

**INDUSTRIAL TRAINING REPORT**  
ON  
**“STUDENT PLACEMENT PREDICTION”**

Submitted in partial fulfilment of the requirement for degree of

**BTECH**  
in  
**COMPUTER SCIENCE & ENGINEERING**  
at  
**ARKA JAIN UNIVERSITY, Jharkhand**

**Submitted By**  
**ABHISHEK**  
**AJU/191607**

**Under the Guidance of**

**JOYJIT GUHA**  
**(Project Engineer, Ardent Computech Private Limited)**



**SCHOOL OF ENGINEERING & I.T,**  
**ARKA JAIN UNIVERSITY, JAMSHEDPUR**  
**2020-2023**

# Industrial Internship Certificate



This certificate is awarded to

**ABHISHEK**

of

ARKA JAIN UNIVERSITY

for successfully completing the **Industrial Internship** on

**AI DATA SCIENCE AND MACHINE LEARNING USING PYTHON**

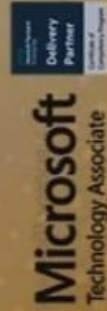
from

10.06.2022 to 09.07.2022

and implementing the project titled

**STUDENT PLACEMENT PREDICTION**

Certificate ID: ARDENT/2022/AD4230  
Issue Date: 09.07.2022



Director  
Technology Services

Director  
Operations





**ARKA JAIN  
University**  
Jharkhand

## CERTIFICATE

This is to certify that the Industrial Training Report titled "Student Placement Prediction" in partial fulfillment of the requirement for the award of the Degree of Btech in Computer Science and Engineering, submitted to ARKA JAIN University, Jharkhand, is an authentic record of bonafide industrial training work carried out by Mr. Joyjit Guha and submitted under my supervision/guidance.

*Joyjit Guha*

**Mr Joyjit Guha**  
Project Engineer  
Ardent Private Limited



*Ashwin*

**Mr Ashwin Kumar**  
Assistant Dean  
School of Engineering



**Date:**  
**Place:**

## ACKNOWLEDGEMENT

I would like to express my utmost gratitude to the AJU for providing an opportunity to pursue the engineering training as partial fulfillment of the requirement for the degree of Btech in Computer Science and Engineering. The internship opportunity I had with **ARDENT COMPUTECH PRIVATE LIMITED** was a great chance for learning and professional development. Therefore, I consider myself as a very lucky individual as I was provided with an opportunity to be a part of it. I am also grateful for having a chance to meet so many wonderful people and professionals who led me through this internship period.

Bearing in mind previous I am using this opportunity to express my deepest gratitude and special thanks to the **Mr. JIT DUTTA**, Head(Training and Operation) of **ARDENT COMPUTECH PRIVATE LIMITED, KOLKATA** who in spite of being extraordinarily busy with his duties, took time out to hear, guide and keep me on the correct path and allowing me to carry out my project at their esteemed organization and extending during the training.

I express my deepest thanks to **Mr. JOYJIT GUHA**, Project Engineer, **ARDENT COMPUTECH PRIVATE LIMITED, KOLKATA** for taking part in useful decision & giving necessary advice and guidance and arranged all facilities to make life easier. I choose this moment to acknowledge his contribution gratefully.

I express my deepest thanks to all staffs and employees of **ARDENT COMPUTECH PRIVATE LIMITED** for taking part in useful decision & giving necessary advice and guidance and arranged all facilities to make life easier. I choose this moment to acknowledge their contribution gratefully. I perceive as this opportunity as a big milestone in my career development. I will strive to use gained skills and knowledge in the best possible way, and I will continue to work on their improvement, in order to attain desired career objectives. Hope to continue cooperation with all of you in the future.

## ABSTRACT

This course dives into the basics of machine learning using an approachable and well-known programming language, Python.

In this course, we will be reviewing two main components: First, you will be learning about the purpose of Machine Learning and where it applies to the real world. Second, you will get a general overview of Machine Learning topics such as supervised vs. unsupervised learning, model evaluation, and Machine Learning algorithms.

By just putting in a few hours a week for the next few weeks, this is what you'll get.

- 1) New skills to add to your resume, such as regression, classification, clustering etc.
- 2) New projects that you can add to your portfolio, including cancer detection, predicting economic trends, predicting customer churn, recommendation engines, and many more.
- 3) And a certificate in machine learning to prove your competency, and share it anywhere you like online or offline, such as LinkedIn profiles and social media.

# TABLE OF CONTENTS

<b>DECLARATION BY THE STUDENT</b>	iii
<b>CERTIFICATE</b>	iv
<b>ACKNOWLEDGEMENT</b>	v
<b>ABSTRACT</b>	vi
<b>TABLE OF CONTENTS</b>	vii
<b>CHAPTER-1 INTRODUCTION</b>	
1.1 About Company	01
1.2 Overview/ features(Python)	02
1.3 Basic Syntax(Python)	03
<b>CHAPTER-2 ARTIFICIAL INTELLIGENCE</b>	
2.1 Basic Introduction	04
2.2 Goals/ Application	05
<b>CHAPTER-3 MACHINE LEARNING</b>	
3.1 Basic Introduction	06
3.2 Supervised VS Unsupervised	07
3.3 Numpy	08
3.4 Scipy	09
3.5 Regression Analysis	10
3.6 Linear, Logistic, Polynomial Regression	11
3.7 Matplotlib	12
3.8 Pandas/ Clustering	13
3.9 Algorithm	14
<b>CHAPTER-4 PROJECT IMPLEMENTATION</b>	
4.1 Importing Packages and Reading Data	15
4.2 Cleaning Data/ Importing Matplotlib	16
4.3 Applying Logistic Regression	17
4.3 Predicting Output	18
<b>CHAPTER-5 CONCLUSION</b>	19

# INTRODUCTION

## ABOUT COMPANY:-

ARDENT-The Passionate Team for your Learning Needs

ARDENT as the name stands, is driven by a passionate team with objective of enhancing the Skilled Human Resources of India. Ardent was an initiative of a group of fresh engineering graduates aiming to transform the higher education methodologies and quality. It started with focus on technology training and project implementation knowledge towards engineering students and job aspirants. The group leveraged on the technical strength and fast growing engineering education growth in India.

Vision: To invent, nurture, and transform talent and educate them to transform the globe.

Mission: To impart training in the area of engineering and management that will increase employability and bridge the gap between academia and corporate.

ARDENT is a well-known name in the Corporate Training, Consulting, Web Application Development and Mobile Application Development Industry. The journey started in the year 2012 and since then ARDENT have served more than 2, 80, 000 students/professionals. ARDENT's training programs target candidates with passion of making their mark in the Technology Industry. We have highly skilled team of Trainers and 24/7 Technical Support. ARDENT is affiliated to different State Govt. bodies, Central Govt. bodies, and International Brands. Learning at ARDENT had always been an exciting journey for the participants.

## OVERVIEW (PYTHON)

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and has fewer syntactical constructions than other languages.

**Python is interpreted:** Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to Perl and PHP.

**Python is Interactive:** You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

**Python is Object-Oriented:** Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

**Python is a Beginner's Language:** Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

## FEATURES OF PYTHON

**Python has a big list of good features, few are listed below:**

- It support functional and structured programming methods as well as OOP.
- It can be used as a scripting language or can be compiled to byte code for building large application
- It provides very high level dynamic data types and supports dynamic type checking.
- It supports automatic garbage collections.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA and JAVA.



## BASIC SYNTAX OF PYTHON PROGRAM

Type the following text at the Python prompt and press the Enter:-

```
>>> print "Hello, Python!"
```

If you are running new version of Python, then you would need to use print statement with parenthesis as in **print ("Hello, Python!");**

However in Python version 2.4.3, this produces the following result  
Hello, Python!

## LINES & INDENTATION

Python provides no braces to indicate blocks of code for class and function definitions or flow control. Blocks of code are denoted by line indentation, which is rigidly enforced.

The number of spaces in the indentation is variable, but all statements within the block must be indented the same amount. For example –

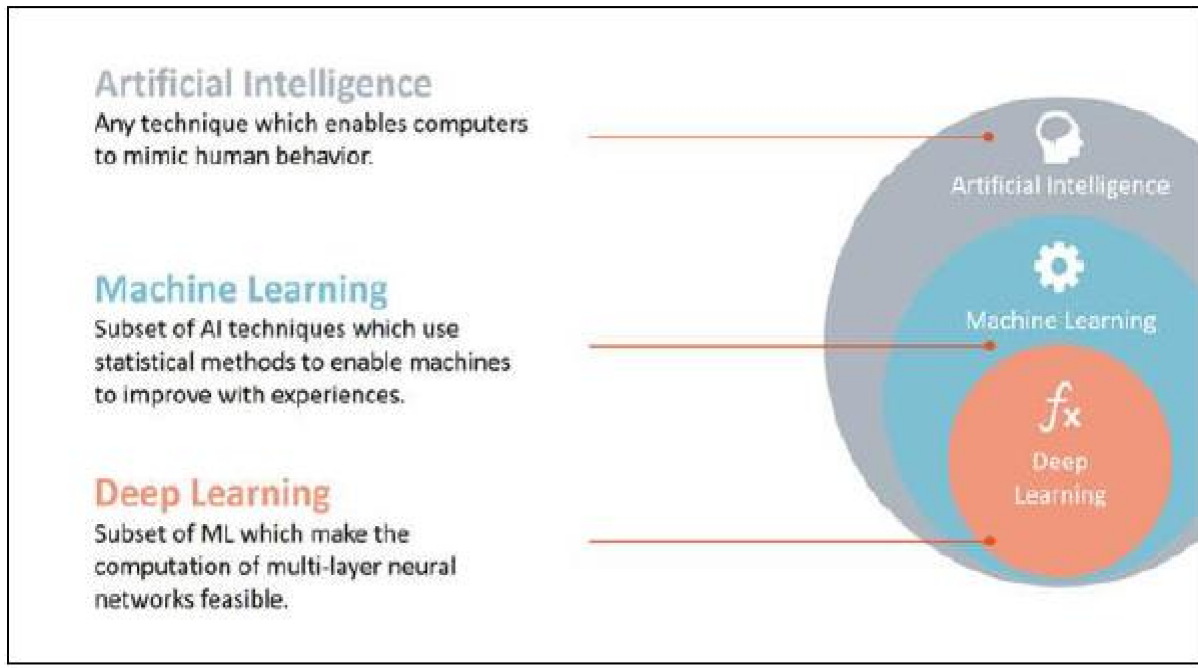
```
if True:
    print -True
else:
    print "False"
```

## STANDARD DATA TYPES

The data stored in memory can be of many types. For example, a person's age is stored as a numeric value and his or her address is stored as alphanumeric characters. Python has five standard data types –

1. String
2. List
3. Tuple
4. Dictionary
5. Number

# ARTIFICIAL INTELLIGENCE



According to the father of Artificial Intelligence, John McCarthy, it is “The science and engineering of making intelligent machines, especially intelligent computer programs”.

Artificial Intelligence is a way of making a computer, a computer-controlled robot, or a software think intelligently, in the similar manner the intelligent humans think. AI is accomplished by studying how human brain thinks and how humans learn, decide, and work while trying to solve a problem, and then using the outcomes of this study as a basis of developing intelligent software and systems.

The development of AI started with the intention of creating similar intelligence in machines that we find and regard high in humans.

## GOALS OF AI

**To Create Expert Systems** – The systems which exhibit intelligent behavior, learn, demonstrate, explain, and advice its users.

**To Implement Human Intelligence in Machines** – Creating systems that understand, think, learn, and behave like humans.

## APPLICATION OF AI

**AI has been dominant in various fields such as:-**

**Gaming** – AI plays crucial role in strategic games such as chess, poker, tic-tac-toe, etc., where machine can think of large number of possible positions based on heuristic knowledge

**Natural Language Processing** – It is possible to interact with the computer that understands natural language spoken by humans.

**Expert Systems** – There are some applications which integrate machine, software, and special information to impart reasoning and advising. They provide explanation and advice to the users

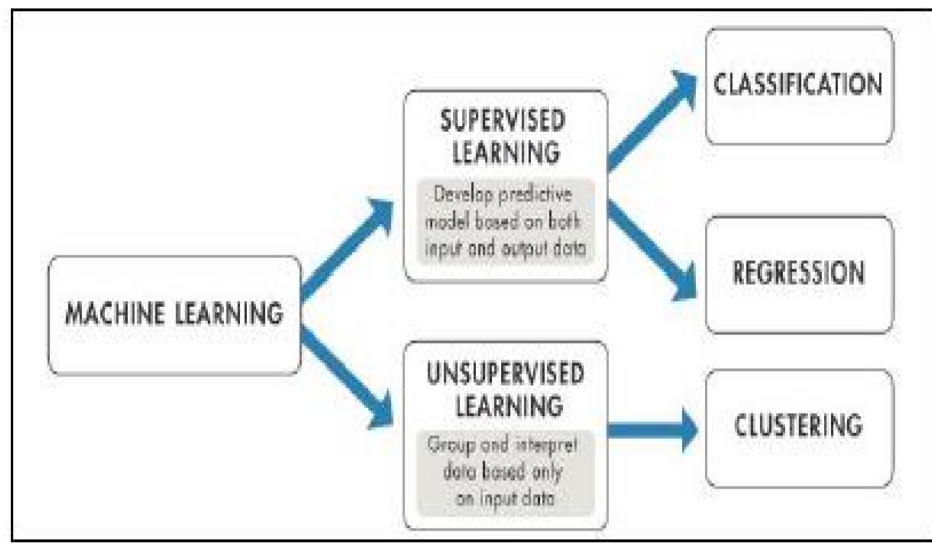
**Vision Systems** – These systems understand, interpret, and comprehend visual input on the computer.

**Speech Recognition** – Some intelligent systems are capable of hearing and comprehending the language in terms of sentences and their meanings while a human talks to it. It can handle different accents, slang words, noise in the background, change in human's voice due to cold, etc.

**Handwriting Recognition** – The handwriting recognition software reads the text written on paper by a pen or on screen by a stylus. It can recognize the shapes of the letters and convert it into editable text.

**Intelligent Robots** – Robots are able to perform the tasks given by a human. They have sensors to detect physical data from the real world such as light, heat, temperature, movement, sound, bump, and pressure. They have efficient processors, multiple sensors and huge memory, to exhibit intelligence. In addition, they are capable of learning from their mistakes and they can adapt to the new environment.

# MACHINE LEARNING



Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed.

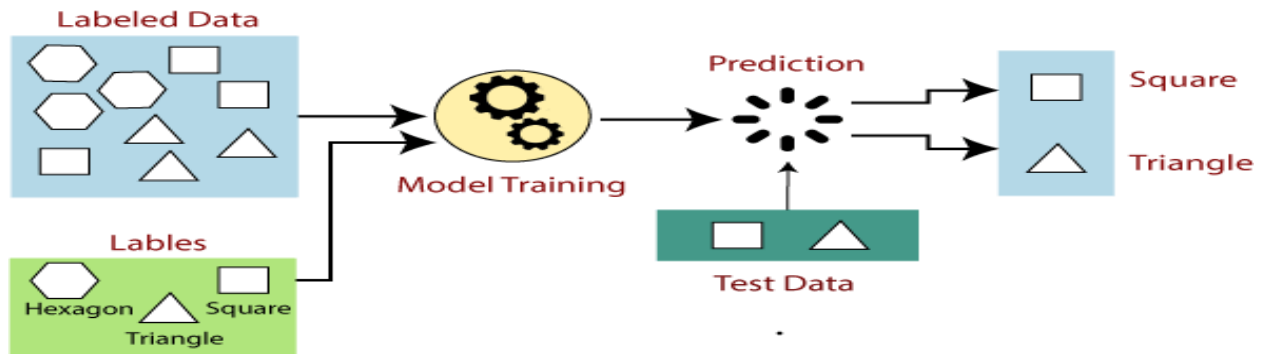
Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data.

**Arthur Samuel** an American pioneer in the field of computer gaming and artificial intelligence, coined the term "Machine Learning" in 1959 while at IBM. Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data

Machine learning tasks are typically classified into two broad categories, depending on whether there is a learning "signal" or "feedback" available to a learning system

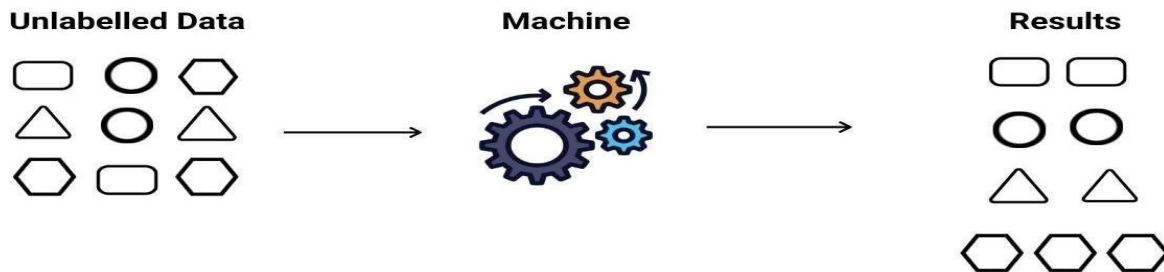
## SUPERVISED LEARNING

**Supervised learning** is the machine learning task of inferring a function from labeled training data. The training data consist of a set of training examples. In supervised learning, each example is a *pair* consisting of an input object (typically a vector) and a desired output value. A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way.



## UNSUPERVISED LEARNING

**Unsupervised learning** is the machine learning task of inferring a function to describe hidden structure from "unlabeled" data (a classification or categorization is not included in the observations). Since the examples given to the learner are unlabeled, there is no evaluation of the accuracy of the structure that is output by the relevant algorithm—which is one way of distinguishing unsupervised learning from supervised learning and reinforcement learning.



## NUMPY

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. The ancestor of NumPy, Numeric, was originally created by Jim Hugunin.

NumPy targets the CPython reference implementation of Python, which is a non-optimizing byte code interpreter. Mathematical algorithms written for this version of Python often run much slower than compiled equivalents.

Using NumPy in Python gives functionality comparable to MATLAB since they are both interpreted, and they both allow the user to write fast programs as long as most operations work on arrays or matrices instead of scalars.

## NUMPY ARRAY

NumPy's main object is the homogeneous multidimensional array. It is a table of elements (usually numbers), all of the same type, indexed by a tuple of positive integers. In NumPy dimensions are called axes. The number of axes is rank.

For example, the coordinates of a point in 3D space [1, 2, 1] is an array of rank 1, because it has one axis. That axis has a length of 3. In the example pictured below, the array has rank 2 (it is 2-dimensional). The first dimension (axis) has a length of 2, the second dimension has a length of 3.

## SLICING NUMPY ARRAY

```
Import numpy as np
a = np.array([[1,2,3],[3,4,5],[4,5,6]])

print 'Our array is:'
print a
print '\n'

print 'The items in the second column are:'
print a[:,1]
print '\n'
```

```
print 'The items in the second row are:'  
print a[1,...]  
print '\n'
```

```
print 'The items column 1 onwards are:'  
print a[...,:1]
```

## **OUTPUT:-**

Our array is:

```
[[1 2 3]  
 [3 4 5]  
 [4 5 6]]
```

The items in the second column are:

```
[2 4 5]
```

The items in the second row are:

```
[3 4 5]
```

The items columns 1 onwards are:

```
[[2 3]  
 [4 5]  
 [5 6]]
```

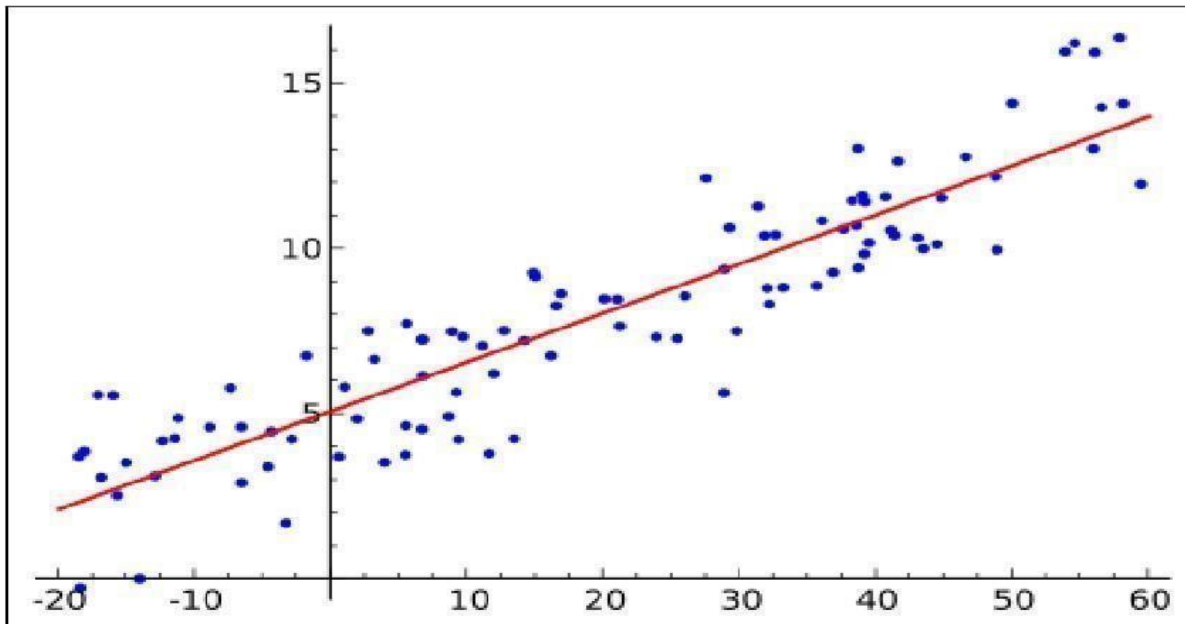
## **SCIPY**

SciPy builds on the NumPy array object and is part of the NumPy stack which includes tools like Matplotlib, pandas and SymPy, and an expanding set of scientific computing libraries. This NumPy stack has similar users to other applications such as MATLAB, GNU Octave, and Scilab. The NumPy stack is also sometimes referred to as the SciPy stack.

## **DATA STRUCTURE**

The basic data structure used by SciPy is a multidimensional array provided by the NumPy module. NumPy provides some functions for linear algebra, Fourier transforms and random number generation, but not with the generality of the equivalent functions in SciPy. NumPy can also be used as an efficient multi-dimensional container of data with arbitrary data-types. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases. Older versions of SciPy used Numeric as an array type, which is now deprecated in favour of the newer NumPy array code.

## REGRESSION ANALYSIS



In statistical modeling, **regression analysis** is a set of statistical processes for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors'). More specifically, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed.

Regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships. In restricted circumstances, regression analysis can be used to infer casual relationships between the independent and dependent variables. However this can lead to illusions or false relationships, so caution is advisable



## **LINEAR REGRESSION**

Linear regression is a linear approach for modeling the relationship between a scalar dependent variable  $y$  and one or more explanatory variables (or independent variables) denoted  $X$ . The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression.

In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models.

## **LOGISTIC REGRESSION**

Logistic regression, or logit regression, or logit model <sup>[1]</sup> is a regression model where the dependent variable (DV) is categorical. This article covers the case of a binary dependent variable—that is, where the output can take only two values, "0" and "1", which represent outcomes such as pass/fail, win/lose, alive/dead or healthy/sick. Cases where the dependent variable has more than two outcome categories may be analysed in multinomial logistic regression, or, if the multiple categories are ordered, in ordinal logistic regression. In the terminology of economics, logistic regression is an example of a qualitative response/discrete choice model.

## **POLYNOMIAL REGRESSION**

Polynomial regression is a form of regression analysis in which the relationship between the independent variable  $x$  and the dependent variable  $y$  is modeled as an  $n^{\text{th}}$  degree polynomial in  $x$ .

Polynomial regression fits a nonlinear relationship between the value of  $x$  and the corresponding conditional mean of  $y$ , denoted  $E(y | x)$ , and has been used to describe nonlinear phenomena such as the growth rate of tissues, the distribution of carbon isotopes in lake sediments, and the progression of disease epidemics.

Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear, in the sense that the regression function  $E(y | x)$  is linear in the unknown parameters that are estimated from the data.

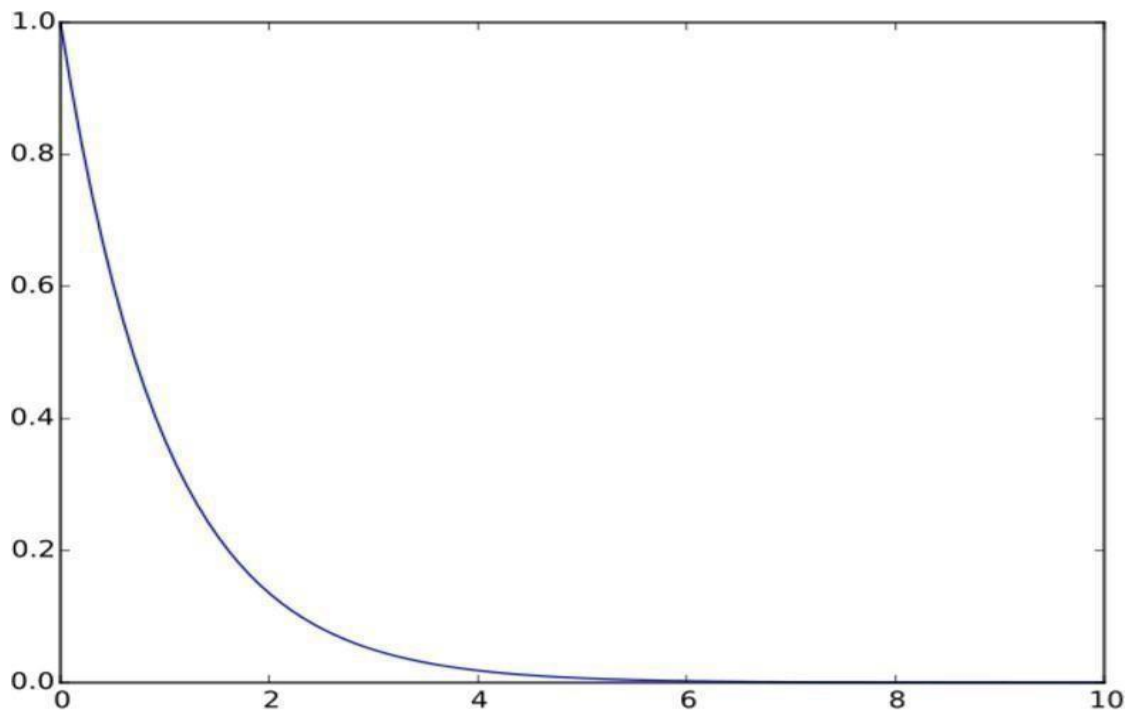
## MATPLOTLIB

**Matplotlib** is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged. SciPy makes use of matplotlib.

### EXAMPLE

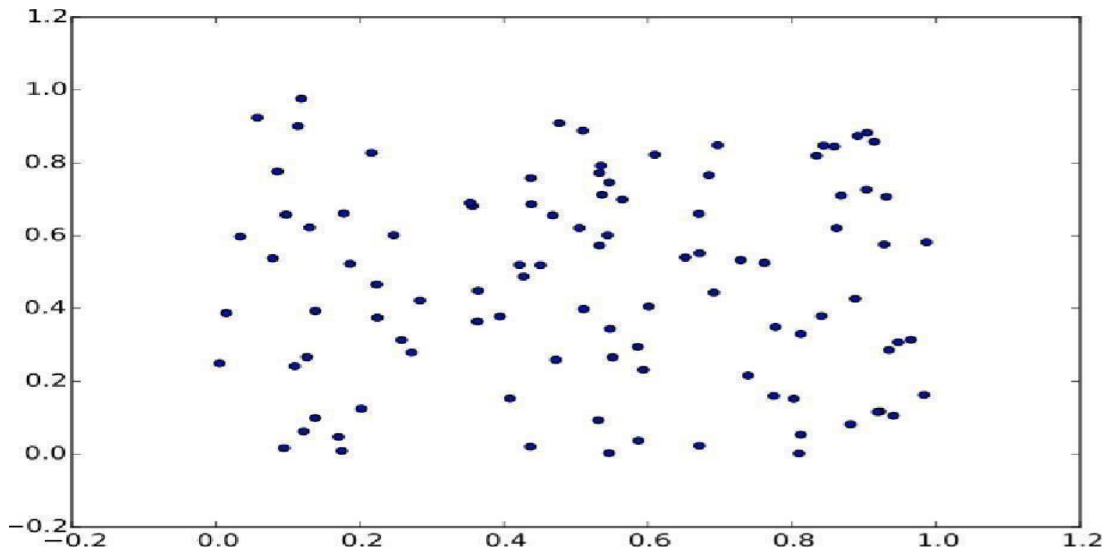
➤ LINE PLOT

```
>>>import matplotlib.pyplot as plt
>>>import numpy as np
>>> a = np.linspace(0,10,100)
>>> b = np.exp(-a)
>>>plt.plot(a,b)
>>>plt.show()
```



## ➤ SCATTER PLOT

```
>>>import matplotlib.pyplot as plt
>>>from numpy.random import rand
>>> a = rand(100)
>>> b = rand(100)
>>>plt.scatter(a,b)
>>>plt.show()
```



## PANDAS

In computer programming, pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. "Panel data", an econometrics term for multidimensional, structured data sets.

## CLUSTERING

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics.

## ALGORITHM

- Data Collection
- Data Formatting
- Model Selection
- Training
- Testing

**Data Collection:** We have collected data sets of weather from online website. We have downloaded the .csv files in which information was present.

**Data Formatting:** The collected data is formatted into suitable data sets. We check the collinearity with mean temperature. The data sets which have collinearity nearer to 1.0 has been selected.

**Model Selection:** We have selected different models to minimize the error of the predicted value. The different models used are Linear Regression Linear Model, Ridge Linear model, Lasso Linear Model and Bayesian Ridge Linear Model.

**Training:** The data sets were divided such that  $x_{train}$  is used to train the model with corresponding  $x_{test}$  values and some  $y_{train}$  kept reserved for testing.

**Testing:** The model was tested with  $y_{train}$  and stored in  $y_{predict}$ . Both  $y_{train}$  and  $y_{predict}$  was compared.

# ACTUAL CODING

## (STUDENT PLACEMENT PREDICTION)

### Importing Package and Reading Data

```
In [66]: import pandas as pd
path = r"C:\Users\HP\OneDrive\Desktop\collegePlace.csv"
df=pd.read_csv(path)
df.head()
```

```
Out[66]:
```

	Age	Gender	Stream	Internships	CGPA	Hostel	HistoryOfBacklogs	PlacedOrNot
0	22	Male	Electronics And Communication	1	8	1	1	1
1	21	Female	Computer Science	0	7	1	1	1
2	22	Female	Information Technology	1	6	0	0	1
3	21	Male	Information Technology	0	8	0	1	1
4	22	Male	Mechanical	0	8	1	0	1

```
In [64]: df.shape
```

```
Out[64]: (2966, 8)
```

```
In [70]: df['Gender'].replace({'Male':0 , 'Female':1},inplace = True)
```

```
In [70]: df['Gender'].replace({'Male':0 , 'Female':1},inplace = True)
```

```
In [71]: df.head()
```

```
Out[71]:
```

	Age	Gender	Stream	Internships	CGPA	Hostel	HistoryOfBacklogs	PlacedOrNot
0	22	0	Electronics And Communication	1	8	1	1	1
1	21	1	Computer Science	0	7	1	1	1
2	22	1	Information Technology	1	6	0	0	1
3	21	0	Information Technology	0	8	0	1	1
4	22	0	Mechanical	0	8	1	0	1

```
In [69]: df['Stream'].unique()
```

```
Out[69]: array(['Electronics And Communication', 'Computer Science',
              'Information Technology', 'Mechanical', 'Electrical', 'Civil'],
              dtype=object)
```

```
In [30]: df['Stream'].replace({'Electronics And Communication':0, 'Computer Science':1, 'Information Technology':2, 'Mechanical':3, 'Electrical':4, 'Civil':5},inplace = True)
```

```
In [31]: df.head()
```

## Cleaning of Data

```
dtype=object)

In [30]: df['Stream'].replace({'Electronics And Communication':0, 'Computer Science':1, 'Information Technology':2, 'Mechanical':3, 'Electronics':4})

In [31]: df.head()

Out[31]:
```

	Age	Gender	Stream	Internships	CGPA	Hostel	HistoryOfBacklogs	PlacedOrNot
0	22	0	0	1	8	1	1	1
1	21	1	1	0	7	1	1	1
2	22	1	2	1	6	0	0	1
3	21	0	2	0	8	0	1	1
4	22	0	3	0	8	1	0	1

```
In [32]: df.isnull().sum()

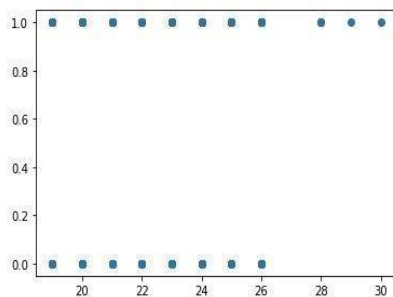
Out[32]: Age          0
Gender          0
Stream          0
Internships     0
CGPA            0
Hostel          0
HistoryOfBacklogs 0
PlacedOrNot    0
dtype: int64
```

## Importing Matplotlib

```
dtype: int64

In [33]: import matplotlib.pyplot as plt
age=df['Age']
stream=df['Stream']
internships=df['Internships']
cgpa=df['CGPA']
hostel=df['Hostel']
backlogs=df['HistoryOfBacklogs']
y=df['PlacedOrNot']
plt.scatter(age,y)

Out[33]: <matplotlib.collections.PathCollection at 0x24c431143d0>
```



```
In [34]: y=df['PlacedOrNot']
y=df[df['PlacedOrNot']==1]
```

## Removed column “Placed or not”

```
In [34]: y=df['PlacedOrNot']  
X=df.drop(['PlacedOrNot'],axis = 1)
```

```
In [35]: X
```

```
Out[35]:
```

	Age	Gender	Stream	Internships	CGPA	Hostel	HistoryOfBacklogs
0	22	0	0	1	8	1	1
1	21	1	1	0	7	1	1
2	22	1	2	1	6	0	0
3	21	0	2	0	8	0	1
4	22	0	3	0	8	1	0
...	...	...	...	...	...	...	...
2961	23	0	2	0	7	0	0
2962	23	0	3	1	7	1	0
2963	22	0	2	1	7	0	0
2964	22	0	1	1	7	0	0
2965	23	0	5	0	8	0	0

2966 rows x 7 columns

## Applying Logistic Regression

```
In [36]: from sklearn.model_selection import train_test_split  
X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.8,random_state=1)
```

```
In [37]: from sklearn.linear_model import LogisticRegression  
from sklearn import metrics
```

```
In [38]: logreg = LogisticRegression()
```

```
In [41]: print(X_train)
```

	Age	Gender	Stream	Internships	CGPA	Hostel	HistoryOfBacklogs
557	21	1	5	1	7	0	1
1039	22	1	2	1	8	0	0
561	21	0	1	1	7	0	0
2819	22	0	3	0	6	1	0
2039	20	0	1	2	7	0	0
...	...	...	...	...	...	...	...
2763	21	1	1	1	8	1	0
905	22	0	4	0	8	0	1
1096	22	0	0	1	7	0	1
235	22	0	1	1	8	0	0
1061	22	0	1	1	7	0	0

[2372 rows x 7 columns]

```
In [49]: logreg.fit(X_train,y_train)
```

## Predicting Output and its Accuracy through Confusion Matrix

```
In [49]: logreg.fit(X_train,y_train)
```

```
Out[49]: LogisticRegression()
```

```
In [50]: y_pred=logreg.predict(X_test)
```

```
In [51]: print(y_pred)
```

```
[0 1 0 0 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 1 0 1 1 1 1 1 1 0 0 1 1 0 0
 0 0 1 0 1 1 0 0 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1
 1 1 0 1 0 1 1 1 1 0 1 0 0 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 0 0 1 0 0 1 0 1
 0 0 0 0 0 1 0 0 1 0 1 0 1 0 0 1 0 1 0 0 1 1 0 1 1 0 0 1 0 1 0 0 1 0 1 0 0 1 0 1 1 1
 1 0 0 1 1 1 1 0 1 1 0 0 0 1 1 0 1 0 1 1 1 0 0 0 0 1 1 1 0 0 0 1 0 1 1 1 0 1 1 1 0 1
 1 0 0 1 0 1 0 1 0 0 0 0 1 1 0 1 0 1 1 1 1 1 0 0 0 0 1 1 1 1 1 0 1 1 1 0 1 1 1 0 1
 0 0 1 1 0 0 1 1 1 0 0 0 1 1 1 0 1 0 1 1 1 1 1 0 1 0 1 1 0 1 0 1 1 0 0 0 1
 1 0 0 0 0 0 1 0 1 0 0 1 0 0 1 1 1 1 1 0 0 1 1 0 1 0 1 1 1 1 0 0 0 0 1 1 0
 0 0 1 0 1 1 1 0 1 1 0 1 1 0 1 1 1 0 1 0 0 1 1 0 1 1 0 1 0 1 0 1 1 0 0 1 0 0
 1 1 0 1 1 0 1 1 0 0 0 1 0 1 1 1 0 1 0 1 0 0 0 0 1 0 1 1 1 1 1 1 0 1 1 0 1 1 1
 0 0 1 1 0 0 1 1 1 1 1 1 1 0 1 1 1 0 1 1 1 1 0 1 1 1 1 0 0 1 1 1 1 1 1 0 1 1
 1 1 1 1 1 1 1 0 1 0 0 0 1 0 0 1 0 0 1 0 0 0 1 0 1 0 0 0 1 1 1 0 0 1 0 1 0
 1 0 0 1 0 0 1 1 0 1 1 1 0 0 0 0 1 0 0 1 1 1 1 1 1 0 1 1 1 0 0 1 1 1 0 1 1 1
 1 1 0 1 0 0 1 1 0 0 0 0 0 1 0 1 1 1 0 1 0 1 0 0 1 1 1 1 1 0 1 0 1 0 1 0
 0 1 0 0 1 1 0 1 1 1 1 1 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 1 1 1 0 0 1 1 1 0 0 1 0 1
 1 0 0 1 0 0 0 1 1 1 1 1 1 0 1 0 0 0 1 1 0 1 0 0 0 1 1 1 0 1 1 1 1 1 0 0 0
 1 1]
```

```
In [48]: print(logreg.score(X_test,y_test))
```

```
1 0 0 1 0 0 1 1 0 1 1 1 0 0 0 0 1 0 0 1 1 1 1 1 0 1 1 0 0 1 1 1 0 1 1 1
1 1 0 1 0 0 1 1 0 0 0 0 0 1 0 1 1 1 0 1 0 1 0 1 0 0 1 1 1 1 0 1 0 1 0 1 0
0 1 0 0 1 1 0 1 1 1 1 0 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 1 1 0 0 1 0 1
1 0 0 1 0 0 0 1 1 1 1 1 0 1 1 0 0 1 1 0 1 0 0 1 1 1 0 1 1 1 1 1 1 0 0 0
1 1]
```

```
In [48]: print(logreg.score(X_test,y_test))
```

```
0.7693602693602694
```

```
In [61]: logreg.predict([[0,2,8,90,85,0,1]])
```

```
C:\Users\HP\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but LogisticRegression was fitted with feature names
  warnings.warn(
```

```
Out[61]: array([1], dtype=int64)
```

```
In [62]: from sklearn.metrics import confusion_matrix, accuracy_score
```

```
In [63]: print(confusion_matrix(y_test,y_pred))
print(accuracy_score(y_test))
```

```
[[189 74]
 [ 63 268]]
0.7693602693602694
```

```
In [ ]:
```



## CONCLUSION

We have collected the raw data from online sources. Then we take this raw data and format it.

Now we have selected few models for error detection. We have used two models namely logistic regression model, classification model.