Project Based Learning Report

on

Gender and Age Detection with Data Science

Submitted in the partial fulfillment of the requirements
For the Project based learning in (**Essentials of Data Science**)
in
Electronics & Communication Engineering

By

2014111124	**Aditya Tandon**
2014111767	**Shivam Shashwat**
2014111111	**Priyanka Singh**

Under the guidance of Course In-charge

Prof. Dnyanesh S.Lavhkare

Department of Electronics & Communication Engineering

Bharati Vidyapeeth
(Deemed to be University)
College of Engineering,
Pune – 4110043

Academic Year: 2021-22

Bharati Vidyapeeth
(Deemed to be University)
College of Engineering,
Pune – 411043

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

CERTIFICATE

Certified that the Project Based Learning report entitled, **"Gender and Age Detection with Data Science"** is work done by

2014111124	**Aditya Tandon**
2014111767	**Shivam Shashwat**
2014111111	**Priyanka Singh**

in partial fulfillment of the requirements for the award of credits for Project Based Learning (PBL) in **Essentials of Data Science Course** of Bachelor of Technology Semester IV, Electronics & Communication Engineering.

Date: 21 May 2022

Prof. Dnyanesh S.Lavhkare
Course In-charge

Dr. Tanuja S.Dhope
PBL Co-Ordinator

Dr. Arundhati A.Shinde

Professor & Head

ELECTRONICS & COMMUNICATION ENGINEERING

Index: -

Page No.	Contents
1-1	Problem Statement with Solution
2-3	Description about project
4-5	Software Used
6-11	Results with Analysis
11-11	Conclusion & Outcome

Problem Statement :-

What is Data Science? Why learn Data Science?

Solution :-

Data science is the domain of study that deals with vast volumes of data using modern tools and techniques to find unseen patterns, derive meaningful information, and make business decisions. Data science uses complex machine learning algorithms to build predictive models. The data used for analysis can come from many different sources and presented in various formats.

Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. Data science practitioners apply machine learning algorithms to numbers, text, images, video, audio, and more to produce artificial intelligence (AI) systems to perform tasks that ordinarily require human intelligence. In turn, these systems generate insights which analysts and business users can translate into tangible business value.

Reasons to learn Data Science are: -

1) Learning about data science provides an opportunity for you to recreate yourself.
2) We live in a digital world, everything is data-driven. There is data science in business, accounting, education, science, engineering, healthcare, technology, energy sector, government, and so on.
3) Data science is also a very promising field with lots of high paying job opportunities.
4) Basic data science skills are important for personal use.
5) Great potential to branch out with different options.
6) Become a decision-maker, not every job opportunity will give you the power to make informed business decisions. For a data scientist, that is the core responsibility.
7) Less competitive because it is a highly analytical role, competition is less, but demand is not. With a limited talent pool, there is always a challenge for businesses to hire in these roles.

Design Gender and Age Detection with Data Science with OpenCV

The enhancing of raw images that are received from the camera sources, from satellites, aircrafts and the pictures captured in day-to-day lives is called image processing. The images have been processed through many different techniques and calculations have been made on the basis and analysis of the studies. There is a need of analyzing and studying the digitally formed images. There are two main and very common steps followed for image processing. The improvement of an image such that the resulted image is of greater quality and can be used by other programs, is called image enhancement [1]. The other technique is the most sought after technique used for extraction of information from an image. There is a division of the image into certain number of parts or objects so that the problem is solved. This process is called segmentation.

Datasets: -

We have downloaded three datasets about the Age and gender detection from
`https://drive.google.com/uc?id=1_aDScOvBeBLCn_iv0oxSO8X1ySQpSbIS`

which are-
model weight, age&genderDetection.ipynb and README.md. We have performed analysis visualization on dataset, age_and_gender_detection on Google colab .

Libraries used: -

1) NumPy library **-** NumPy is used to perform various mathematical operations on arrays.
2) Pandas Library **-** pandas provides various data structures and operations for manipulating numerical data and time series.
3) Matplotlib library from which pyplot module is used for plotting library used for 2D graphics.
4) Seaborn library - seaborn is a library for making statistical graphics in Python.

LITERATURE SURVEY: -

We already have several approaches to detect gender and age through facial images. We can do this classification on Gender Based on Human faces has been detected. We have collected certain data of equipped work and
worked through it to detect age and gender and mentioned the methods used below. Here fig 1 indicates the Proposed age and gender detection.

III.METHODOLOGY

In the section, we present the methodology of the proposed age and gender detection system. The first step is to input the data. The second step is tokenization and extraction of the feature sets that we will use later to build the classifier, where tokenization of the data means chopping the text into words, The third step is applying string to word vector which is very important as it cleans the data by removing unnecessary information in order to improve system performance. The fourth step is applying feature selection to the data. The fifth step is applying the classifier using different algorithm namely (random forest, naive Bayes, decision tree).The last step is producing the output class and evaluating the results.

Software Used: -

Google colab-

Google Colab was developed by Google to provide free access to GPU’s and TPU’s to anyone who needs them to build a machine learning or deep learning model. Google Colab can be defined as an improved version of Jupiter Notebook..

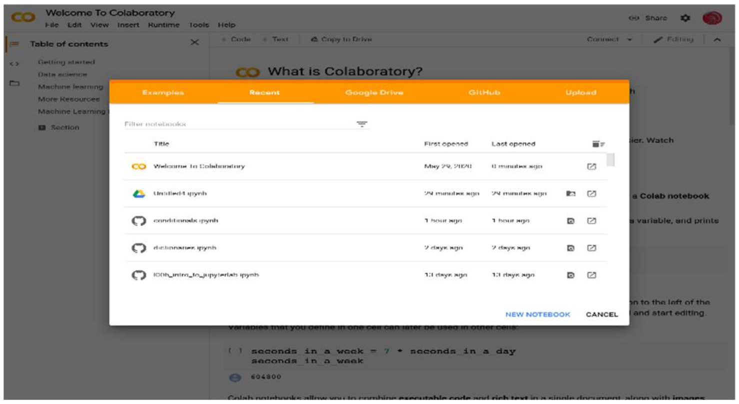
Programming Languages are an intermediate form between human-understandable language and machine understandable language. Every application is built using one of the many programming languages available. Maybe a person with a computer science background can understand, but not everyone can. Remember, as Software Developers, we develop applications for people with little computer science knowledge.

Consider you are creating a machine learning model to improve customer satisfaction for a local store, in that case you will have to explain how the model can do this task, and you can’t just explain him with your code base. Most people facing this situation will prepare a separate presentation. Notebooks were created so that it is not necessary. Notebook documents can include executable lines of code along with text, images, figures, tables, graphs, equations, and much more graphical data. In simple words, Notebook documents are a way of creating human-readable executable documents.

Google Colab Features

Google Colab provides tons of exciting features that any modern IDE offers, and much more. Some of the most exciting features are listed below.

- Interactive tutorials to learn machine learning and neural networks.
- Write and execute Python 3 code without having a local setup.
- Execute terminal commands from the Notebook.



Title	First opened	Last opened
Welcome to Colaboratory	May 29, 2020	0 minutes ago
Untitled4.ipynb	29 minutes ago	29 minutes ago
conditionals.ipynb	1 hour ago	1 hour ago
dictionaries.ipynb	2 days ago	2 days ago
100h_intro_to_jupyterlab.ipynb	13 days ago	13 days ago

```
[ ] seconds_in_a_week = 7 * seconds_in_a_day
    seconds_in_a_week
604800
```

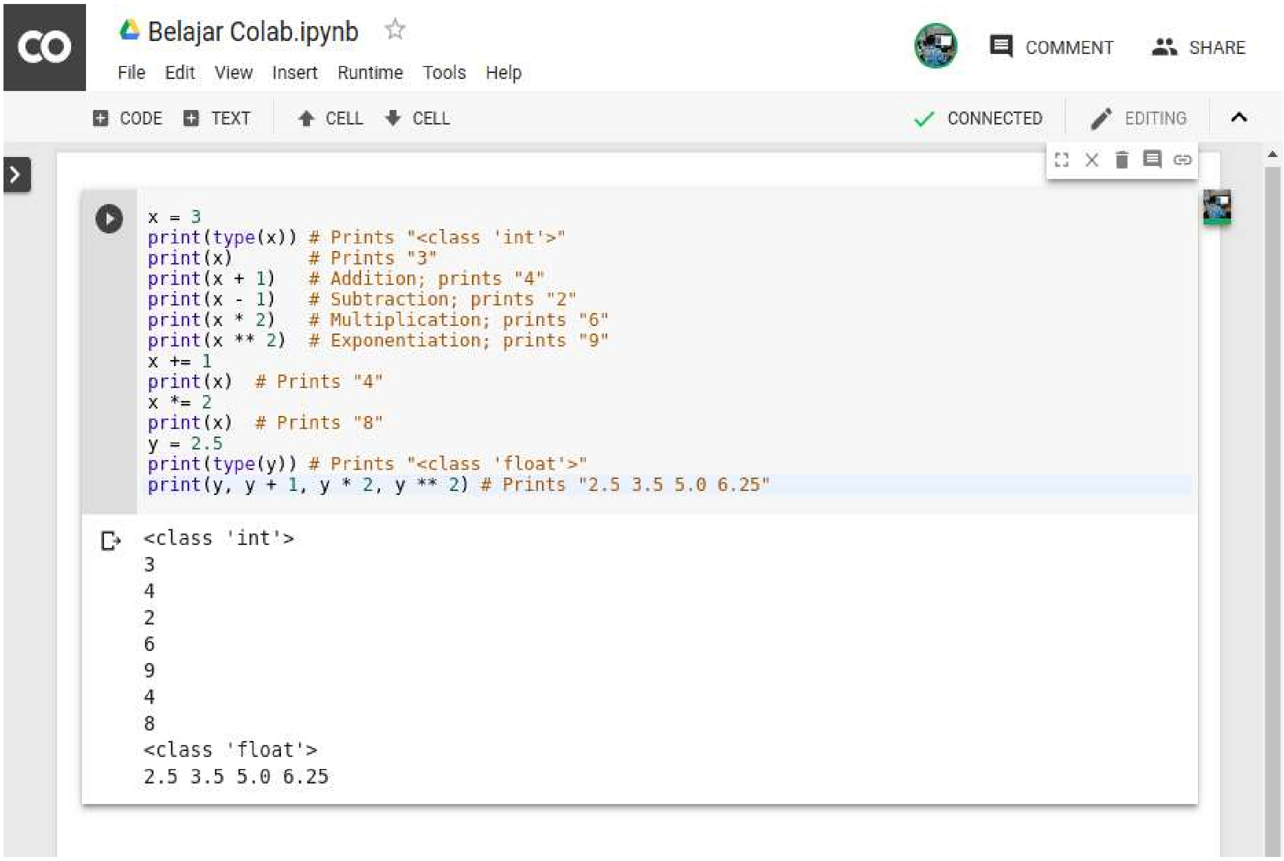
- Import datasets from external sources such as Kaggle.
- Save your Notebooks to Google Drive.
- Import Notebooks from Google Drive.

- Free cloud service, GPUs and TP
- Integrate with PyTorch, Tensor Flow, Open CV.
- Import or publish directly from/to GitHub.

How to Use Google Colab?

Just like any other product from Google, you need a Google account to get started. You can access Google Colab using this link Welcome to Collaboratory – Collaboratory (google.com) and signing in through your Google account.

You will land on a page similar to the one shown below.



```
x = 3
print(type(x)) # Prints "<class 'int'>"
print(x)       # Prints "3"
print(x + 1)   # Addition; prints "4"
print(x - 1)   # Subtraction; prints "2"
print(x * 2)   # Multiplication; prints "6"
print(x ** 2)  # Exponentiation; prints "9"
x += 1
print(x)  # Prints "4"
x *= 2
print(x)  # Prints "8"
y = 2.5
print(type(y)) # Prints "<class 'float'>"
print(y, y + 1, y * 2, y ** 2) # Prints "2.5 3.5 5.0 6.25"
```

```
<class 'int'>
3
4
2
6
9
4
8
<class 'float'>
2.5 3.5 5.0 6.25
```

Result with Analysis

Analysis of the code: -

- First, we have imported four libraries – NumPy as np, pandas as pd, matplotlib.pyplot as plt and seaborn library as sns.
- Secondly, we have loaded our dataset – tracks.csv using read_csv() function of pandas library and used head() function for displaying first five rows of the dataset.
- Then, for checking null values in the dataset, we have used isnull() function of pandas library.
- After this, we have created a Convolution Map using heatmap() function of Seaborn library and then , we have created a Regression Plot using regplot() function of Seaborn library.

Then, we used displot() function of Seaborn library, to create a Distribution plot for the 'number of songs per year'. At last, we used barplot() function of Seaborn library to create a barplot.

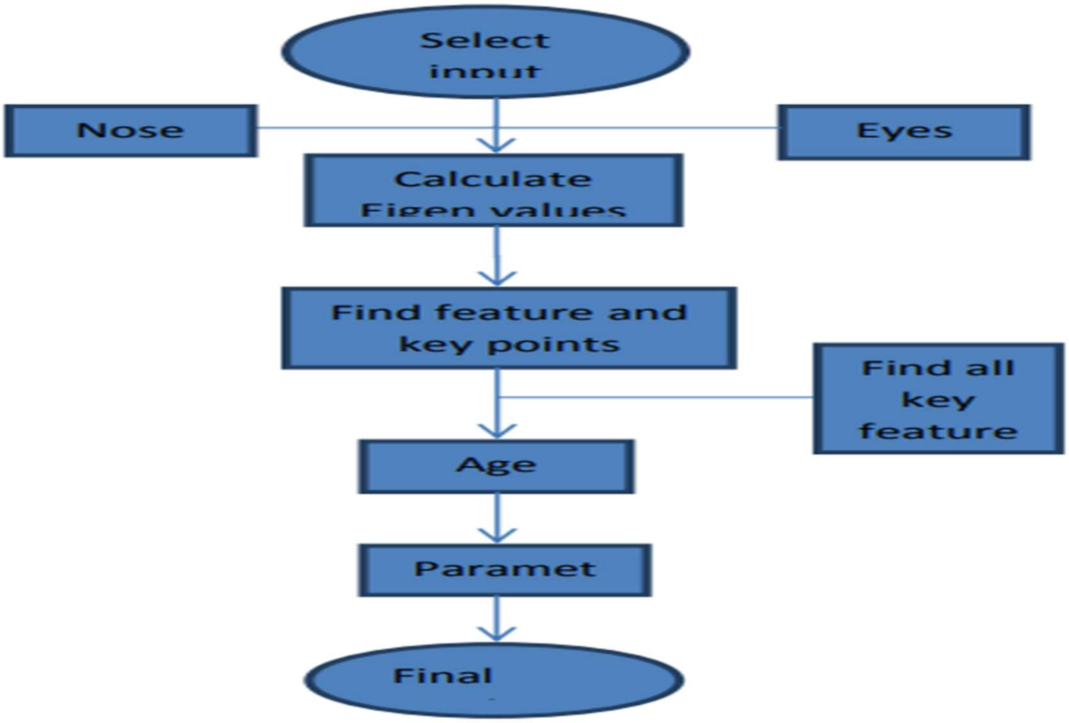
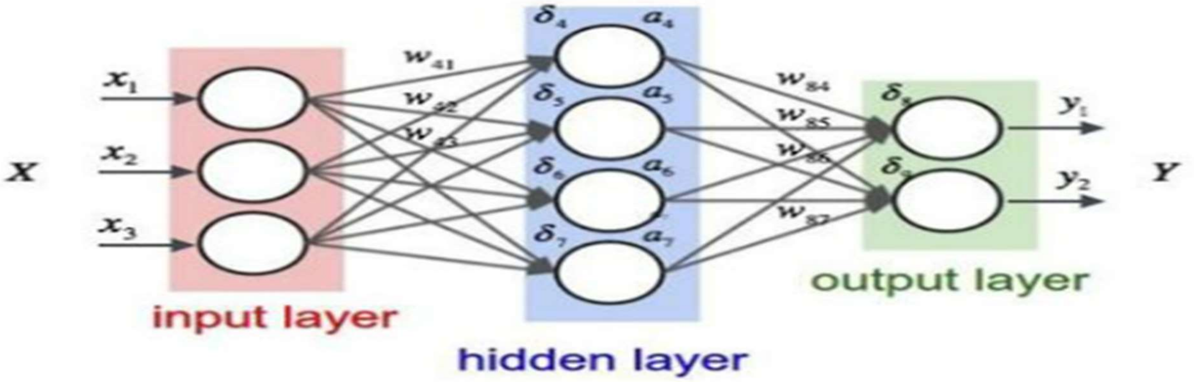
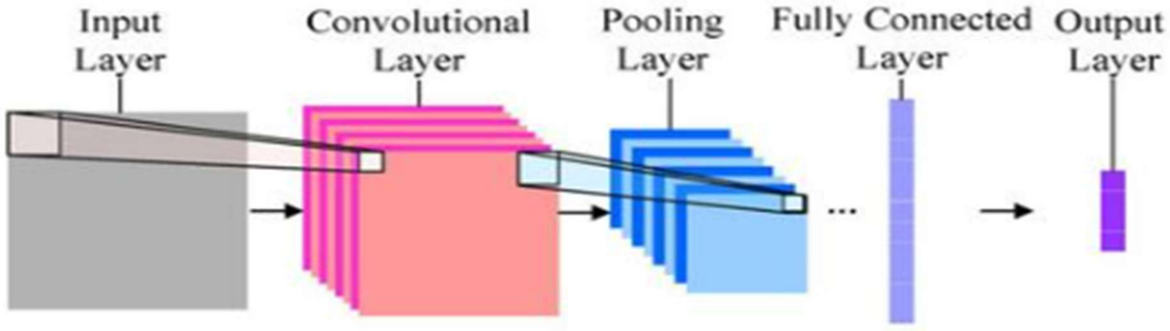


Fig 1: Flowchart of Proposed Technique

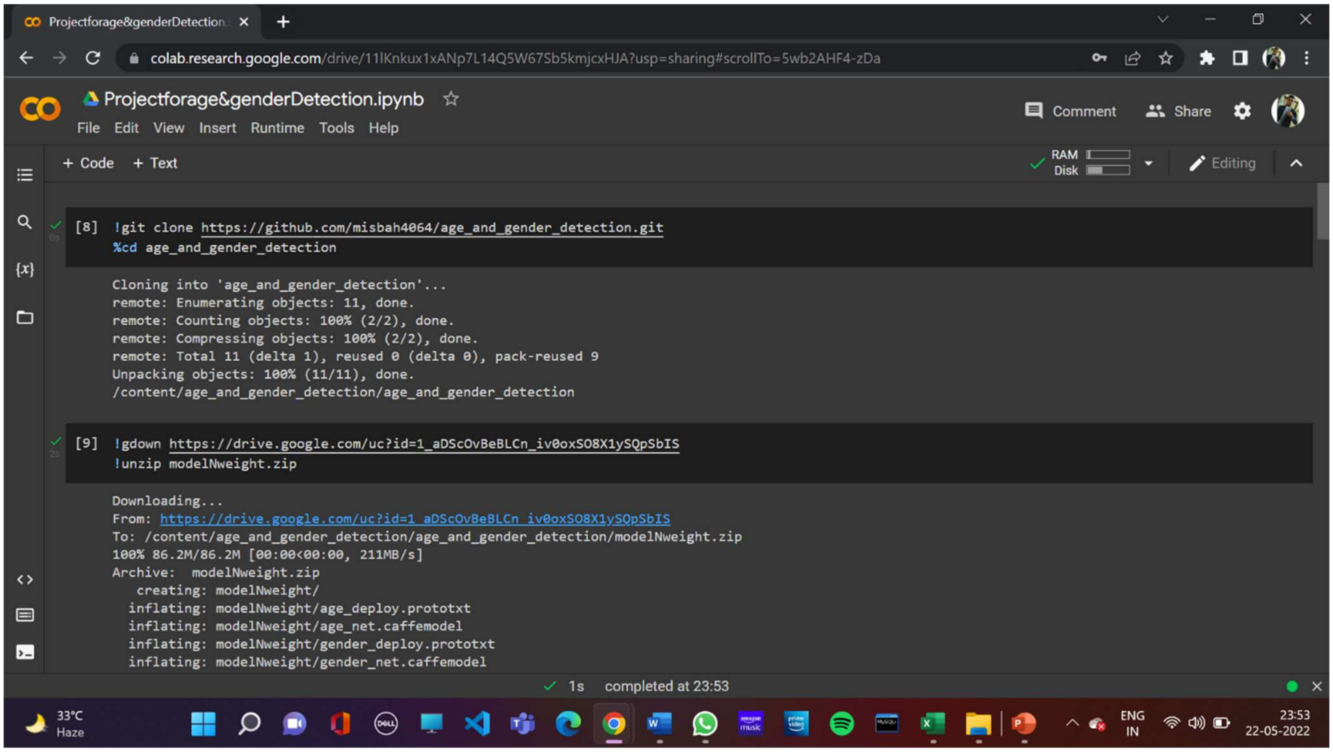
Fully Connected Neural Network



Convolutional Neural Network

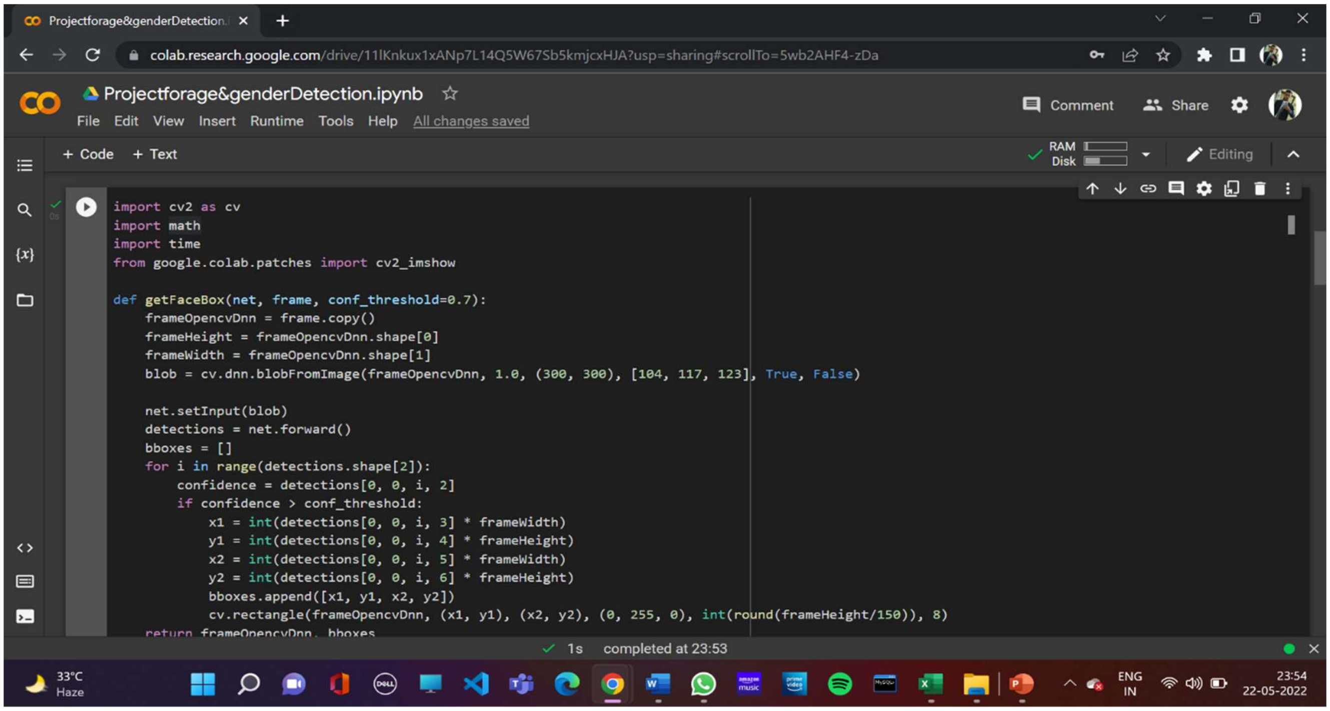


Screenshots of code: -

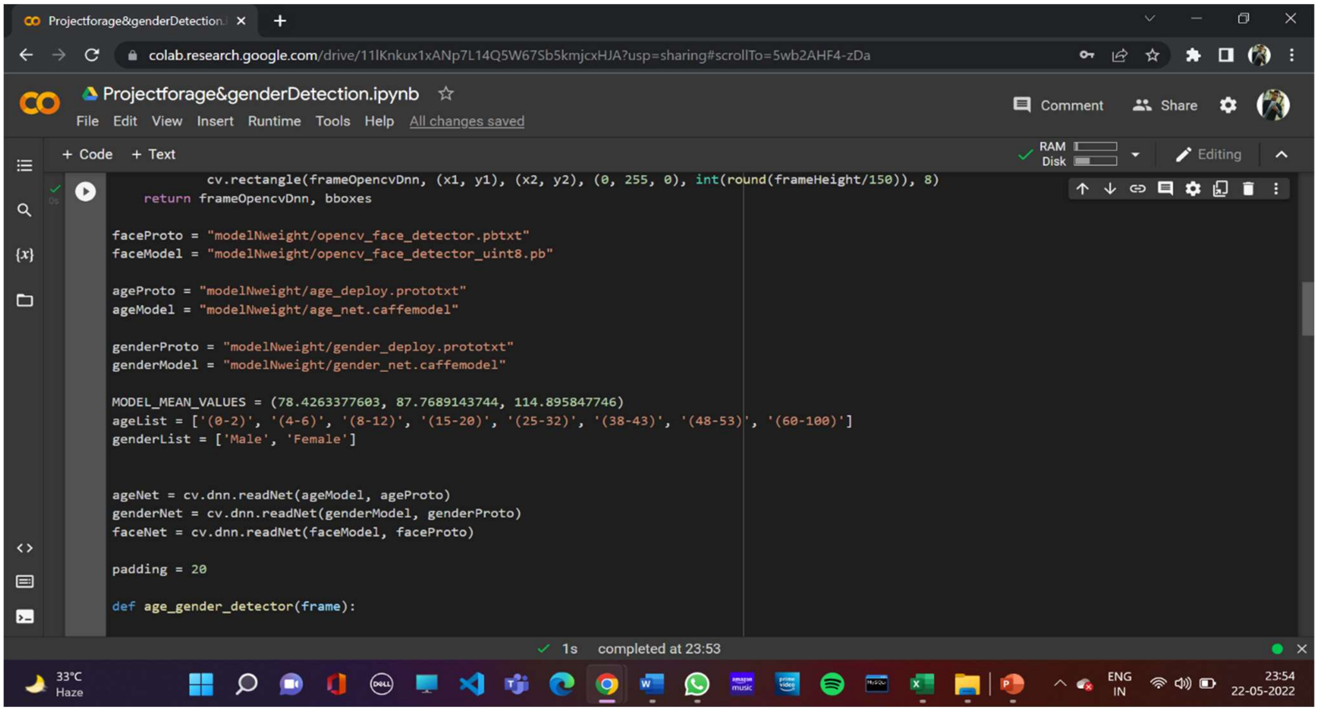

Projectforage&genderDetection.i
colab.research.google.com/drive/11lKnkux1xANp7L14Q5W67Sb5kmjcxHJA?usp=sharing#scrollTo=5wb2AHF4-zDa
Projectforage&genderDetection.ipynb
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code
+ Text
RAM
Disk
Editing
1s completed at 23:53
33°C
Haze
ENG
IN
23:53
22-05-2022

```
[8] !git clone https://github.com/misbah4064/age_and_gender_detection.git
    %cd age_and_gender_detection

    Cloning into 'age_and_gender_detection'...
    remote: Enumerating objects: 11, done.
    remote: Counting objects: 100% (2/2), done.
    remote: Compressing objects: 100% (2/2), done.
    remote: Total 11 (delta 1), reused 0 (delta 0), pack-reused 9
    Unpacking objects: 100% (11/11), done.
    /content/age_and_gender_detection/age_and_gender_detection
```

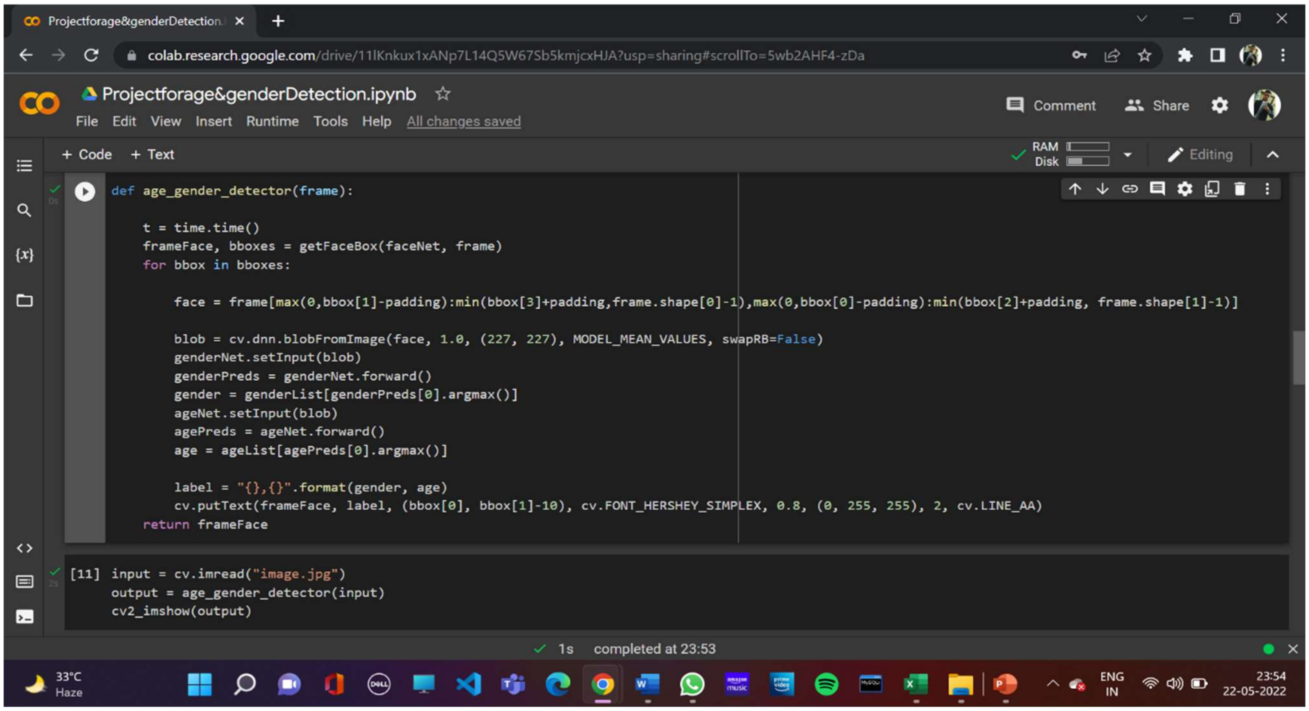
```
[9] !gdown https://drive.google.com/uc?id=1_aDScOvBeBLCn_iv0oxSO8X1ySQpSbIS
    !unzip modelNweight.zip

    Downloading...
    From: https://drive.google.com/uc?id=1_aDScOvBeBLCn_iv0oxSO8X1ySQpSbIS
    To: /content/age_and_gender_detection/age_and_gender_detection/modelNweight.zip
    100% 86.2M/86.2M [00:00<00:00, 211MB/s]
    Archive:  modelNweight.zip
       creating: modelNweight/
      inflating: modelNweight/age_deploy.prototxt
      inflating: modelNweight/age_net.caffemodel
      inflating: modelNweight/gender_deploy.prototxt
      inflating: modelNweight/gender_net.caffemodel
```


Projectforage&genderDetection.i
colab.research.google.com/drive/11lKnkux1xANp7L14Q5W67Sb5kmjcxHJA?usp=sharing#scrollTo=5wb2AHF4-zDa
Projectforage&genderDetection.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code
+ Text
RAM
Disk
Editing
1s completed at 23:53
33°C
Haze
ENG
IN
23:54
22-05-2022

```python
import cv2 as cv
import math
import time
from google.colab.patches import cv2_imshow

def getFaceBox(net, frame, conf_threshold=0.7):
    frameOpencvDnn = frame.copy()
    frameHeight = frameOpencvDnn.shape[0]
    frameWidth = frameOpencvDnn.shape[1]
    blob = cv.dnn.blobFromImage(frameOpencvDnn, 1.0, (300, 300), [104, 117, 123], True, False)

    net.setInput(blob)
    detections = net.forward()
    bboxes = []
    for i in range(detections.shape[2]):
        confidence = detections[0, 0, i, 2]
        if confidence > conf_threshold:
            x1 = int(detections[0, 0, i, 3] * frameWidth)
            y1 = int(detections[0, 0, i, 4] * frameHeight)
            x2 = int(detections[0, 0, i, 5] * frameWidth)
            y2 = int(detections[0, 0, i, 6] * frameHeight)
            bboxes.append([x1, y1, x2, y2])
            cv.rectangle(frameOpencvDnn, (x1, y1), (x2, y2), (0, 255, 0), int(round(frameHeight/150)), 8)
    return frameOpencvDnn, bboxes
```

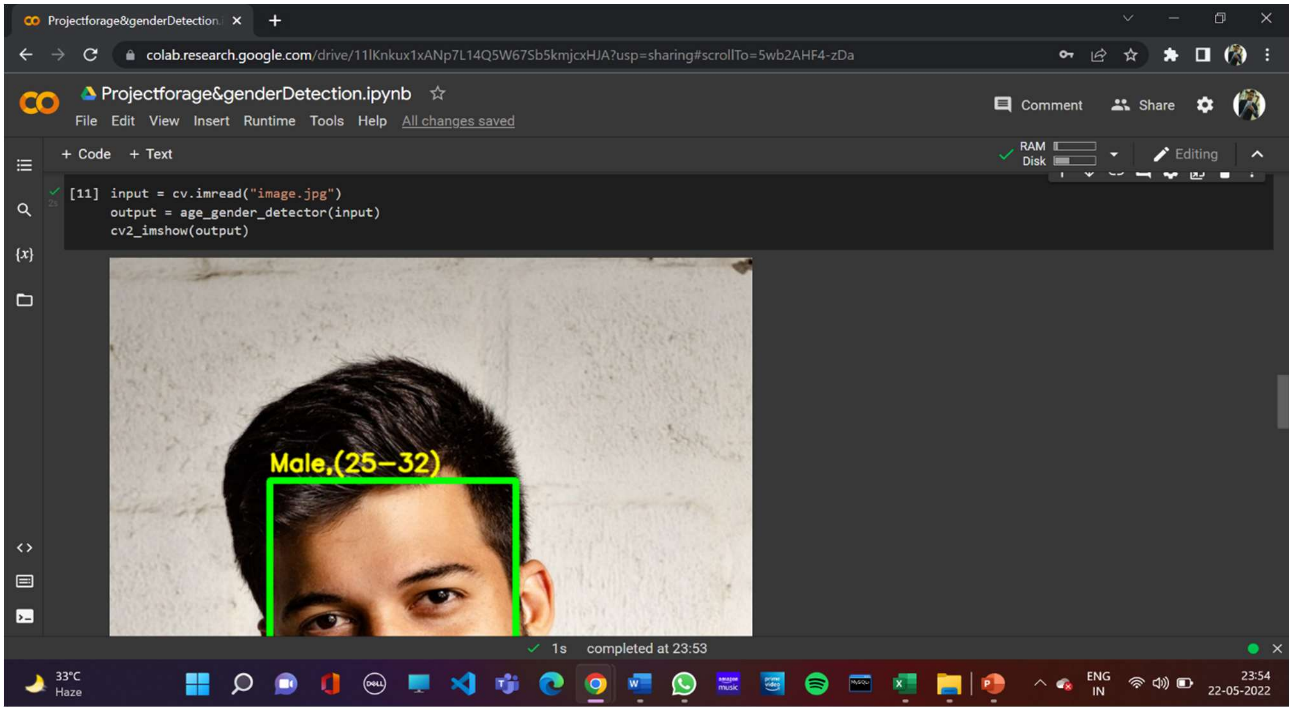

Projectforage&genderDetection.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
Comment Share Editing
1s completed at 23:53
33°C Haze
ENG IN 23:54 22-05-2022

```
        cv.rectangle(frameOpencvDnn, (x1, y1), (x2, y2), (0, 255, 0), int(round(frameHeight/150)), 8)
    return frameOpencvDnn, bboxes

faceProto = "modelNweight/opencv_face_detector.pbtxt"
faceModel = "modelNweight/opencv_face_detector_uint8.pb"

ageProto = "modelNweight/age_deploy.prototxt"
ageModel = "modelNweight/age_net.caffemodel"

genderProto = "modelNweight/gender_deploy.prototxt"
genderModel = "modelNweight/gender_net.caffemodel"

MODEL_MEAN_VALUES = (78.4263377603, 87.7689143744, 114.895847746)
ageList = ['(0-2)', '(4-6)', '(8-12)', '(15-20)', '(25-32)', '(38-43)', '(48-53)', '(60-100)']
genderList = ['Male', 'Female']


ageNet = cv.dnn.readNet(ageModel, ageProto)
genderNet = cv.dnn.readNet(genderModel, genderProto)
faceNet = cv.dnn.readNet(faceModel, faceProto)

padding = 20

def age_gender_detector(frame):
```


Projectforage&genderDetection.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
Comment Share Editing
1s completed at 23:53
33°C Haze
ENG IN 23:54 22-05-2022

```
def age_gender_detector(frame):

    t = time.time()
    frameFace, bboxes = getFaceBox(faceNet, frame)
    for bbox in bboxes:

        face = frame[max(0,bbox[1]-padding):min(bbox[3]+padding,frame.shape[0]-1),max(0,bbox[0]-padding):min(bbox[2]+padding, frame.shape[1]-1)]

        blob = cv.dnn.blobFromImage(face, 1.0, (227, 227), MODEL_MEAN_VALUES, swapRB=False)
        genderNet.setInput(blob)
        genderPreds = genderNet.forward()
        gender = genderList[genderPreds[0].argmax()]
        ageNet.setInput(blob)
        agePreds = ageNet.forward()
        age = ageList[agePreds[0].argmax()]

        label = "{},{}".format(gender, age)
        cv.putText(frameFace, label, (bbox[0], bbox[1]-10), cv.FONT_HERSHEY_SIMPLEX, 0.8, (0, 255, 255), 2, cv.LINE_AA)
    return frameFace
```

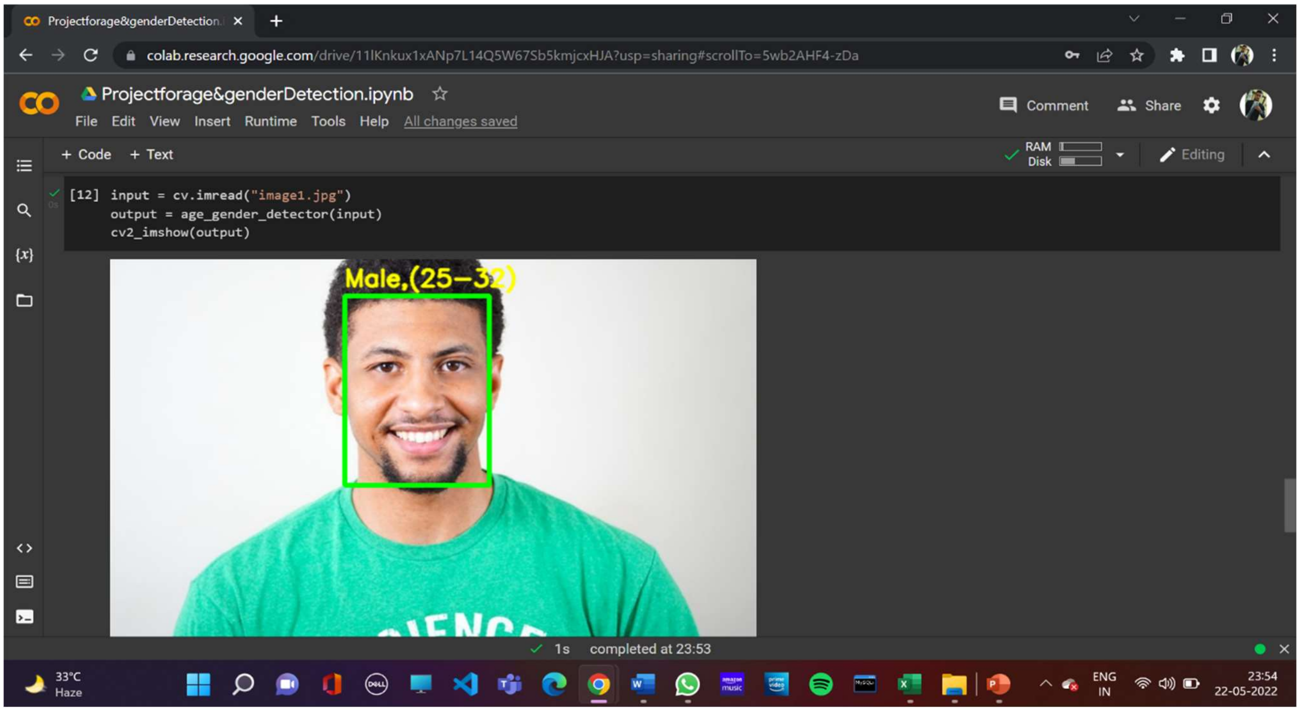
```
[11] input = cv.imread("image.jpg")
     output = age_gender_detector(input)
     cv2_imshow(output)
```



Projectforage&genderDetection.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
Editing
Male,(25–32)

```
[11] input = cv.imread("image.jpg")
     output = age_gender_detector(input)
     cv2_imshow(output)
```

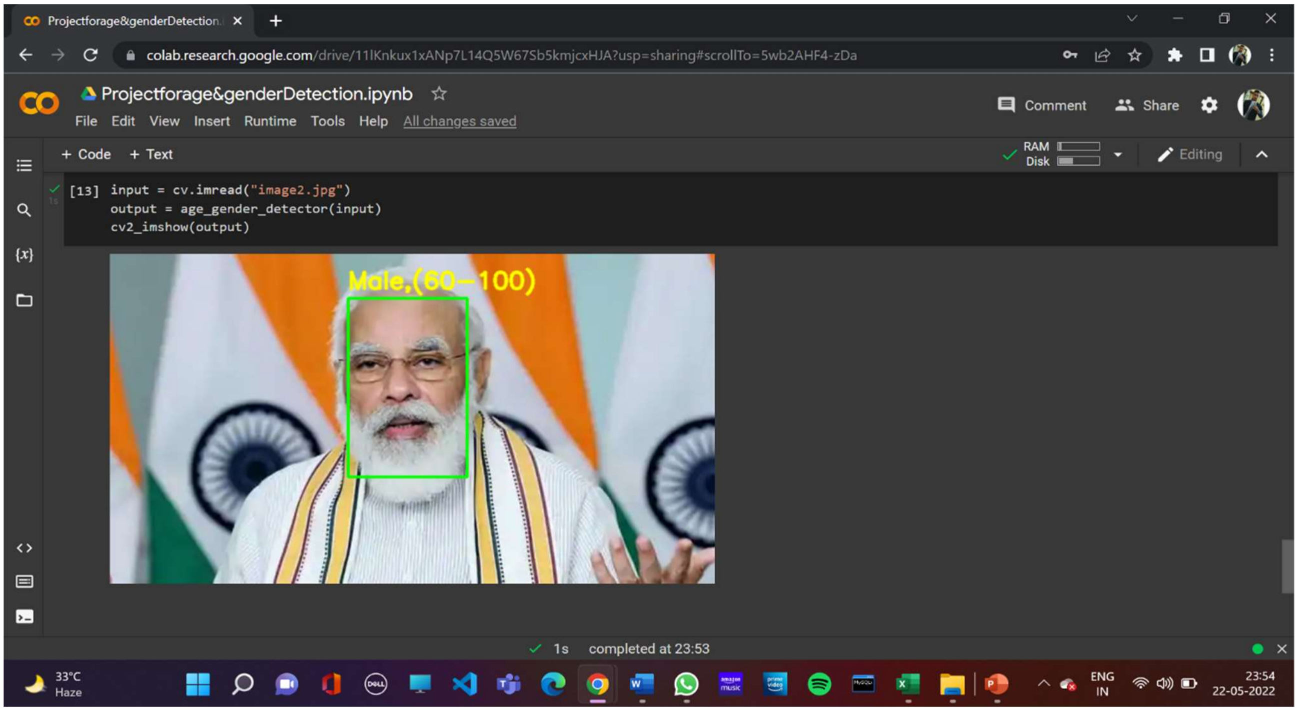
1s completed at 23:53



Projectforage&genderDetection.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
Editing
Male,(25–32)

```
[12] input = cv.imread("image1.jpg")
     output = age_gender_detector(input)
     cv2_imshow(output)
```

1s completed at 23:53



Projectforage&genderDetection.
colab.research.google.com/drive/11lKnkux1xANp7L14Q5W67Sb5kmjcxHJA?usp=sharing#scrollTo=5wb2AHF4-zDa
Projectforage&genderDetection.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Editing
Male,(60–100)
1s completed at 23:53
33°C
Haze
ENG
IN
23:54
22-05-2022

```
[13] input = cv.imread("image2.jpg")
     output = age_gender_detector(input)
     cv2_imshow(output)
```

Project Outcome: -

From this project, we learnt to describe a flow process for data science problems and classified data science problems into standard typology. We also learnt about correlating results to the solution approach followed and assessing the solution approach.

Project Conclusion: -

In this work, it is been concluded that detection of age and gender take consideration of research few years ago. In this work, technique of morphological and SIFT is applied to search key features from the images. The key features of the images are the color and texture of the image. The simulation results shows that proposed algorithm performed well in terms of fault detection rate and accuracy. In future, further improvement will be done in proposed work for iris reorganization for batter reorganization