

# PROJECT REPORT

MAJOR-PROJECT-1

GENDER CLASSIFICATION

DATA SCIENCE AND MACHINE LEARNING

Name: YASH YADAV

College: VELLORE INSTITUTE OF TECHNOLOGY BHOPAL

Programme: B-TECH COMPUTER SCIENCE AND ENGINEERING (CSE)

Branch: CORE

Year: (2021-2025)

(2<sup>nd</sup> Year)

# SCREENSHOTS

The screenshot shows a Google Colab notebook titled "Untitled4.ipynb". The first code cell, labeled "#1. Take the data and create dataframe", contains the following Python code:

```
import pandas as pd
df = pd.read_csv('/content/gender_classification_v7.csv')
df
```

The output of the code is a preview of the DataFrame, showing 5001 rows and 8 columns. The columns are: long\_hair, forehead\_width\_cm, forehead\_height\_cm, nose\_wide, nose\_long, lips\_thin, distance\_nose\_to\_lip\_long, and gender. The gender column contains values "Male" and "Female".

	long_hair	forehead_width_cm	forehead_height_cm	nose_wide	nose_long	lips_thin	distance_nose_to_lip_long	gender
0	1	11.8	6.1	1	0	1		1 Male
1	0	14.0	5.4	0	0	1		0 Female
2	0	11.8	6.3	1	1	1		1 Male
3	0	14.4	6.1	0	1	1		1 Male
4	1	13.5	5.9	0	0	0		0 Female
...	...	...	...	...	...	...	...	...
4996	1	13.6	5.1	0	0	0		0 Female
4997	1	11.9	5.4	0	0	0		0 Female
4998	1	12.9	5.7	0	0	0		0 Female
4999	1	13.2	6.2	0	0	0		0 Female
5000	1	15.4	5.4	1	1	1		1 Male

The notebook interface shows the file "gender\_classification\_v7.csv" is located in the "/content" directory. The status bar at the bottom indicates the code was completed at 10:47 PM.

The screenshot shows the same Google Colab notebook, now at the second code cell, labeled "#2. Exploratory Data Analysis". The code executed is:

```
df.info()
```

The output of `df.info()` is as follows:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5001 entries, 0 to 5000
Data columns (total 8 columns):
 #   column              Non-Null Count  Dtype
---  -
 0   long_hair            5001 non-null   int64
 1   forehead_width_cm    5001 non-null   float64
 2   forehead_height_cm   5001 non-null   float64
 3   nose_wide            5001 non-null   int64
 4   nose_long            5001 non-null   int64
 5   lips_thin            5001 non-null   int64
 6   distance_nose_to_lip_long 5001 non-null   int64
 7   gender               5001 non-null   object
dtypes: float64(2), int64(5), object(1)
memory usage: 312.7+ KB
```

The next code cell, labeled "[3] df.isnull().sum()", shows the output:

```
long_hair            0
forehead_width_cm    0
forehead_height_cm   0
nose_wide            0
nose_long            0
lips_thin            0
distance_nose_to_lip_long 0
gender               0
dtype: int64
```

The final code cell, labeled "[4] #3. Visualization- not required", is currently empty. The notebook interface shows that all changes have been saved. The status bar at the bottom indicates the code was completed at 10:47 PM.

```
Untitled4.ipynb - Colaboratory
colab.research.google.com/drive/1SGc6k8gSuq4Pvu7FSinjQZevMqfqZyCd#scrollTo=WMOnyvfg0xr4

Untitled4.ipynb
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text
[4] #3.visualization- not required
[5] #4.Divide the data into input and output
x = df.iloc[:,0:7].values
y = df.iloc[:,7].values
[6] #5.Train and Test variables
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,random_state = 0)
[7] #6.Normalization- not required
[8] #7.Apply a CLASSIFIER
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
[9] #8.Fitting the model
model.fit(x_train,y_train)
LogisticRegression()
[10] #9.PREDICT THE OUTPUT
y_pred = model.predict(x_test)
y_pred #PREDICTED OUTPUT VALUES
array(['Female', 'Male', 'Male', ..., 'Male', 'Male', 'Female'],
      dtype=object)
[11] v. test #ACTUAL OUTPUT VALUES
0s completed at 10:47 PM
```

```
Untitled4.ipynb - Colaboratory
colab.research.google.com/drive/1SGc6k8gSuq4Pvu7FSinjQZevMqfqZyCd#scrollTo=WMOnyvfg0xr4

Untitled4.ipynb
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text
dtype=object)
[11] y_test #ACTUAL OUTPUT VALUES
array(['Female', 'Male', 'Male', ..., 'Male', 'Male', 'Female'],
      dtype=object)
[12] #10.Accuracy
from sklearn.metrics import accuracy_score
accuracy_score(y_pred,y_test)*100
96.72262190247881
[12]
```

# GITHUB

GITHUB LINK

[https://github.com/yashyadav02/Rinex\\_Major\\_Project\\_1.git](https://github.com/yashyadav02/Rinex_Major_Project_1.git)