



Gujarat Technological University

Chandkheda, Ahmedabad



G H Patel College of Engineering and Technology

Bakrol gate, Vallabh Vidhyanagar, Anand (Affiliated with GTU)

A Mini Project Report on

“Placement Prediction”

Under the subject

Summer Internship (3170001)

Department of Information Technology

Submitted By

1 Gajera Jagdishkumar

190110116019

Academic Year

(2022-2023)

Internal Guide

Prof. Rahul Patel

Head of the Department

Dr. Nikhil Gondaliya



G H Patel College of Engineering and Technology

Bakrol gate, Vallabh Vidhyanagar, Anand (Affiliated with GTU)

DEPARTMENT OF INFORMATION TECHNOLOGY

CERTIFICATE

This is to certify that the mini project on **Placement Prediction** has been satisfactorily carried by **Gajera Jagdishkumar** (190110116019) under my guidance in the fulfillment of the course Summer Internship (3170001) work during the academic year 2022-2023.

Internal Guide

Prof. Rahul Patel

Head of the Department

Dr. Nikhil Gondaliya

Acknowledgement

We wish to express our sincere gratitude to our project guide Prof. **Rahul Patel** and all the faculty members for helping us through our project by giving us the necessary suggestions and advices along with their valuable co-ordination in completing this work.

We also thank our parents, friends and all the members of the family for their precious support and encouragement which they had provided in completion of our work. In addition to that, we would also like to mention the college personals who gave us the permission to use and experience the valuable resources required for the project from the college premises.

Thus, In conclusion to the above said, we once again thank the faculties and members of **G H Patel College of Engineering and Technology** for their valuable support in completion of the project.

Thank You

Gajera Jagdishkumar

190110116019

INDEX

Certificate.....	2
Acknowledgement.....	3
List of Figures.....	5
1. Abstract.....	6
2. Introduction.....	7
3. Design methodology.....	8
3.1 Data Gathering.....	8
3.2 Pre-Processing.....	9
3.3 Processed data.....	11
4. Implementation.....	12
5. Conclusion.....	17
6. Bibliography.....	18

List of Figures

1. Design and Methodology.....	8
2. Trainnig and Testing.....	9
3. Implementation.....	12
Figure 1.....	12
Figure 2.....	12
Figure 3.....	13
Figure 4.....	13
Figure 5.....	14
Figure 6.....	14
Figure 7.....	15
Figure 8.....	15
Figure 9.....	16
Figure 10.....	16

Chapter 1- Abstract

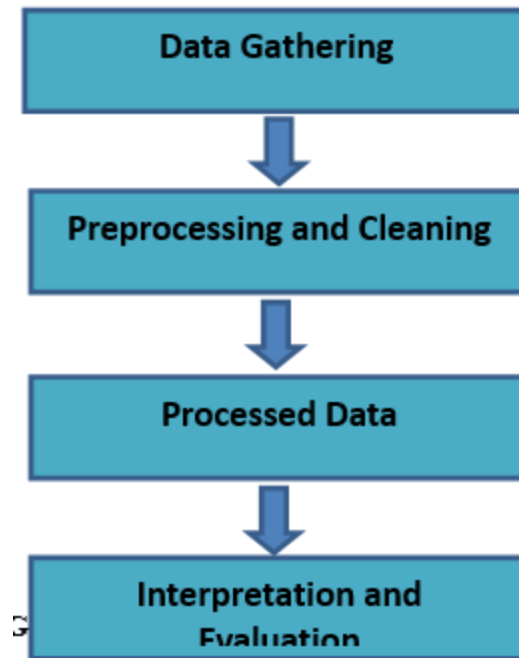
- Placement of students is one of the most important objectives of an educational institution. Reputation and yearly admissions of an institution invariably depend on the placements it provides its students with. That is why all the institutions, arduously, strive to strengthen their placement department so as to improve their institution on a whole. Any assistance in this particular area will have a positive impact on an institution's ability to place its students. This will always be helpful to both the students, as well as the institution.
- In this study, the objective is to analyze previous year's student's data and use it to predict the placement chance of the current students.
- This model is proposed with an algorithm to predict the same. Data pertaining to the study were collected from the same institution for which the placement prediction is done, and also suitable data pre-processing methods were applied. This proposed model is also compared with other traditional classification algorithms such as Decision tree and Random Forest with respect to accuracy, precision and recall.
- From the results obtained it is found that the proposed algorithm performs significantly better in comparison with the other algorithms mentioned.

Chapter 2- Introduction

- Placements are considered to be very important for each and every college. The basic success of the college is measured by the campus placement of the students. Every student takes admission to the colleges by seeing the percentage of placements in the college.
- Hence, in this regard the approach is about the prediction and analyses for the placement necessity in the colleges that helps to build the colleges as well as students to improve their placements.
- In Placement Prediction system predicts the probability of an undergrad students getting placed in a company by applying classification algorithms such as Decision tree and Random Forest.
- The main objective of this model is to predict whether the student he/she gets placed or not in campus recruitment. For this the data consider is the academic history of student like overall percentage, backlogs, credits. The algorithms are applied on the previous year's data of the students.

Chapter 3- Design Methodology

The whole approach is depicted by the following flowchart.



1. Data Gathering:

In this phase gather the data from colleges. The combination of various attributes determines whether the student is placed or not. There are many places to find your free datasets for your projects. Like Kaggle, Earth Data, Data.Gov, etc. For my project I took dataset from Kaggle.

The dataset contains various information about the students.

- I. Age
- II. Gender
- III. Stream
- IV. History of backlogs
- V. Ssc_p
- VI. Hsc_p
- VII. Hostel
- VIII. Internships

IX. CGPA

X. Placed or not

	A	B	C	D	E	F	G	H	I	J	K
1	Age	Gender	Stream	ssc_p	hsc_p	Hostel	HistoryOfE	Internship	CGPA	PlacedOrNot	
2	22	Male	Electronics	67	91	1	1	1	8	1	
3	21	Female	Computer	79.33	78.33	1	1	0	7	1	
4	22	Female	Informatic	65	68	0	0	1	6	1	
5	21	Male	Informatic	56	52	0	1	0	8	1	
6	22	Male	Mechanica	85.8	73.6	1	0	0	8	1	
7	22	Male	Electronics	55	49.8	0	0	0	6	0	
8	21	Male	Computer	46	49.2	0	1	0	7	0	
9	21	Male	Informatic	82	64	0	0	1	7	0	
10	21	Male	Computer	73	79	0	0	2	6	1	
11	21	Female	Computer	58	70	1	0	1	6	0	
12	22	Male	Computer	58	61	0	0	1	7	0	
13	22	Female	Electrical	69.6	68.4	0	1	1	8	1	
14	21	Female	Computer	47	55	1	1	2	6	0	
15	21	Male	Computer	77	87	0	1	1	8	1	
16	21	Female	Electronics	62	47	0	0	2	8	1	
17	22	Male	Mechanica	65	75	1	0	0	8	1	
18	22	Female	Computer	63	66.2	1	0	1	7	0	
19	21	Male	Civil	55	67	0	0	0	6	0	
20	21	Male	Computer	63	66	0	0	1	7	0	
21	22	Male	Computer	60	67	0	0	1	7	0	
22	22	Male	Civil	62	65	0	0	0	8	1	
23	21	Male	Computer	79	76	0	0	1	6	0	
24	22	Male	Civil	69.8	60.8	0	0	0	7	0	
25	22	Male	Electronics	77.4	60	0	1	1	7	0	
26	22	Male	Electronics	76.5	97.7	0	0	0	7	0	
27	21	Male	Electrical	52.58	54.6	1	0	0	8	1	

collegePlace (+)

Ready Accessibility: Unavailable

Dataset

2. Pre-Processing and Cleaning:

When it comes to creating a Machine Learning model, data preprocessing is the first step marking the initiation of the process. Data preprocessing in Machine Learning refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for a building and training Machine Learning models. In simple words, data preprocessing in Machine Learning is a data mining technique that transforms raw data into an understandable and readable format.

Steps in Data Preprocessing:

1. Import all the crucial libraries:

Python is a very powerful and flexible programming language. This is an open-source language has created quite a few tools to efficiently work with a python. There are some basic libraries which are essential for building a project:

- Pandas: Pandas is used for data manipulation and analysis.
- Seaborn: It is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
- Scikit-learn: It is used to build a machine learning model. It is built on NumPy, SciPy, and matplotlib.
- NumPy: NumPy is a python library also called as Numeric python which can execute scientific computing. with the help of a NumPy library you can create and perform manipulations on an array.

2. Import the dataset:

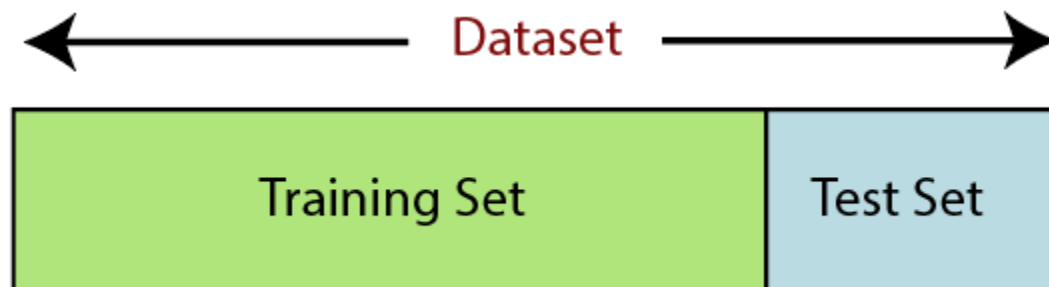
In this step, you need to import the dataset/s that you have gathered for the ML project at hand. Importing the dataset is one of the important steps in data preprocessing in machine learning.

3. Encoding the categorical data

In this step, Machine Learning techniques transforms raw data into an understandable and readable format. Like example in our project Gender and Stream causes problem; So we must convert in numerical values.

4. Splitting the dataset

It is the next step in data preprocessing in machine learning. Every dataset for Machine Learning model must be split into two separate sets – training set and test set.



Training & Testing (2)

3. Processed Data:

Processing is a method in which different algorithms applied on data to find the best results:

1. Decision Tree algorithm
2. Random Forest algorithm
3. Logistic Regression algorithm
4. K-neighbors algorithm
5. Support Vector Machine (SVM) algorithm

Chapter 4 – Implementation

The screenshot shows a Google Colab notebook titled 'MINIPROJECT.ipynb'. The interface includes a menu bar (File, Edit, View, Insert, Runtime, Tools, Help) and a toolbar with options like 'Code' and 'Text'. The notebook content is divided into two sections: 'IMPORT LIBRARIES' and 'IMPORT DATASET'.

IMPORT LIBRARIES

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import sklearn
import seaborn as sns

from google.colab import drive
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

IMPORT DATASET

```
df = pd.read_csv('/content/drive/MyDrive/data/science/miniproject/collegePlace.csv')
df
```

The output of the dataset import is a DataFrame with the following columns: Age, Gender, Stream, ssc_p, hsc_p, Hostel, HistoryOfBacklogs, Internships, CGPA, and PlacedOrNot. The first few rows of data are shown below:

	Age	Gender	Stream	ssc_p	hsc_p	Hostel	HistoryOfBacklogs	Internships	CGPA	PlacedOrNot
0	22	Male	Electronics And Communication	67.00	91.00	1	1	1	8	1
1	21	Female	Computer Science	79.33	78.33	1	1	0	7	1
2	22	Female	Information Technology	65.00	68.00	0	0	1	6	1
3	21	Male	Information Technology	56.00	52.00	0	1	0	8	1
4	22	Male	Mechanical	85.80	73.60	1	0	0	8	1

Figure (1)

The screenshot shows the same Google Colab notebook, now displaying the 'TOTAL NUMBER OF FEATURES AND ATTRIBUTES IN DATASET' section. The code executed includes:

```
df.shape
df.size
df.head()
df.tail()
```

The output of `df.shape` is `(2966, 10)`, indicating 2966 rows and 10 columns. The output of `df.size` is `29660`. The `df.head()` and `df.tail()` outputs show the first and last rows of the dataset, respectively, with the same column structure as in Figure 1.

Figure (2)

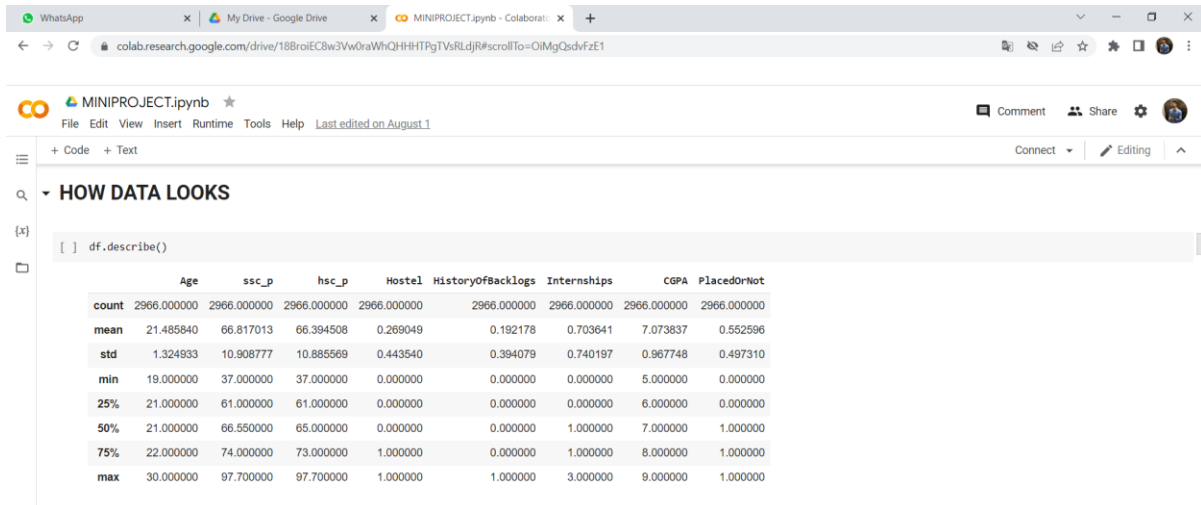


Figure (3)

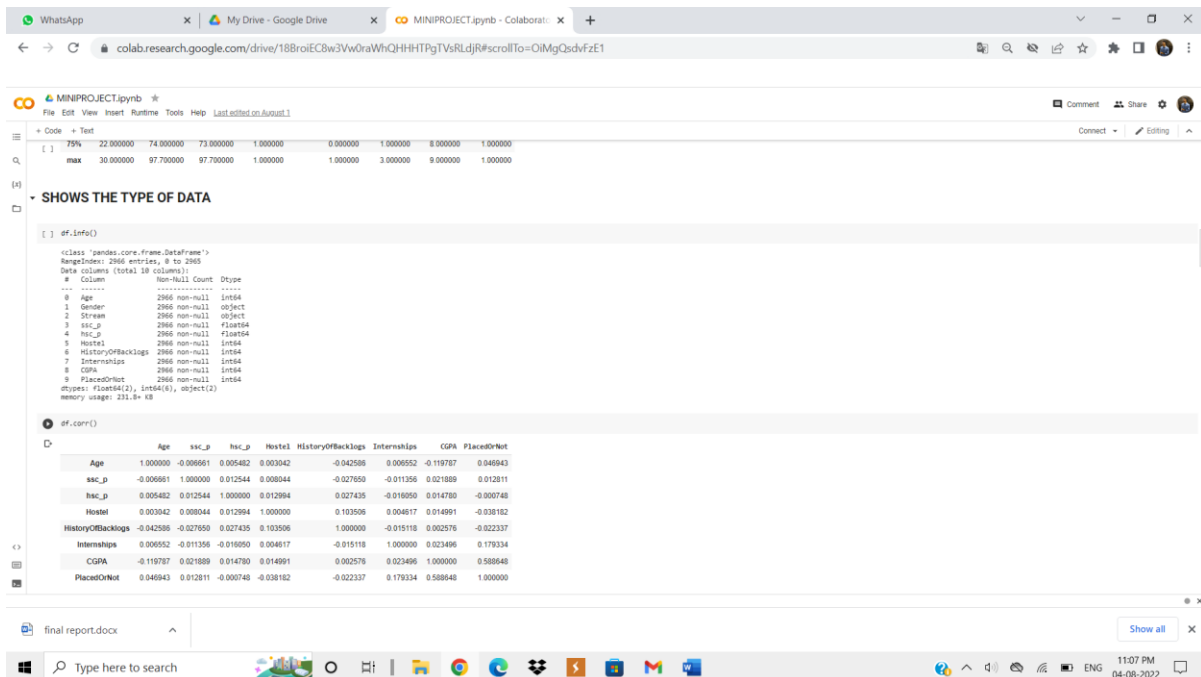


Figure (4)

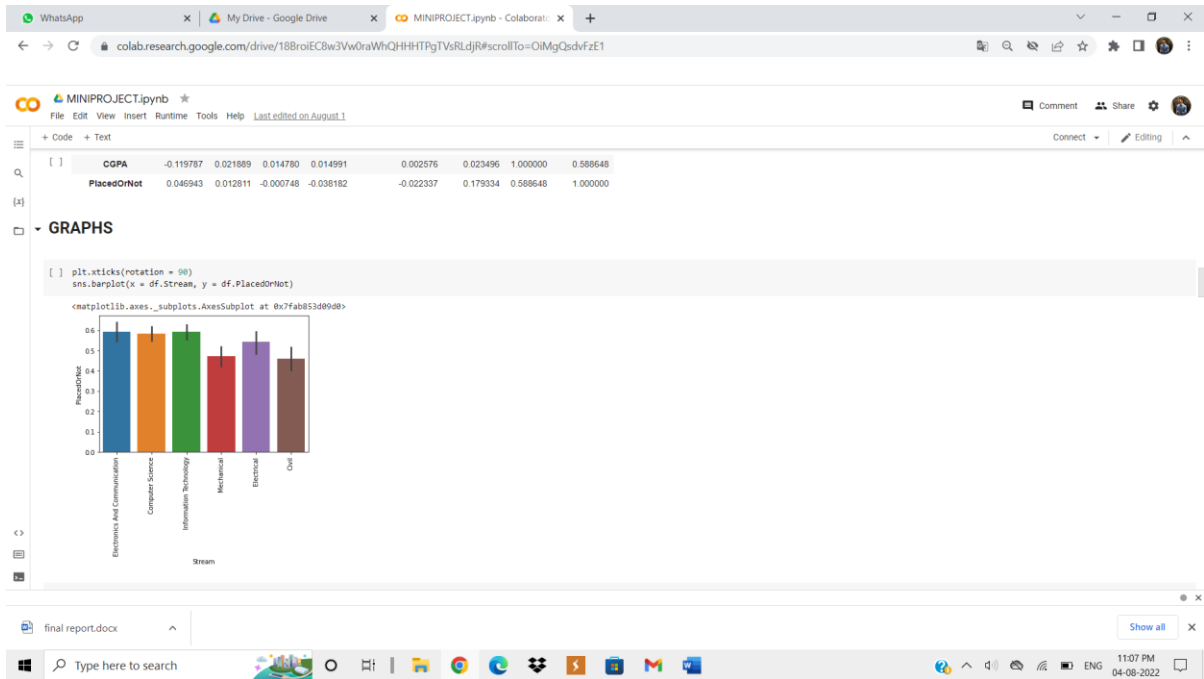


Figure (5)

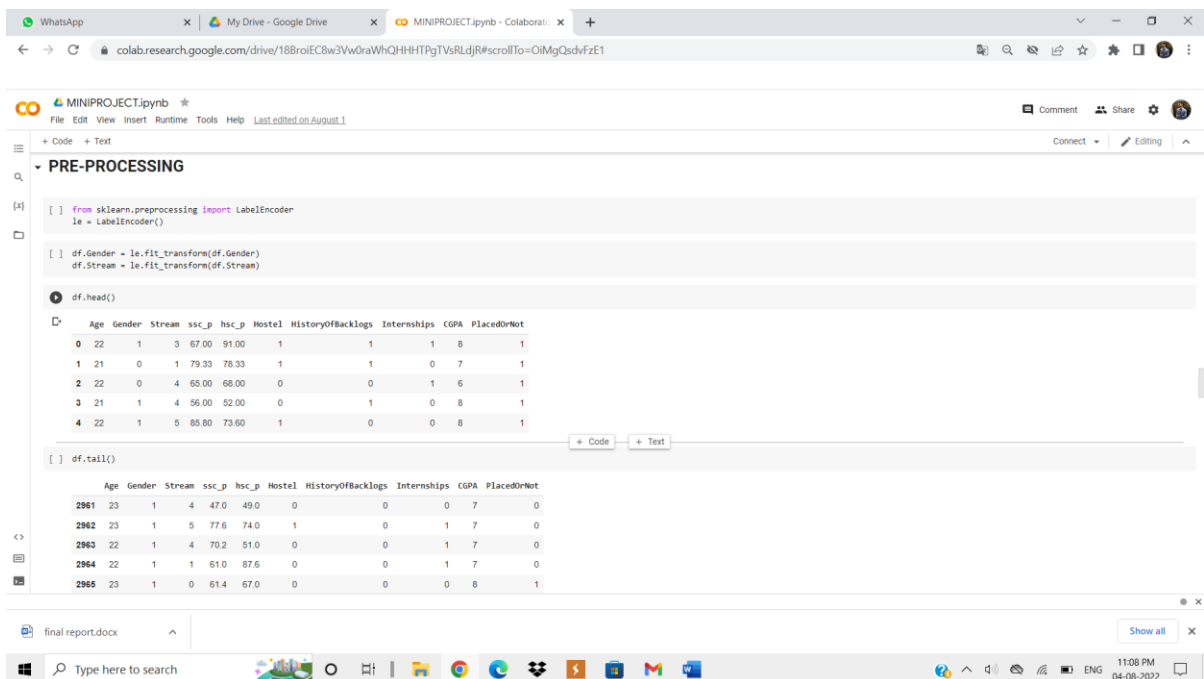
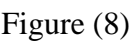
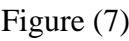
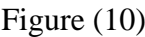
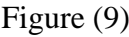


Figure (6)





Chapter 5 – Conclusion

- This system is beneficial for institutions to predict student's campus placement and placement officers can work on identifying the weakness of each student.
- They can also suggest improvements so that the student can overcome the weakness and supply to the best of their abilities.
- Algorithms like random forest and decision tree will give maximum accuracy to the prediction.

Chapter 6 – Bibliography

<https://www.kaggle.com/>

<http://ijetmr.org/2020/06/21/campus-placement-prediction-using-data-science/>

[https://www.google.com/imgres?imgurl=https%3A%2F%2Fstatic.javatpoint.com%2Ftutorial%2Fmachine-learning%2Fimages%2Ftrain-and-test-datasets-in-machine-](https://www.google.com/imgres?imgurl=https%3A%2F%2Fstatic.javatpoint.com%2Ftutorial%2Fmachine-learning%2Fimages%2Ftrain-and-test-datasets-in-machine-learning.png&imgrefurl=https%3A%2F%2Fwww.javatpoint.com%2Ftrain-and-test-datasets-in-machine-learning&tbnid=zmljNn8meUmqMM&vet=12ahUKEwihlK3D1rH5AhW__jgGHdXdBE4QMygvegUIARCeAg..i&docid=qWjWXh4ZYsIC7M&w=550&h=178&q=training%20and%20testing%20dataset&ved=2ahUKEwihlK3D1rH5AhW__jgGHdXdBE4QMygvegUIARCeAg)

[learning.png&imgrefurl=https%3A%2F%2Fwww.javatpoint.com%2Ftrain-and-test-datasets-in-machine-](https://www.google.com/imgres?imgurl=https%3A%2F%2Fstatic.javatpoint.com%2Ftutorial%2Fmachine-learning%2Fimages%2Ftrain-and-test-datasets-in-machine-learning.png&imgrefurl=https%3A%2F%2Fwww.javatpoint.com%2Ftrain-and-test-datasets-in-machine-learning&tbnid=zmljNn8meUmqMM&vet=12ahUKEwihlK3D1rH5AhW__jgGHdXdBE4QMygvegUIARCeAg..i&docid=qWjWXh4ZYsIC7M&w=550&h=178&q=training%20and%20testing%20dataset&ved=2ahUKEwihlK3D1rH5AhW__jgGHdXdBE4QMygvegUIARCeAg)

[learning&tbnid=zmljNn8meUmqMM&vet=12ahUKEwihlK3D1rH5AhW__jgGHdXdBE4QMygvegUIARCeAg..i&docid=qWjWXh4ZYsIC7M&w=550&h=178&q=training%20and%20testing%20dataset&ved=2ahUKEwihlK3D1rH5AhW__jgGHdXdBE4QMygvegUIARCeAg](https://www.google.com/imgres?imgurl=https%3A%2F%2Fstatic.javatpoint.com%2Ftutorial%2Fmachine-learning%2Fimages%2Ftrain-and-test-datasets-in-machine-learning.png&imgrefurl=https%3A%2F%2Fwww.javatpoint.com%2Ftrain-and-test-datasets-in-machine-learning&tbnid=zmljNn8meUmqMM&vet=12ahUKEwihlK3D1rH5AhW__jgGHdXdBE4QMygvegUIARCeAg..i&docid=qWjWXh4ZYsIC7M&w=550&h=178&q=training%20and%20testing%20dataset&ved=2ahUKEwihlK3D1rH5AhW__jgGHdXdBE4QMygvegUIARCeAg)