**Live Class Monitoring System [Face Emotion Recognition]**

**Kartik Pandey**

**Sagar Khekale**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

This project aims to classify the emotion on a person's face into one of **seven categories**, using deep convolutional neural networks. The model is trained on the **FER-2013** dataset which was published on International Conference on Machine Learning (ICML). This dataset consists of 35887 grayscale, 48x48 sized face images with **seven emotions** - angry, disgusted, fearful, happy, neutral, sad and surprised.

**1.Problem Statement**

We will solve the above-mentioned challenge by applying deep learning algorithms to live video data. The solution to this problem is by recognizing facial emotions. Through facial emotion recognition, we are able to measure the effects that content and services have on the audience/users through an easy and low-cost procedure. For example, retailers may use these metrics to evaluate customer interest. Healthcare providers can provide better service by using additional information about patients' emotional state during treatment. Entertainment producers can monitor audience engagement in events to consistently create desired content.

**2. Introduction**

The Indian education landscape has been undergoing rapid changes for the past 10 years owing to the advancement of web-based learning services, specifically, eLearning platforms. Global E-learning is estimated to witness an 8X over the next 5 years to reach USD 2B in 2021. India is expected to grow with a CAGR of 44% crossing the 10M users mark in 2021. Although the market is growing on a rapid scale, there are major challenges associated with digital learning when compared with brick and mortar classrooms. One of many challenges is how to ensure quality learning for students. Digital platforms might overpower physical classrooms in terms of content quality but when it comes to understanding whether students are able to grasp the content in a live class scenario is yet an open-end challenge. In a physical classroom during a lecturing teacher can see the faces and assess the emotion of the class and tune their lecture accordingly, whether he is going fast or slow. He can identify students who need special attention. Digital classrooms are conducted via video telephony software program (exZoom) where it’s not possible for medium scale class (25-50) to see all students and access the mood. Because of this drawback, students are not focusing on content due to lack of surveillance. While digital platforms have limitations in terms of physical surveillance but it comes with the power of data and machines which can work for you. It provides data in the form of video, audio, and texts which can be analysed using deep learning algorithms. Deep learning backed system not only solves the surveillance issue, but it also removes the human bias from the system, and all information is no longer in the teacher’s brain rather translated in numbers that can be analysed and tracked.

## **3. Dataset Information**

We have built a deep learning model which detects the real time emotions of students through a webcam so that teachers can understand if students are able to grasp the topic according to students' expressions or emotions and then deploy the model. The model is trained on the FER-2013 dataset .This dataset consists of 35887 grayscale, 48x48 sized face images with seven emotions - angry, disgusted, fearful, happy, neutral, sad and surprised. Here is the dataset link:- <https://www.kaggle.com/msambare/fer2013>

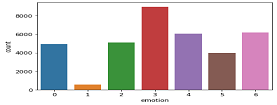
**4. Reasons for Face Recognition**

The reasons for face recognition:

* Student emotion during live class
* Entering in offices with face detection
* To help access personal device
* Face attendance based system and so on

# **5. Emotion Database**

##### **In the data collection steps, this is used both in real-world media and online media to collect as much data as that could. Real-world includes different types of emotional pictures of friends and family members, relatives, some known unknown people’s different kinds of facial expressions. They culled data was initially stored for future analysis. From online media, the data is collected data set from kaggle.com. This site uploaded this data set 6years ago. This site most trusted data set of emotions. This converted the data into 48×48 pixel grayscale images of faces. It contains two sections pixels and feelings. The feeling section contains a numeric code which runs from 0 to 6 in figure 1. What's more, the pixel section contains a string incorporated in statements for each picture. Furthermore, the picture should be only the picture of a face. So the collected pictures are resized and cropped picture of a fac**



**6. Steps involved:**

* **Exploratory Data Analysis**

After we use the fer dataset we load in kaggle kernal we use small EDA we check which emotion is has large image

# **Realtime Local Video Face Detection**

**We created one patterns for detecting and predicting single faces and as well as multiple faces using OpenCV videocapture in local. For Webapp , OpenCV can’t be used. Thus, using Streamlit-Webrtc for front-end application.**

* **Real time video detector usi ng haarcascade file**

Haar Cascade is a machine learning-based approach where a lot of positive and negative images are used to train the classifier. Positive images – These images contain the images which we want our classifier to identify. Negative Images – Images of everything else, which do not contain the object we want to detect

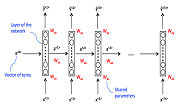
* **Feature Selection**

We use data augmentation it useful Data augmentation is a strategy that enables practitioners to significantly increase the diversity of data available for training models, without actually collecting new data. Data augmentation techniques such as cropping, padding, and horizontal flipping are commonly used to train large neural networks.

**7.1.Model :**

1. **Sequential Model:**

**Sequence modeling, put simply, is the process of generating a sequence of values by analyzing a series of input values. These input values could be time series data where a specific variable, say the demand for a particular product, varies over a period of time. The output could be the prediction of demand for consequent periods. Another example could be text prediction, where based on the sequence of the last phrase and a set of pre-loaded conditions and rules, the sequence modeling algorithm predicts what the next word could be. By using sequence modeling, businesses can achieve more than just pattern generation and prediction.**

****

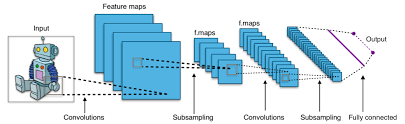
.

1. **Deep Convolution Network(DCNN):**

Deep learning is a machine learning technique used to build artificial intelligence (AI) systems. It is based on the idea of ​​artificial neural networks (ANN), designed to perform complex analysis of large amounts of data by passing it through multiple layers of neurons.

There is a wide variety of deep neural networks (DNN). Deep convolutional neural networks (CNN or DCNN) are the type most commonly used to identify patterns in images and video. DCNNs have evolved from traditional artificial neural networks, using a three-dimensional neural pattern inspired by the visual cortex of animals.

Deep convolutional neural networks are mainly focused on applications like object detection, image classification, recommendation systems, and are also sometimes used for natural language processing.

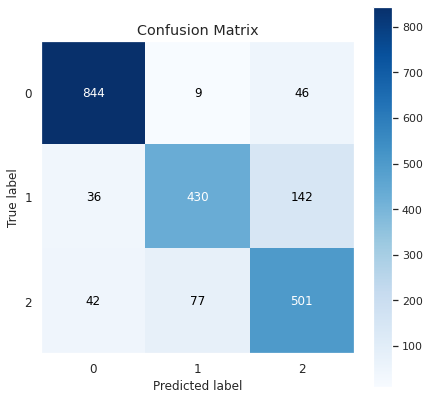


**7.2. Model performance:**

Model can be evaluated by various metrics such as:

1. **Confusion Matrix**-

The confusion matrix is a table that summarizes how successful the classification modelis at predicting examples belonging to various classes. One axis of the confusion matrix is the label that the model predicted, and the other axis is the actual label.



1. **Precision/Recall**-

Precision is the ratio of correct positive predictions to the overall number of positive predictions : TP/TP+FP

Recall is the ratio of correct positive predictions to the overall number of positive examples in the set: TP/FN+TP

1. **Accuracy**-

Accuracy is given by the number of correctly classified examples divided by the total number

of classified examples. In terms of the confusion matrix, it is given by: TP+TN/TP+TN+FP+FN

1. **Area under ROC Curve(AUC)**-

ROC curves use a combination of the true positive rate (the proportion of positive examples predicted correctly, defined exactly as recall) and false positive rate (the proportion of negative examples predicted incorrectly) to build up a summary picture of the classification performance

**8. Conclusion:**

That's it! We reached the end of our exercise.

Starting with loading the data so far we have done EDA , , feature selection and then model building.

In all of these models our accuracy revolves

Finally we build the webapp and deployed which has training accuracy of 72.15% and test accuracy of 66.64% .

**References-**

1. https://arxiv.org/pdf/1902.01019.pdf
2. Towardsdatascience
3. Analytics Vidhya