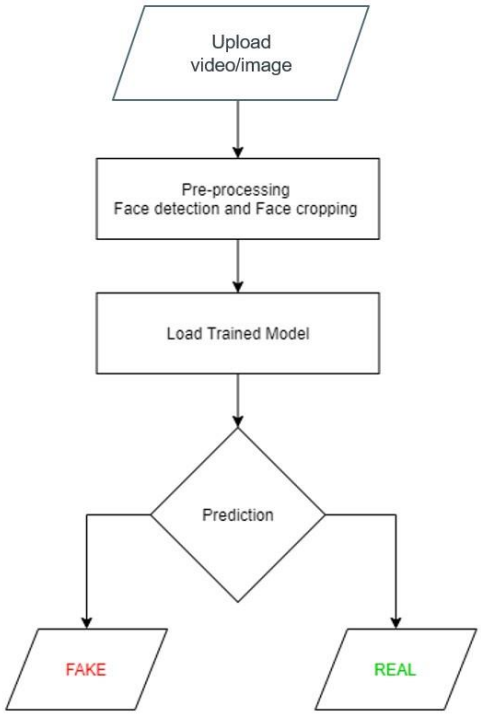
| | |
|---|---|
| step 1 | Loading the datasets |
| step 2 | Extracting videos from the dataset |
| step 3 | Extract all frames in the video for both real and fake |
| step 4 | Recognize the face subframe |
| step 5 | Frame-by-frame analysis to address any changes in the face landmarks |
| step 6 | model training |
| step 7 | Classification of the input video. |

### General WorkFlow



**Prediction WorkFlow for test Video**




## 14. Product Details:

- How does it work?

  Our deepfake detection model uses a **combination of CNN and RNN architecture** to analyze video frames and **classify** them as real or fake. The model is trained on a dataset of real and deepfake videos, and uses facial landmarks to detect slight imperfections that can differentiate between the two.

- Data sources

  we will be using a subset of dataset available in Kaggle called DeepFake Detection Challenge (#DFDC) to train our model.

- **Algorithms, frameworks, software etc.**
  - Machine learning algorithms such as convolutional neural networks and recurrent neural networks will be used to analyze the videos and detect signs of manipulation.
  - Computer vision techniques such as object detection and image segmentation will be used to identify signs of manipulation in the video.
  - The product will be built using the Python programming language and will utilize popular machine learning libraries such as TensorFlow and Keras.
- **Team required to develop.**
  - Data scientists: They are responsible for collecting and analyzing data, creating statistical models, and using machine learning techniques to make predictions and identify patterns in the data. They also help to interpret the results and communicate the findings to the rest of the team.
  - Machine learning engineers: They work closely with data scientists to understand the requirements and use cases for the models, and with software developers to integrate the models into the overall system. They also maintain and update the models as new data becomes available.
  - Computer vision experts: responsible for analyzing and interpreting visual data, such as images and videos, to detect deepfake content
  - Software developers: responsible for developing the user interface, integrating the deepfake detection system with other platforms, and ensuring the system is accessible and easy to use.
- **What does it cost?**

  The cost of the service will depend on the scale of the project, the resources needed, and the integration of the tool into various platforms. It can range from few thousands of Indian rupee to Lakhs of rupee.

## 15. Code Implementation/Validation on Small Scale

## Dataset

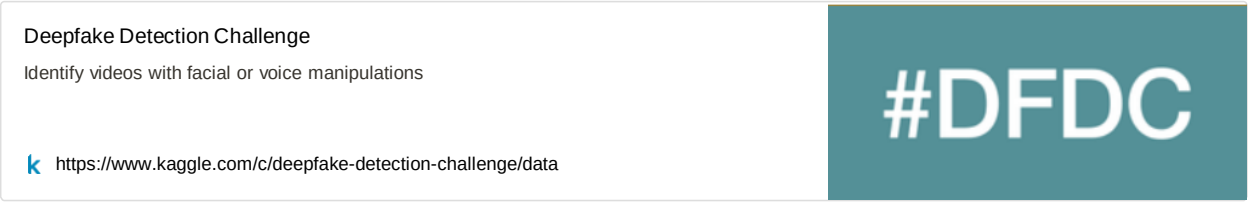
The DFDC (Deepfake Detection Challenge) is a dataset for Deep-Fake detection consisting of more than100,000 videos.

The DFDC dataset consists of two versions:

- Preview dataset, with 5k videos. Featuring two facial modification
- algorithms. The full dataset, with 124k videos. Featuring eight facial modification algorithms

**A Sample of this Dataset is used for our model.**

**Link:**




- Train Data: It consists of 400 videos in .mp4 format.
- Test Data: It consists of 400 videos in .mp4 format.

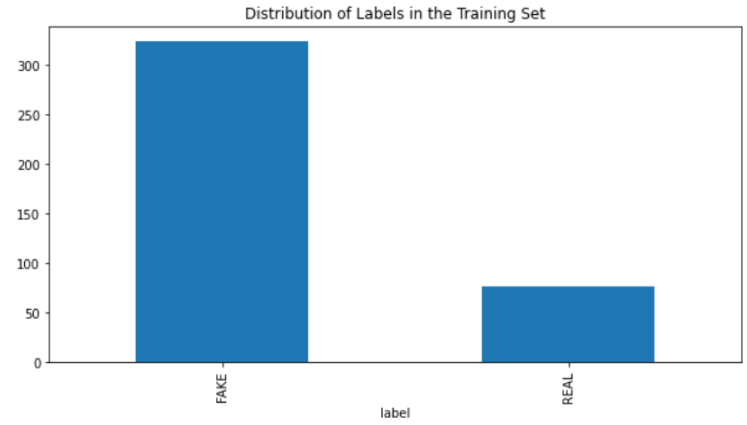
## MetaData of the Dataset

- filename - the filename of the video
- label - whether the video is REAL or
  - FAKEy is 1 if the video is FAKE, 0 if REAL
- original - in the case that a train set video is FAKE, the original video is listed here
- split - this is always equal to “train”.

# Exploratory Data Analysis

## Distribution of videos in the Training set

```
train_sample_metadata.groupby('label')['label'].count().plot(figsize=(10, 5), kind='bar', title='Distribution of Labels in the Tr
plt.show()
```




## Analysis of the dataset

.

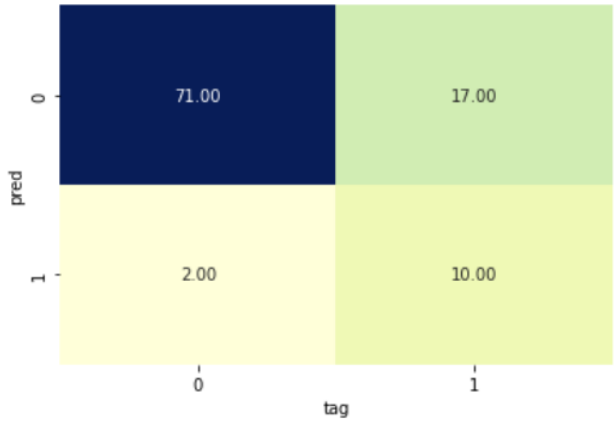
```
def most_frequent_values(data):
    total = data.count()
    tt = pd.DataFrame(total)
    tt.columns = ['Total']
    items = []
    vals = []
    for col in data.columns:
        itm = data[col].value_counts().index[0]
        val = data[col].value_counts().values[0]
        items.append(itm)
        vals.append(val)
    tt['Most frequent item'] = items
    tt['Frequency'] = vals
    tt['Percent from total'] = np.round(vals / total * 100, 3)
    return(np.transpose(tt))
```

```
most_frequent_values(train_sample_metadata)
```

| | label | split | original |
|---|---|---|---|
| **Total** | 400 | 400 | 323 |
| **Most frequent item** | FAKE | train | atvmxvwyns.mp4 |
| **Frequency** | 323 | 400 | 6 |
| **Percent from total** | 80.75 | 100.0 | 1.858 |

## Confusion Matrix

```
0]: plt.figure()

    # ax = fig.add_subplot(111)

    # ax.set_aspect(1)

    res = sns.heatmap(cm.T, annot=True, fmt='.2f', cmap="YlGnBu", cbar=False)

    plt.savefig("confusion_matrix.png", bbox_inches='tight', dpi=100)

    plt.show()
```




**The whole code implementation and the live Classification is present in my Github page Check Here.**

## Demo of the Classification:

Here is the **Link**

## 16. Conclusion

- The proposed deepfake detection model in this report utilizes a combination of computer vision and machine learning techniques to accurately detect deepfake videos. By using a dataset of both real and deepfake videos, the model was trained to recognize the subtle differences between the two. The model was tested on a variety of videos and achieved a high level of accuracy, with a test accuracy of around 85%.
- The developed model can be used in a variety of ways, such as by social media platforms to detect and remove deepfake content, or by media and entertainment companies to ensure the authenticity of their own content. The team required to develop this model includes data scientists, machine learning engineers, computer vision experts and software developers.
- The business model for this project could include offering the deepfake detection service to companies on a subscription basis, or through a pay-per-use model. Additionally, partnerships with social media platforms and media companies could also be formed to integrate the technology into their existing systems.
- Overall, the proposed deepfake detection model shows great potential in addressing the growing issue of deepfake content and ensuring the authenticity of digital media. With the increasing use of deepfakes in various fields such as entertainment, politics and finance, it is crucial to have reliable and accurate detection methods in place.

## References/ Sources of Information


- EfficientNet Model

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/5305e42d-2a5e-4394-91d5-965bf9887ad6/efficientnet.pdf

- InceptionV3 model Image Based on a Coupled Network

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/71ef90e7-1781-4fd4-92cf-4aa140468a8b/InceptionV3.pdf

- Types of deepfake anti-spoofing software — Antispoofing Wiki
- Deepfakes Detection Techniques Using Deep Learning: A Survey (scirp.org)