Design Credit Report (MEN3010)

Indian Institute of Technology, Jodhpur

Prof:- Amit Singh

Student:- Aman Goyal (B19ME004) , Rohit Nagar (B19ME069)

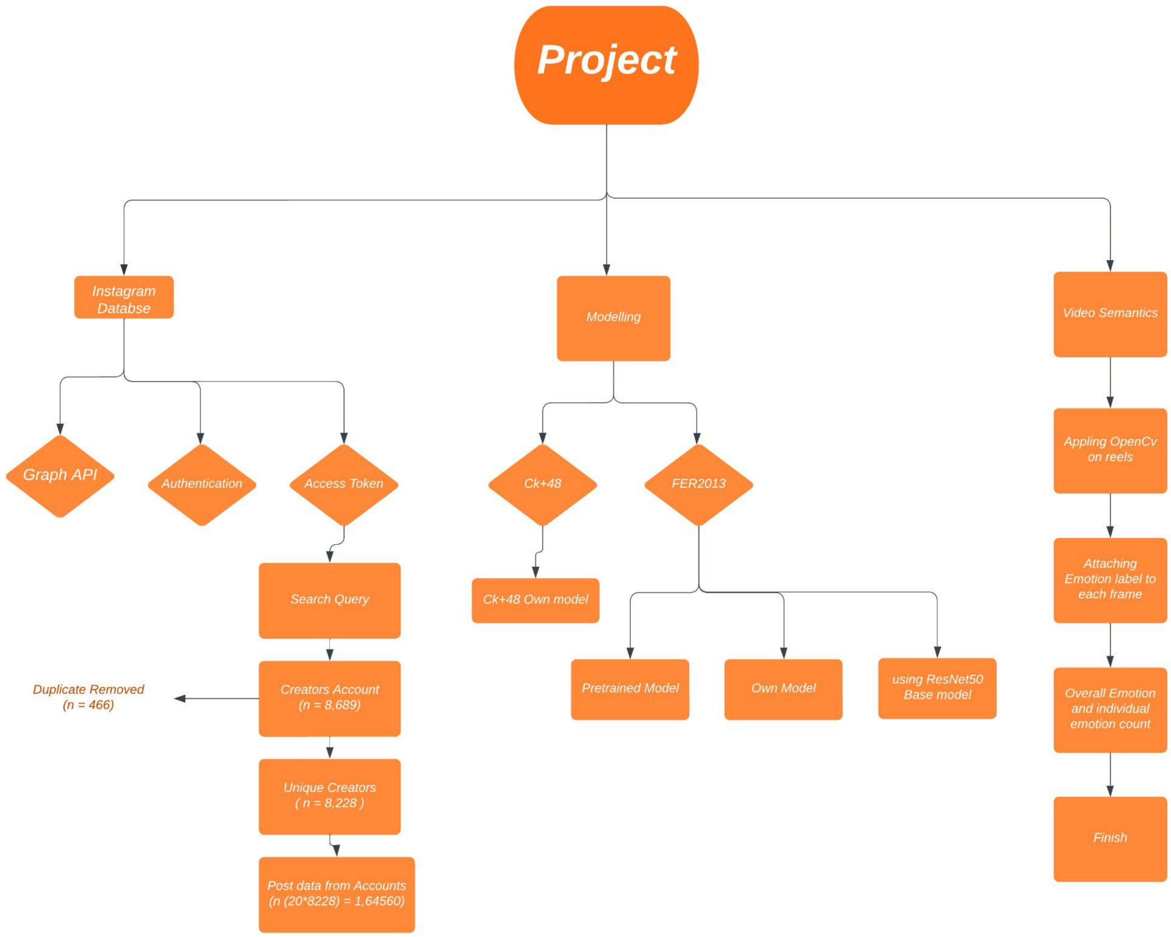
Problem Statement

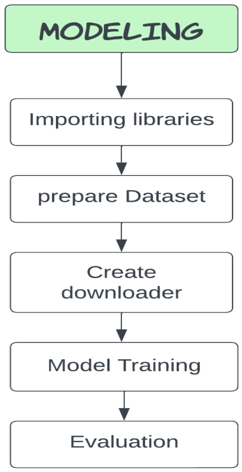
Problem Statement: Video analytics model on SMI videos

With the rise of short-form video content such as reels, TikTok, and Instagram Stories, there is a growing need for tools that can help users analyze and express emotions in their videos. Emotions play a key role in how we communicate and connect with others, and can have a significant impact on the success of social media content.
Traditional methods such as manual annotation or sentiment analysis may not be accurate or scalable enough to handle the volume and diversity of social media data.

To address this problem, we propose a Deep Learning-based approach for emotion detection in social media reels. Our goal is to develop a model that can accurately and automatically detect emotions such as happiness, sadness, anger, fear, and surprise in short-form video content. The model will be trained on a large and diverse dataset of social media reels, and will use state-of-the-art techniques such as deep learning, computer vision, and natural language processing to extract relevant features and patterns.

Mentioning Steps


Project
Instagram Databse
Modelling
Video Semantics
Graph API
Authentication
Access Token
Search Query
Duplicate Removed (n = 466)
Creators Account (n = 8,689)
Unique Creators (n = 8,228)
Post data from Accounts (n (20*8228) = 1,64560)
Ck+48
FER2013
Ck+48 Own model
Pretrained Model
Own Model
using ResNet50 Base model
Appling OpenCv on reels
Attaching Emotion label to each frame
Overall Emotion and individual emotion count
Finish



Work done

Dataset

- Instagram code database

We use the instagram graph api for extracting the data from instagram code base. First setup the graph-api app and authenticate with our business account. Then after we generate the access token from the api-app so that instagram will know about validity. We also wrote some python script to keep following the dry principle. After then we run our get_bussiness_query.py file in order to gain the information about influencers. Again run an another search query to get the recursive follower and following of the particular influencer.
At the end we have an 8,869 number of creators account in which we have 466 duplicates so we remove them. ultimately we have 8,228 number of unique creators accounts. We extract 20 posts from each creator and having 20*8228 = 164560 number of post in csv.

- CK+48 Dataset

The CK+48 dataset contains 593 image sequences of 123 subjects, with a total of 4,620 images.
Each image sequence captures the facial expressions of the subjects, which include seven different expressions: anger, contempt, disgust, fear, happiness, sadness, and surprise. CK+48 dataset also provide 3 different columns in which usages column mentioned that for upto what images we can use for training, validation and testing purpose. So with these steps we finally prepare our dataset in training, validations and testing to train the models and evaluate their performances.

- FER2013 Dataset

FER dataset consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centred and occupies about the same amount of space in each image.
The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

Categories	Emotion
0	Angry
1	Disgust
2	Fear
3	Happy
4	Sad
5	Surprise
6	Neutral

Model

Batch size = 32
Epochs = 50

CK+48 dataset
Model used:

- Own model

Own Model	• Sequential model with Conv2D, MaxPooling2D, Flatten, Dropout, and Dense layers. • Three Conv2D layers with ReLU activation and 3x3 kernels, followed by MaxPooling2D layers. • Two Dense layers with ReLU activation and a Dropout layer for regularization. • Output layer with 7 neurons corresponding to the number of classes

FER dataset
Model used:

- Pretrained model on fer
- Created our own Scratch model
- Create another optimal model with ResNet50 as base model

PreTrained Model	● This is a Sequential model with 24 layers. ● It has 4 Conv2D layers, each followed by BatchNormalization and MaxPooling2D layers. ● It also has 5 Dropout layers, and ends with 3 Dense layers (2 with Dropout). It has a total of 7 output classes. ● The model has a total of 3,880,455 trainable parameters
Own model	● This is a convolutional neural network model with a sequential architecture. ● The model has a total of 14 layers, including 5 convolutional layers, 5 batch normalization layers, 4 max pooling layers, 5 dropout layers, and 2 fully connected layers. ● The number of filters in each convolutional layer is 64, 128, 256, 512, and 1024 respectively. ● The output of the final convolutional layer is flattened and passed through two fully connected layers with 256 and 128 units respectively, before producing the final classification result.
Optimal Model	● This is a Keras sequential model with ResNet50 as the base model. ● The ResNet50 model has 23,587,712 parameters. ● The model has three dense layers with 32 units each, and the output layer has 7 units. ● The first dense layer has 262,176 parameters. ● The model also has three dropout layers, four batch normalization layers, and two activation layers. ● The model's total number of parameters is 23,848,103.

After training and evaluating these models , we compare their performance and used the best model for extracting the emotions from smi videos. For comparison between models we used confusion matrix and classification report that tell us which model is better for precision and recall for specific emotion category and how far one model is better from other in terms of f1-score.

Video Semantics

We apply openCv on reel. Function first convert an reel into frame after that attach each frame with the emotion label. Finally we get the dictionary having key-value that shows emotion as key and emotion-count as value as final outcome.

```python
emotion_labels = ['Angry', 'Disgust', 'Fear', 'Happy', 'Sad', 'Surprise', 'Neutral']

# Define function to preprocess images
def preprocess_image(image):
    # Convert image to grayscale
    gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
    # Resize image to match FER model input shape (48x48 pixels)
    resized = cv2.resize(gray, (48, 48))
    # Expand dimensions to create a batch of size 1
    expanded = np.expand_dims(resized, axis=0)
    # Convert image to float32 and normalize pixel values
    preprocessed = expanded.astype('float32') / 255.0
    return preprocessed

# Define function to predict emotion from image
def predict_emotion(image):
    preprocessed_image = preprocess_image(image)
    prediction = model_1.predict(preprocessed_image)
    emotion_idx = np.argmax(prediction)
    if emotion_idx < len(emotion_labels):
        emotion_label = emotion_labels[emotion_idx]
    else:
        emotion_label = "Unknown"
    return emotion_label

# Define function to classify emotion in video
def classify_emotion(video_path):
    cap = cv2.VideoCapture(video_path)
    emotions = []
    while True:
        ret, frame = cap.read()
        if not ret:
            break
        emotion = predict_emotion(frame)
```

Outcomes

Instagram-graph API

- Run following command in terminal

Python business_discovery.py

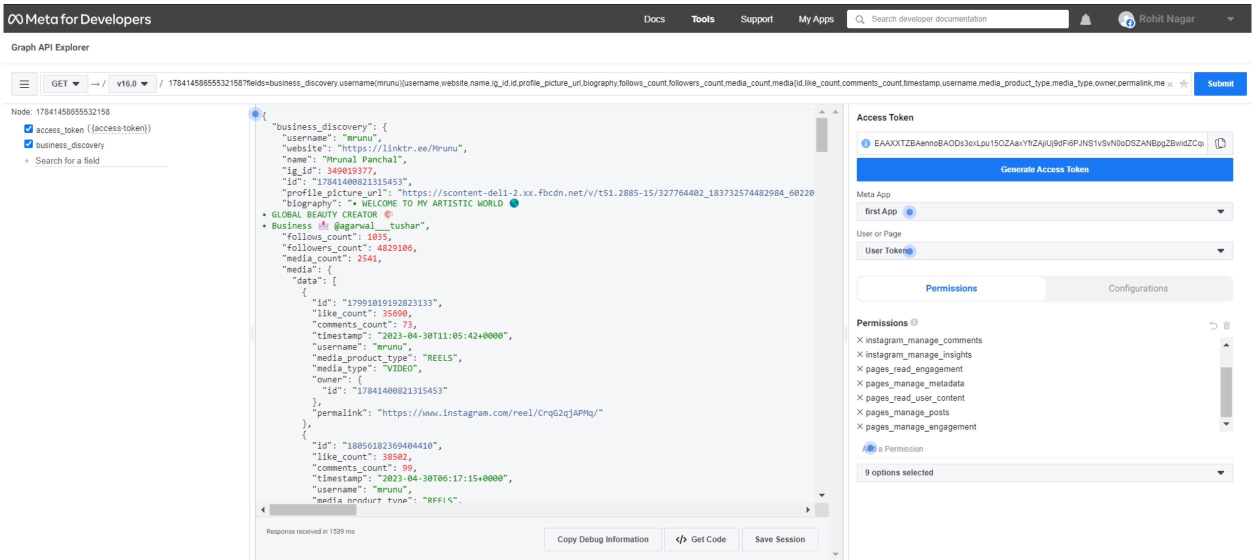
```
OUTPUT   DEBUG CONSOLE   PROBLEMS   TERMINAL

rohit@LAPTOP-QTDT3A6M MINGW64 ~/Desktop/Dc1
$ python business_discovery.py

---- ACCOUNT INFO -----

username: mrunu
website: https://linktr.ee/Mrunu
number of posts: 2541
followers: 4828976
following: 1035
biography:
 • WELCOME TO MY ARTISTIC WORLD 🌎
• GLOBAL BEAUTY CREATOR 🎨
• Business 📩 @agarwal___tushar
```

- Search query for extracting post data of an influencer (from Graph API)



- Search query for extracting post data of an influencer (from python script)

Run following command in terminal

Python get_users_post_data.py

r

Model Results

Model-1 (Ck+48 Own Model)

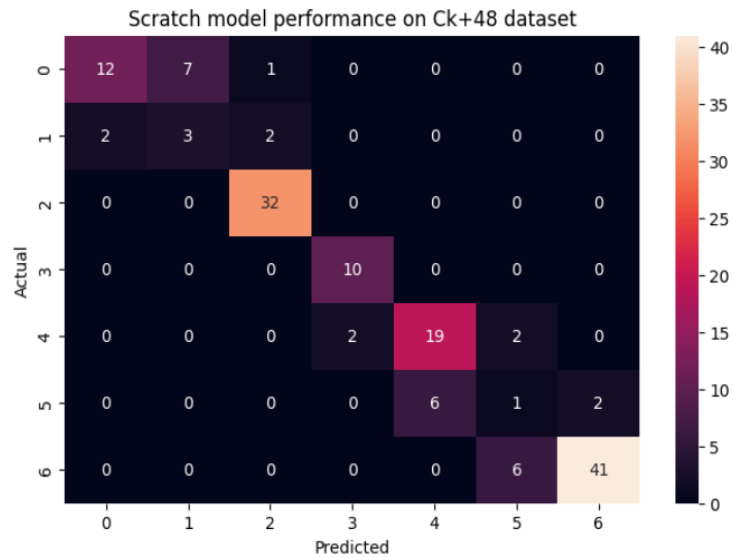
Classification Report

```
5/5 [==============================] - 0s 31ms/step
              precision    recall  f1-score   support

           0       0.86      0.60      0.71        20
           1       0.30      0.43      0.35         7
           2       0.91      1.00      0.96        32
           3       0.83      1.00      0.91        10
           4       0.76      0.83      0.79        23
           5       0.11      0.11      0.11         9
           6       0.95      0.87      0.91        47

    accuracy                           0.80       148
   macro avg       0.68      0.69      0.68       148
weighted avg       0.81      0.80      0.80       148
```

Classification Matrix



Model-2 (FER Pretrained)

Classification Report

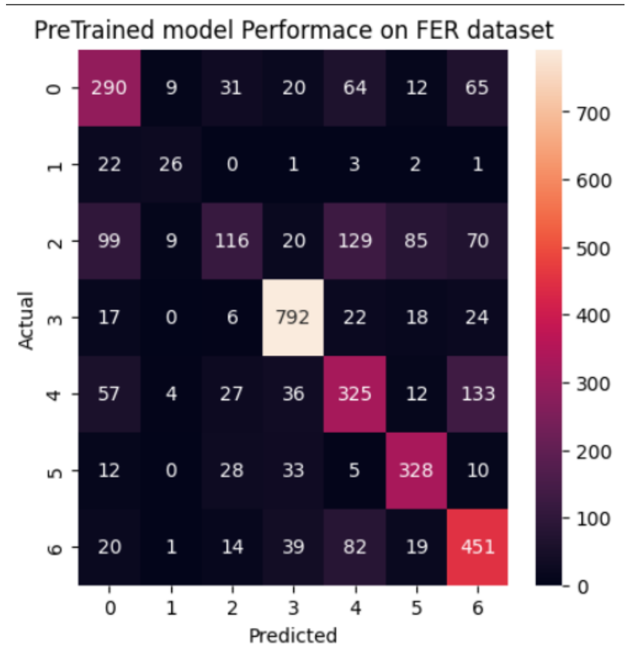
```
113/113 [==============================] - 66s 582ms/step
              precision    recall  f1-score   support

           0       0.56      0.59      0.58       491
           1       0.53      0.47      0.50        55
           2       0.52      0.22      0.31       528
           3       0.84      0.90      0.87       879
           4       0.52      0.55      0.53       594
           5       0.69      0.79      0.74       416
           6       0.60      0.72      0.65       626

    accuracy                           0.65      3589
   macro avg       0.61      0.61      0.60      3589
weighted avg       0.64      0.65      0.63      3589
```

Overal Precision:-

Classification Matrix



Model-3 (FER Own Model)

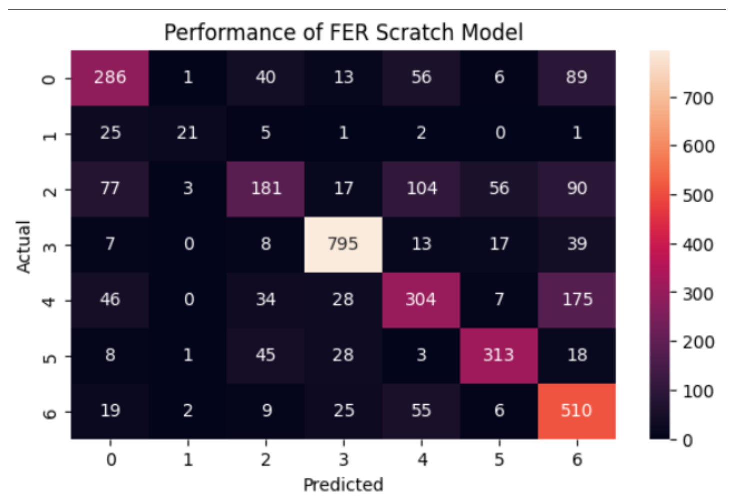
Classification Report	Classification Matrix
See report below	See matrix below

```
113/113 [==============================] - 2s 17ms/step
              precision    recall  f1-score   support

           0       0.61      0.58      0.60       491
           1       0.75      0.38      0.51        55
           2       0.56      0.34      0.43       528
           3       0.88      0.90      0.89       879
           4       0.57      0.51      0.54       594
           5       0.77      0.75      0.76       416
           6       0.55      0.81      0.66       626

    accuracy                           0.67      3589
   macro avg       0.67      0.61      0.63      3589
weighted avg       0.67      0.67      0.66      3589
```

Overall Precision :-



Model-4 (Optimal Model on top of the ResNet50 base model)

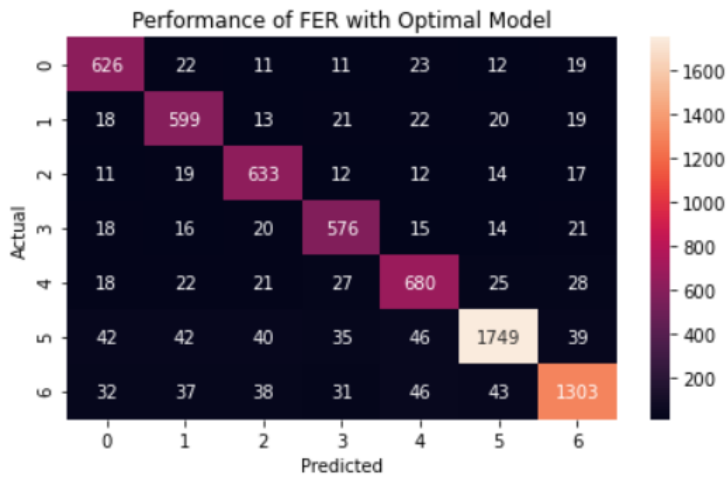
Classification Report	Classification Matrix

```
              precision    recall  f1-score   support

       angry       0.82      0.86      0.84       724
     disgust       0.79      0.84      0.82       712
        fear       0.82      0.88      0.85       718
       happy       0.81      0.85      0.83       680
     neutral       0.81      0.83      0.82       821
         sad       0.93      0.88      0.90      1993
    surprise       0.90      0.85      0.88      1530

    accuracy                           0.86      7178
   macro avg       0.84      0.86      0.85      7178
weighted avg       0.86      0.86      0.86      7178
```

Overall Precision :-



Video Semantics on Optimal Model

We feed the instagram reel to the optimal model , that attach each frame an emotion , so overall we get the dictionary as key-value pair having emotion as key and emotion-count as value to them.

Below is the final output

```
1/1 [==============================] - 0s 54ms/step
1/1 [==============================] - 0s 51ms/step
1/1 [==============================] - 0s 50ms/step
1/1 [==============================] - 0s 50ms/step
1/1 [==============================] - 0s 51ms/step
1/1 [==============================] - 0s 51ms/step
1/1 [==============================] - 0s 49ms/step
1/1 [==============================] - 0s 50ms/step
1/1 [==============================] - 0s 54ms/step
1/1 [==============================] - 0s 49ms/step
1/1 [==============================] - 0s 50ms/step
1/1 [==============================] - 0s 52ms/step
1/1 [==============================] - 0s 49ms/step
```

```
import collections

# using Counter to find frequency of elements
frequency = collections.Counter(emotions)

# printing the frequency
print(dict(frequency))
```

```
{'Neutral': 38, 'Happy': 122, 'Sad': 128, 'Angry': 12}
```

Benefits to the Stakeholder

There could be several benefits of this project to stakeholders, depending on the specific details and context of the project. Here are some examples:

1. Improved user engagement: One possible benefit is that the application can improve user engagement and retention on social media platforms, by providing users with a new and engaging way to analyze and express emotions in their reels. This can lead to increased user satisfaction and loyalty, and can help social media platforms differentiate themselves from competitors.

2. Enhanced content quality: Another possible benefit is that the application can help users improve the quality and relevance of their reels, by providing them with insights into the emotions that resonate with their audience. This can lead to more engaging and viral content, and can help users build their brand or reputation on social media.

3. Better emotional intelligence: A third possible benefit is that the application can help users develop their emotional intelligence and empathy, by providing them with feedback on their own emotional expression and the emotions of others. This can lead to improved communication, collaboration, and relationships both online and offline.

4. Improved mental health: A fourth possible benefit is that the application can improve mental health and well-being for users, by providing them with a new and accessible tool for monitoring and managing their emotions. This can lead to reduced stress, anxiety, and depression, and can help users develop healthy coping mechanisms and habits.

5. Competitive advantage: A fifth possible benefit is that the application can provide social media platforms with a competitive advantage over other platforms or applications that do not offer emotion detection or analysis. This can help social media platforms attract new users and retain existing ones, and can enhance their brand image and reputation.

6. Social media platforms: Social media platforms can use the emotion detection model to improve user engagement and retention, as users are more likely to engage with emotionally engaging content. This can help platforms to attract and retain users, and potentially increase their revenue through advertising or other business models.

7. Advertisers: Advertisers can use the emotion detection model to create more emotionally resonant and effective advertisements, which can lead to higher conversion rates and sales. This can help them to achieve a better return on their advertising investment, and potentially increase their market share or revenue.

8. Researchers: The emotion detection project can provide valuable insights into the emotional dynamics of social media platforms, including how emotions are expressed and perceived across different cultures

and languages. This can help researchers to better understand the social and psychological processes underlying social media use, and potentially inform the development of interventions or policies aimed at improving well-being or reducing harm.

9. Society at large: By improving the emotional quality and engagement of social media content, the emotion detection project can potentially contribute to a more positive and supportive online environment. This can help to reduce social isolation, promote social connections, and improve overall mental health and well-being.

Assumptions

- The model will only be able to detect emotions based on facial expressions in the video, not be applicable to videos or content that do not feature human faces or emotions expressed through facial expressions.
- Other factors such as body language, tone of voice, or environmental cues may not be taken into account.

Limitations

- Data imbalance, if the dataset used for training the emotion detection model is not representative of the broader population, leading to potential biases or inaccuracies in the model's predictions.
- Videos longer than 90 seconds may not be applicable for emotion detection as the model is specifically designed for detecting emotions in social media reels, which typically have a maximum length of 60 seconds.
- The model may not be able to accurately detect emotions related to objects or tools in the background of the videos, as it is designed to focus on facial expressions of humans in the foreground.
- The accuracy of the model may be limited by the quality of the video, including resolution, lighting, and camera angles, which could impact the visibility and clarity of facial expressions.
- The model may not be able to detect emotions for individuals who do not display typical facial expressions, such as individuals with certain neurological or developmental conditions.
- The model may not be able to accurately detect nuanced emotions, as the dataset used for training the model may not have enough examples of those emotions or expressions.
- The model may not be able to detect emotions accurately in cases where the facial expressions of individuals in the video are intentionally concealed or modified, such as with facial masks or filters.

Conclusion

We have presented a machine learning-based approach for detecting emotions in social media reels, with the goal of enhancing emotional expression and engagement in short-form video content. Through the use of state-of-the-art techniques such as deep learning, computer vision, and natural language processing, we have developed a model that can accurately and automatically detect emotions such as happiness, sadness, anger, fear, and surprise in social media reels.

The proposed application has several potential benefits for various stakeholders, including social media users, businesses, and advertisers. For example, social media users can use the application to analyze and improve the emotional quality of their reels, while businesses and advertisers can use it to better understand their audience and target their campaigns more effectively. Additionally, the application can help users develop their emotional intelligence and empathy, and can contribute to improved mental health and well-being.

While the model has certain limitations, including its reliance on visible facial expressions and potential biases related to the dataset used for training, it has the potential to be a valuable tool for content creators and marketers looking to optimize their messaging and understand audience responses.

Overall, the proposed project represents a significant contribution to the field of emotion detection and analysis in social media, and has the potential to enhance emotional expression and engagement in short-form video content. With further development and refinement, the proposed application can provide valuable insights and benefits for users, businesses, and researchers alike.

References

1) https://keras.io/
2) https://developers.facebook.com/docs/instagram-api
3) https://paperswithcode.com/dataset/fer2013
4) https://paperswithcode.com/dataset/ck
5) https://opencv.org/