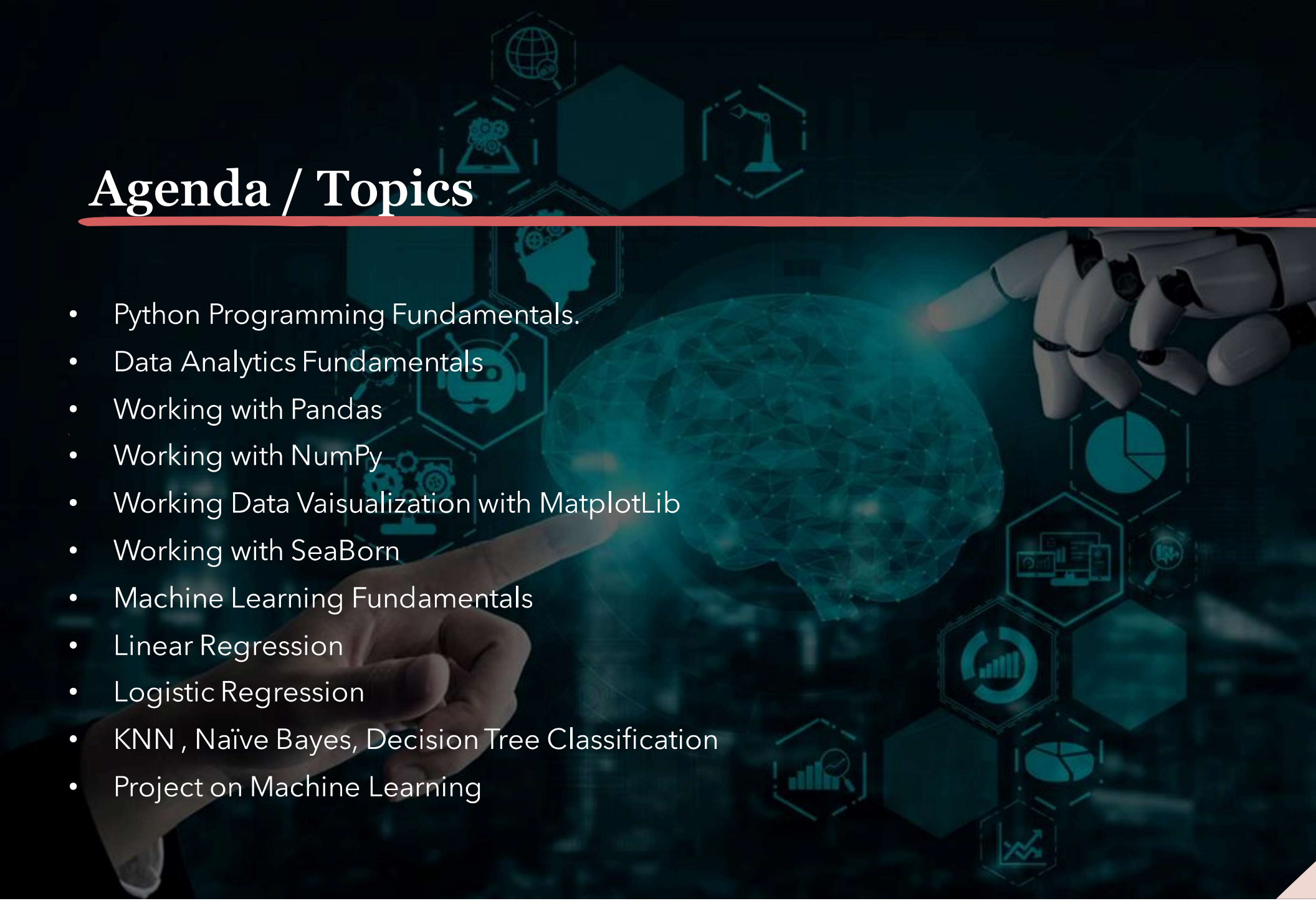
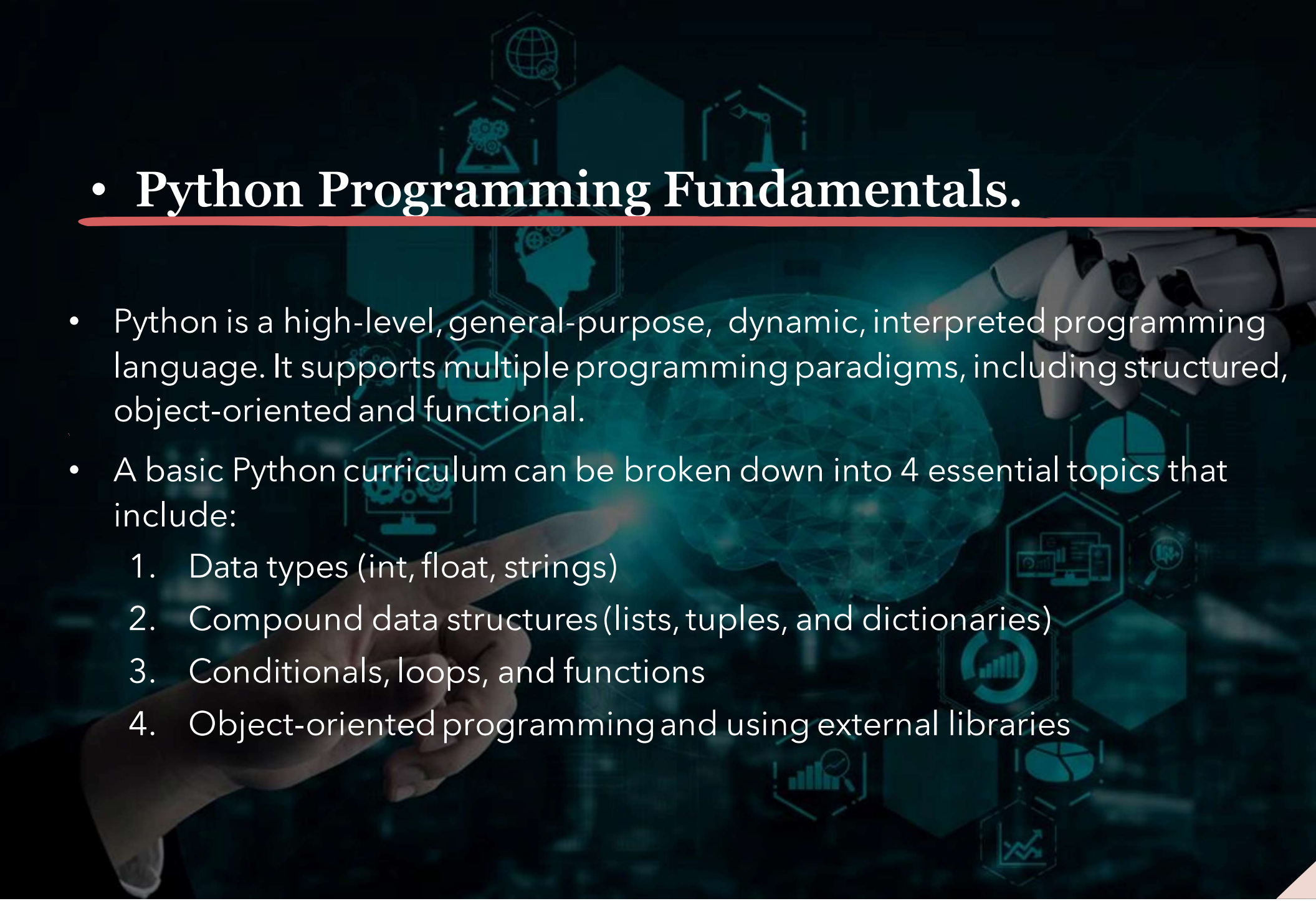

Python-DA-ML-AI
MENTORED BY : ----------
TRAINING ORGANIZATION : Froyo Technology
PRESENTED BY : Bharat Yadav
SUBMITTED TO : Ajay Kumar Garg Engineering College
TRAINING DURATION - 6 WEEKS (1 JULY TO 31 JULY)



Agenda / Topics

- Python Programming Fundamentals.
- Data Analytics Fundamentals
- Working with Pandas
- Working with NumPy
- Working Data Vaisualization with MatplotLib
- Working with SeaBorn
- Machine Learning Fundamentals
- Linear Regression
- Logistic Regression
- KNN , Naïve Bayes, Decision Tree Classification
- Project on Machine Learning



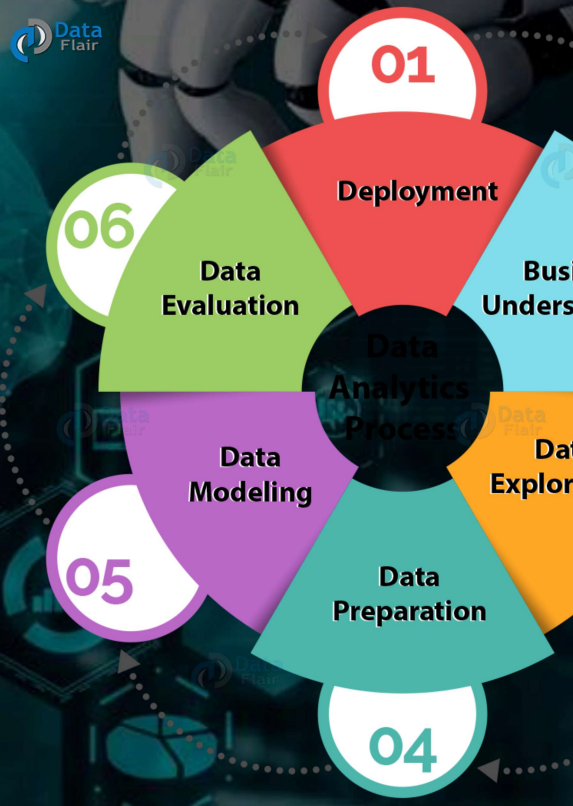
• Python Programming Fundamentals.

- Python is a high-level, general-purpose, dynamic, interpreted programming language. It supports multiple programming paradigms, including structured, object-oriented and functional.
- A basic Python curriculum can be broken down into 4 essential topics that include:
  1. Data types (int, float, strings)
  2. Compound data structures (lists, tuples, and dictionaries)
  3. Conditionals, loops, and functions
  4. Object-oriented programming and using external libraries

• Data Analytics Fundamentals

- Data analytics is the science of integrating heterogeneous data from diverse sources, drawing inferences, and making predictions to enable innovation, gain competitive business advantage, and help strategic decision-making.

- **Data Analysis in 5 Steps**
  - ❑ STEP 1: DEFINE QUESTIONS & GOALS.
  - ❑ STEP 2: COLLECT DATA.
  - ❑ STEP 3: DATA WRANGLING.
  - ❑ STEP 4: DETERMINE ANALYSIS.
  - ❑ STEP 5: INTERPRET RESULTS.



Use of Data Analy

Working with Pandas

- Pandas is **an open source Python package that is most widely used for data science/data analysis and machine learning tasks**. It is built on top of another package named Numpy , which provides support for multi-dimensional arrays.

- **Use of Pandas:-**
  - ❑ Data fill
  - ❑ Data normalization
  - ❑ Merges and joins
  - ❑ Data visualization
  - ❑ Statistical analysis
  - ❑ And much more

Working with NumPy

- NumPy is a general-purpose array-processing package. It provides a high-perfo multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python.

- Use of Numpy :-
  - A powerful N-dimensional array object
  - Sophisticated (broadcasting) functions
  - Tools for integrating C/C++ and Fortran code
  - Useful linear algebra, Fourier transform, and random number capabilities

Working Data Vaisualization with MatplotLib

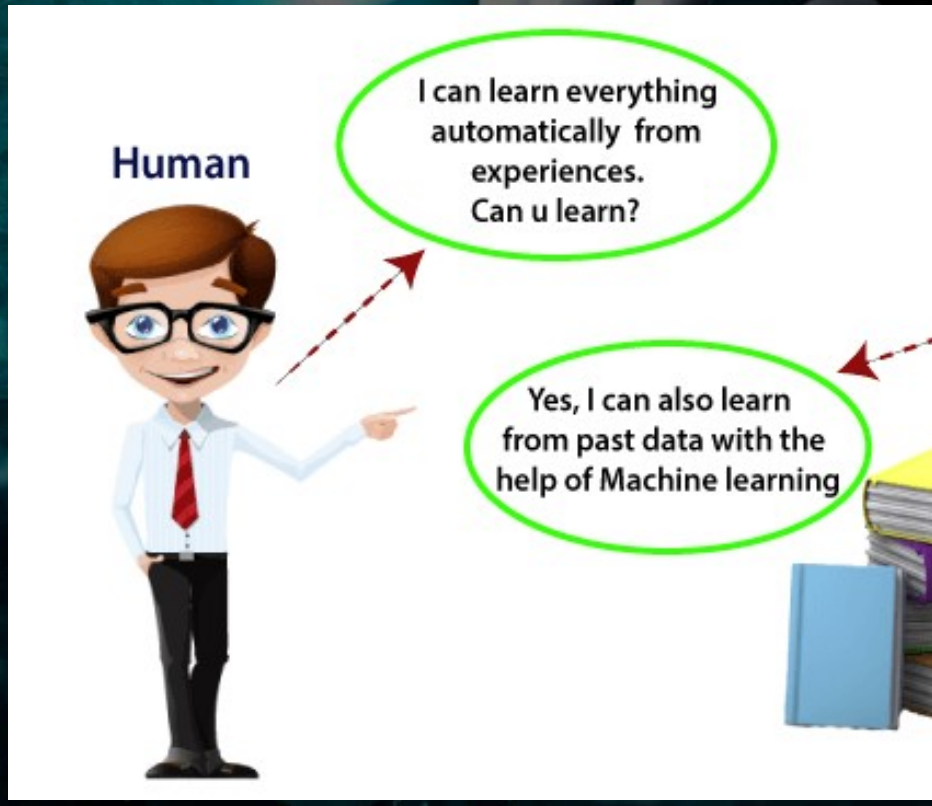
- Data and information visualization is an interdisciplinary field that deals with the representation of data and information. It is a particularly efficient way of communicating when the data or information is numerous as for example a time
- Matplotlib is a plotting library for the Python programming language and its nu mathematics extension NumPy. It provides an object-oriented API for embeddi into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, o

Working with SeaBorn

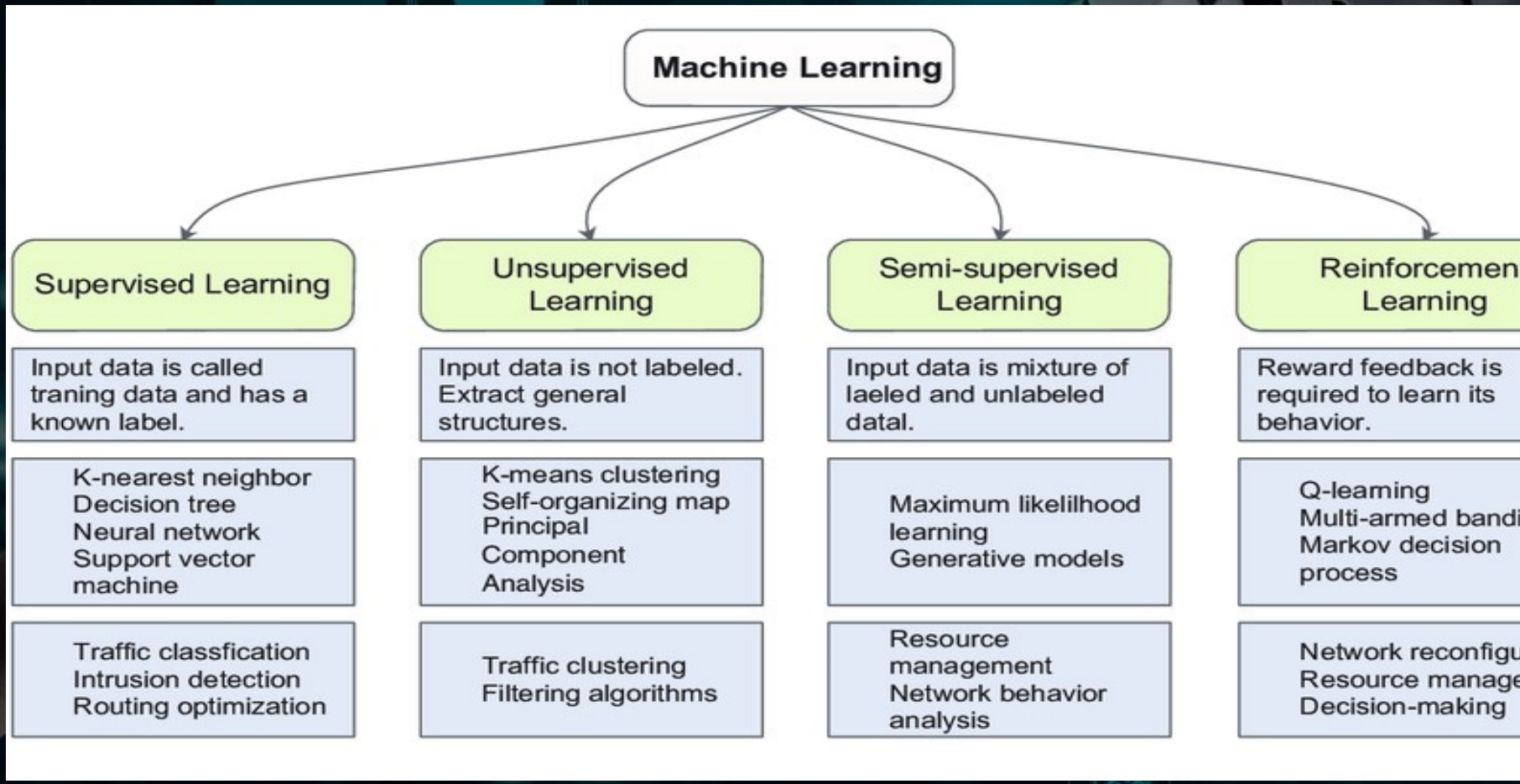
- Seaborn is **a library that uses Matplotlib underneath to plot graphs**. It will be to visualize random distributions.
- To work with seaborn one have to import seaborn as well as matplotlib as seabo based on matplotlib.
- Seaborn helps to visualize the statistical relationships, To understand how variab dataset are related to one another and how that relationship is dependent on ot variables.
- This library is used to visualize our data we do not need to take care of the interr details; we just have to pass our data set or data inside the relplot() function, and calculate and place the value accordingly.

• Machine Learning Fundamentals

- Machine learning is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks. It is seen as an application of artificial intelligence.
- Its primary aim is to allow the computers to learn automatically without human intervention and adjust actions accordingly.



- **Machine Learning Fundamentals**



Linear Regression

- Linear relationship is a simplest mathematical relationship between two variables x a
- In a cause and effect relationship, the independent variable is the cause , and the de variable is the effect.
- Linear regression is a method for predicting the value for dependent variable Y, base value of an independent variable X.
- Mathematically we can represent linear regression as, $y= a_0+a_1x+ \varepsilon.$
- When working with linear regression, our main goal is to find the best fit line that me error between predicted values and actual values should be minimized. The best fit l have the least error.
- Linear regression can be further divided into two types of the algorithm:
- **Simple Linear Regression**
- **Multiple Linear regression**

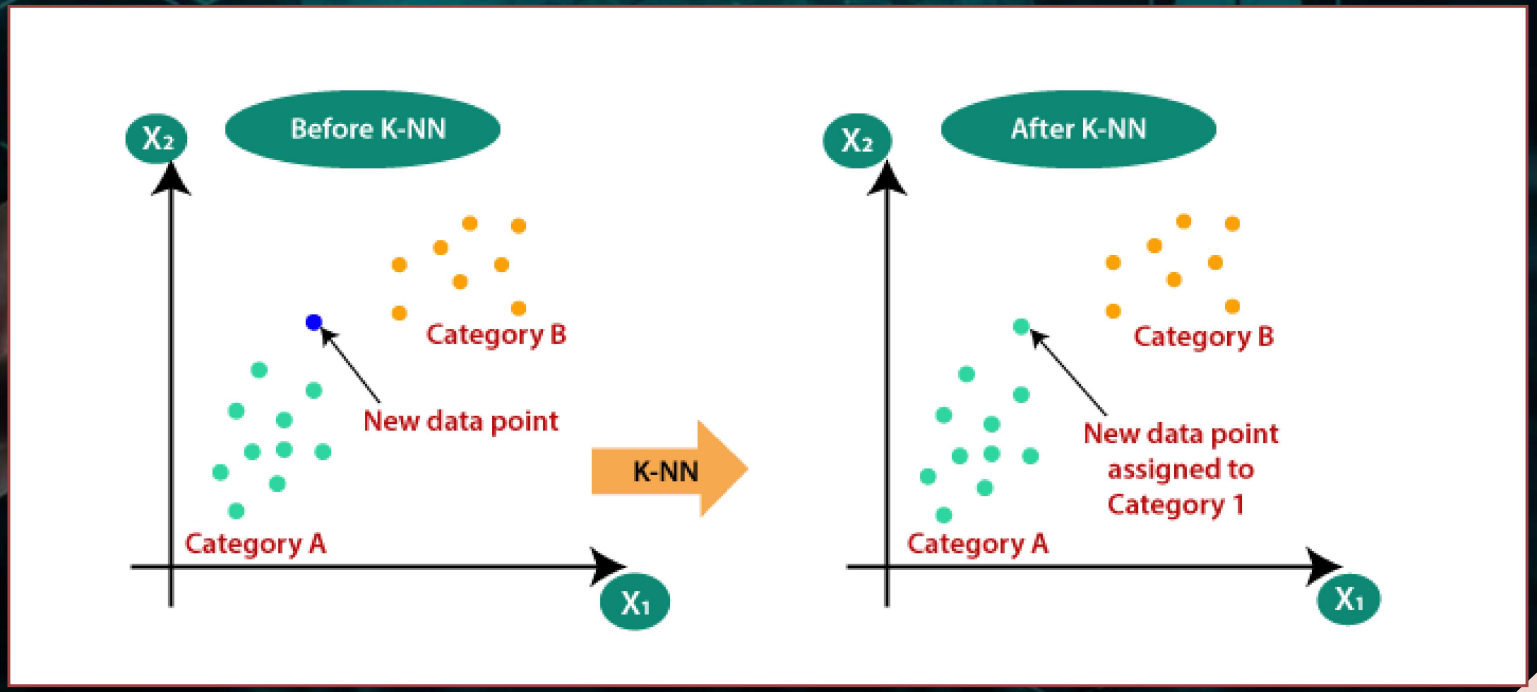
Logistic Regression

- Logistic regression is **a statistical analysis method to predict a binary outcom as yes or no, based on prior observations of a data set.**
- A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables.

- There are three main types of logistic regression:-
  - **binary,**
  - **multinomial**
  - **ordinal**.

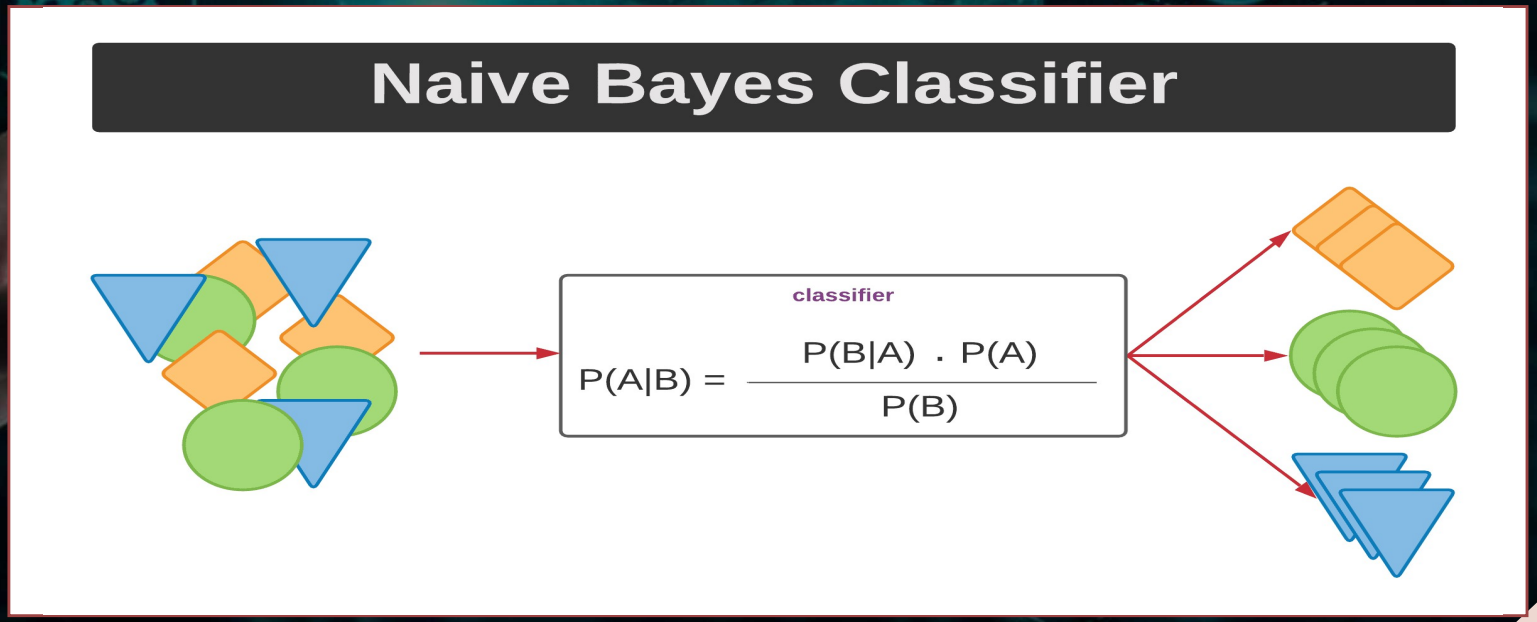
KNN

- The abbreviation KNN stands for "**K-Nearest Neighbour**". It is a supervised mac learning algorithm. The algorithm can be used to solve both classification and re problem statements.



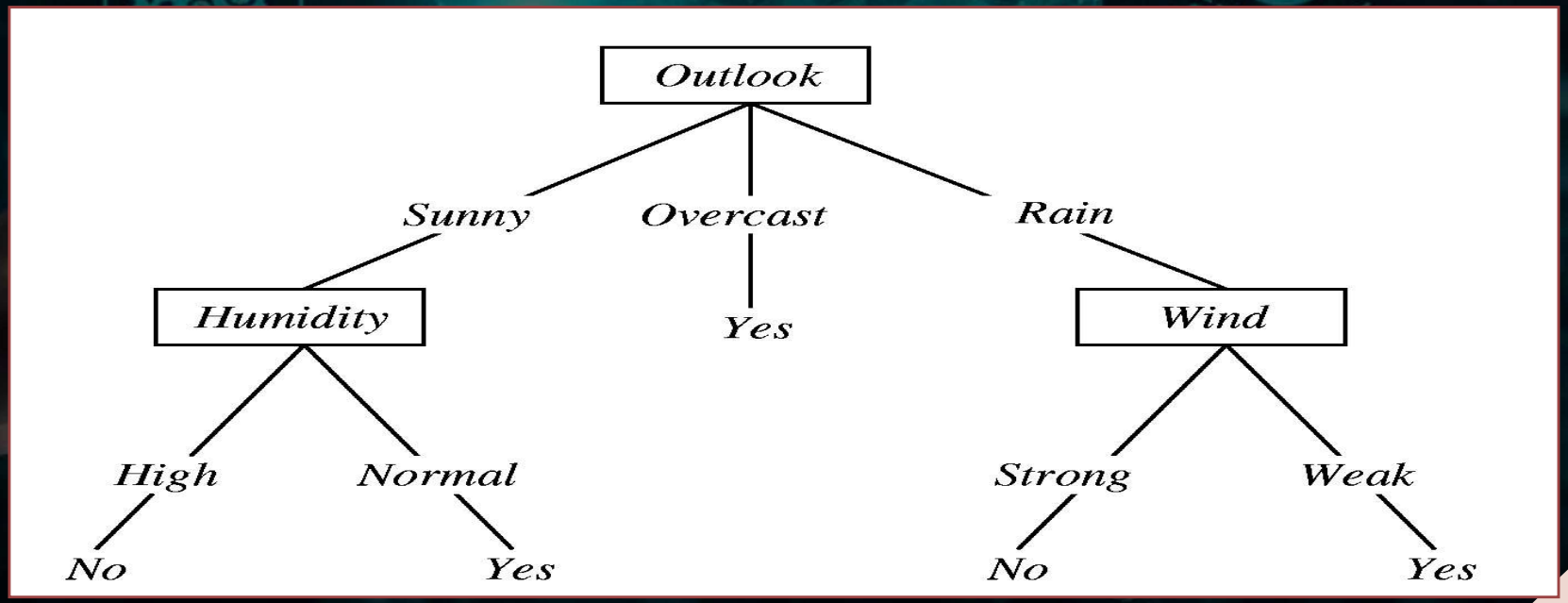
Naïve Bayes

- Naïve Bayes is **one of the fast and easy ML algorithms to predict a class of da** can be used for Binary as well as Multi-class Classifications. It performs well in M predictions as compared to the other Algorithms. It is the most popular choice f classification problems.



Decision Tree Classification

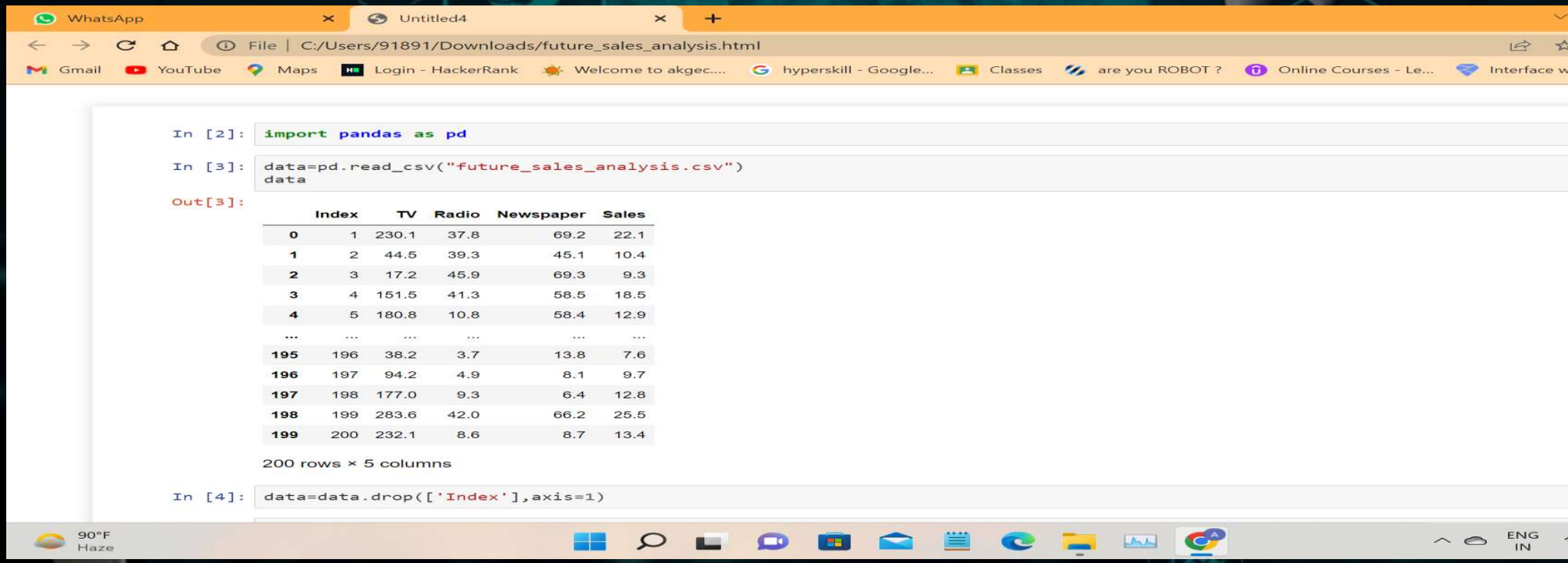
- A decision tree is a flowchart-like structure in which each internal node represen a `test` on a feature (e.g. whether a coin flip comes up heads or tails) , each leaf r represents a `class label` (decision taken after computing all features) and bran represent conjunctions of features that lead to those class labels. The paths from leaf represent `classification rules`.




Project On Machine Learning
Future Sales Analysis

Importing Dataset Using Pandas

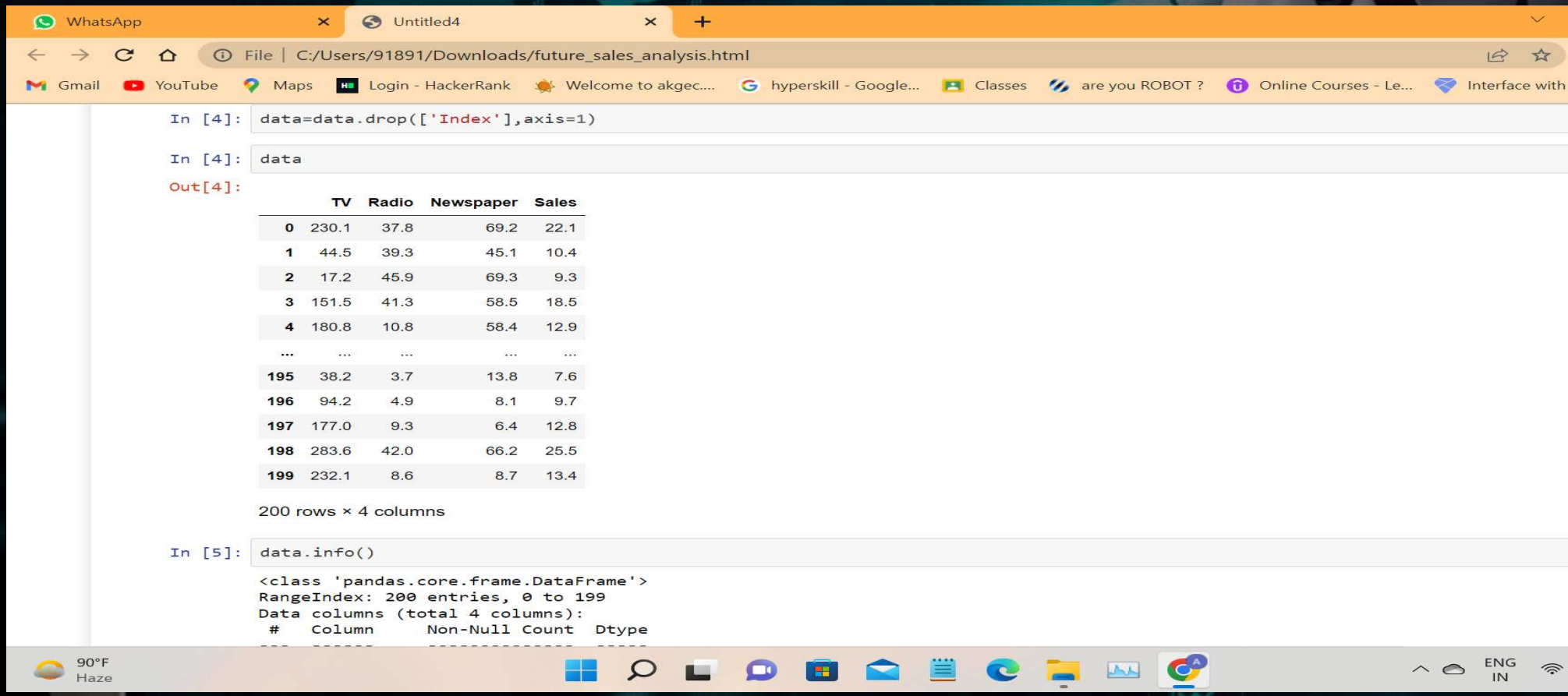
Using read method of pandas we load the dataset into jupyter notebook.



Data Preprocessing/Data Wrangling

- In this step the acquired data get cleaned by removing the unwanted data and s the errors in the data if any.
- This is one of the most important and time consuming step in Data analysis and tidy data.
- This uses most of the python libraries for visualizing the data like matplotlib , sea
- It uses drop method to drop any column with the axis method if axis value is 0 th along row the data will get drop and if value is one then vice-versa.

Data Preprocessing Drop Step



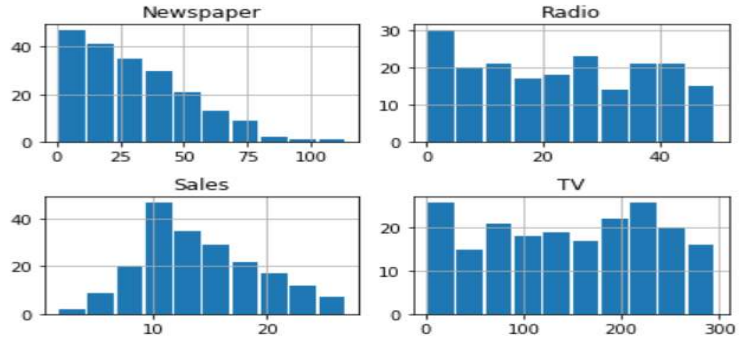
Data Preprocessing visualizing Step

WhatsApp | Untitled4

File | C:/Users/91891/Downloads/future_sales_analysis.html

Gmail YouTube Maps Login - HackerRank Welcome to akgec.... hyperskill - Google... Classes are you ROBOT ? Online Courses - Le... Interface with

In [9]:
```
data.hist(rwidth=0.9)
plt.tight_layout()
```



In [5]:
```
import matplotlib.pyplot as plt
```

In [20]:
```
plt.subplot(2,2,1)
plt.title('Newspaper vs sales')
plt.scatter(data['Newspaper'],data['Sales'],s=10,c='r')

plt.subplot(2,2,2)
plt.title('TV vs Sales')
plt.scatter(data['TV'],data['Sales'],s=10,c='g')

plt.subplot(2,2,3)
```

90°F Haze

ENG IN

Data Preprocessing Visualizing Step

WhatsApp | Untitled4

File | C:/Users/91891/Downloads/future_sales_analysis.html

Gmail | YouTube | Maps | Login - HackerRank | Welcome to akgec.... | hyperskill - Google... | Classes | are you ROBOT ? | Online Courses - Le... | Interface with

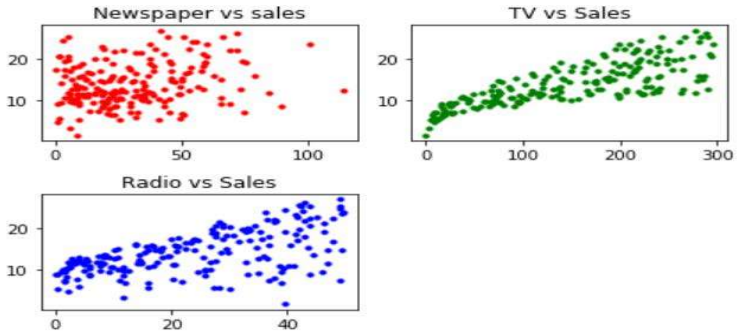
In [5]:
```
import matplotlib.pyplot as plt
```

In [20]:
```
plt.subplot(2,2,1)
plt.title('Newspaper vs sales')
plt.scatter(data['Newspaper'],data['Sales'],s=10,c='r')

plt.subplot(2,2,2)
plt.title('TV vs Sales')
plt.scatter(data['TV'],data['Sales'],s=10,c='g')

plt.subplot(2,2,3)
plt.title('Radio vs Sales')
plt.scatter(data['Radio'],data['Sales'],s=10,c='b')

plt.tight_layout()
```



90°F Haze

ENG IN

Data Preprocessing Visualizing Step

WhatsApp | Untitled4

File | C:/Users/91891/Downloads/future_sales_analysis.html

Gmail | YouTube | Maps | Login - HackerRank | Welcome to akgec.... | hyperskill - Google... | Classes | are you ROBOT ? | Online Courses - Le... | Interface wit

In [21]:
```
import seaborn as sns
sns.set()
```

In [22]:
```
correlation=data.corr()
correlation
```

Out[22]:

	TV	Radio	Newspaper	Sales
TV	1.000000	0.054809	0.056648	0.782224
Radio	0.054809	1.000000	0.354104	0.576223
Newspaper	0.056648	0.354104	1.000000	0.228299
Sales	0.782224	0.576223	0.228299	1.000000

In [24]:
```
sns.heatmap(correlation,annot=True)
```

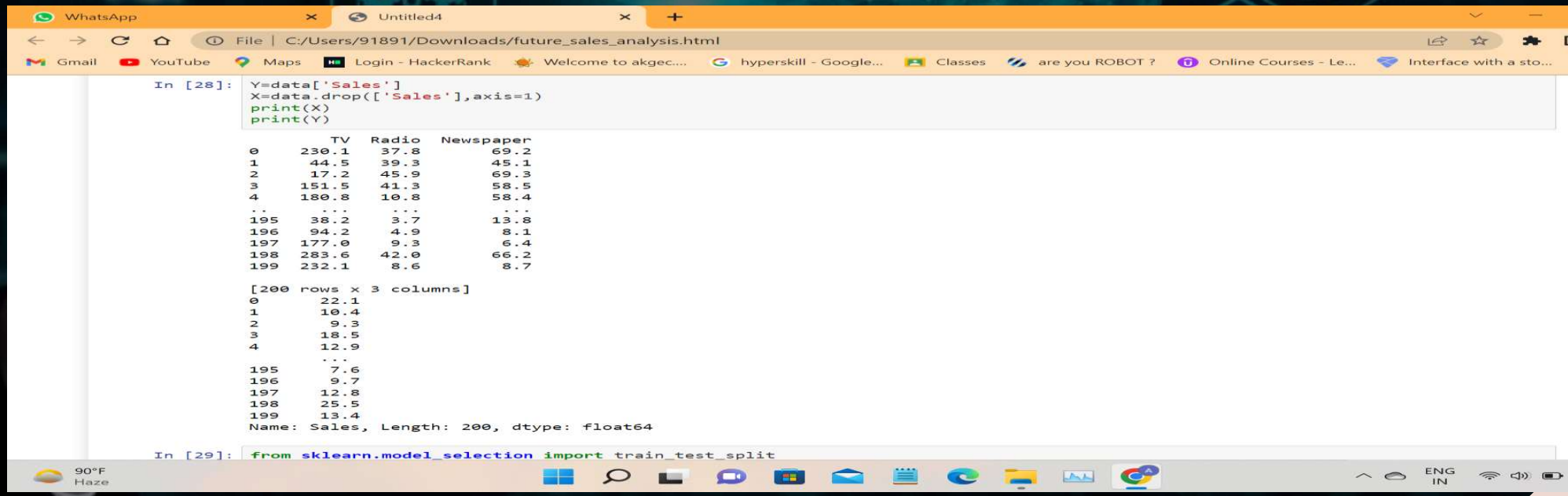
Out[24]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1cfc04b5788>
```

90°F Haze

ENG IN

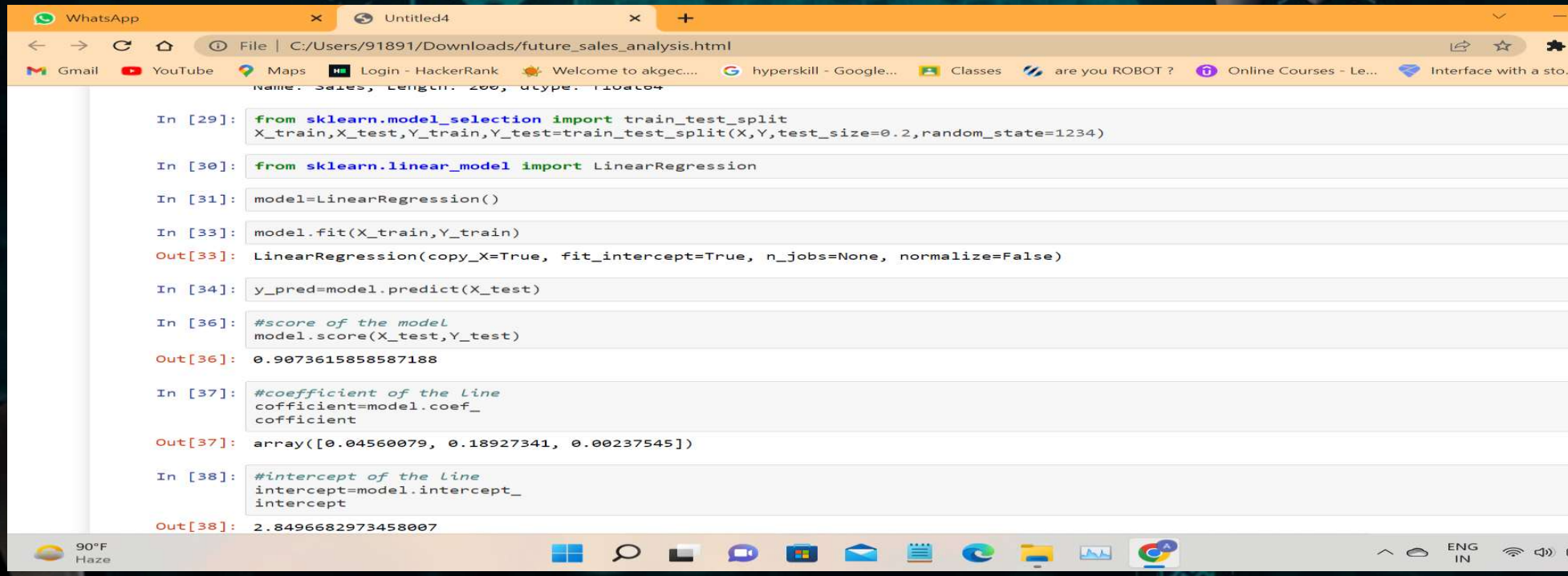
Applying Machine learning Algorithms

- After getting the tidy data we have to build the model by ML algorithms ba the need of the data analysis.
- In this project we have used linear regression ML algorithm to analyse the f sales. We have two test sets X and Y.

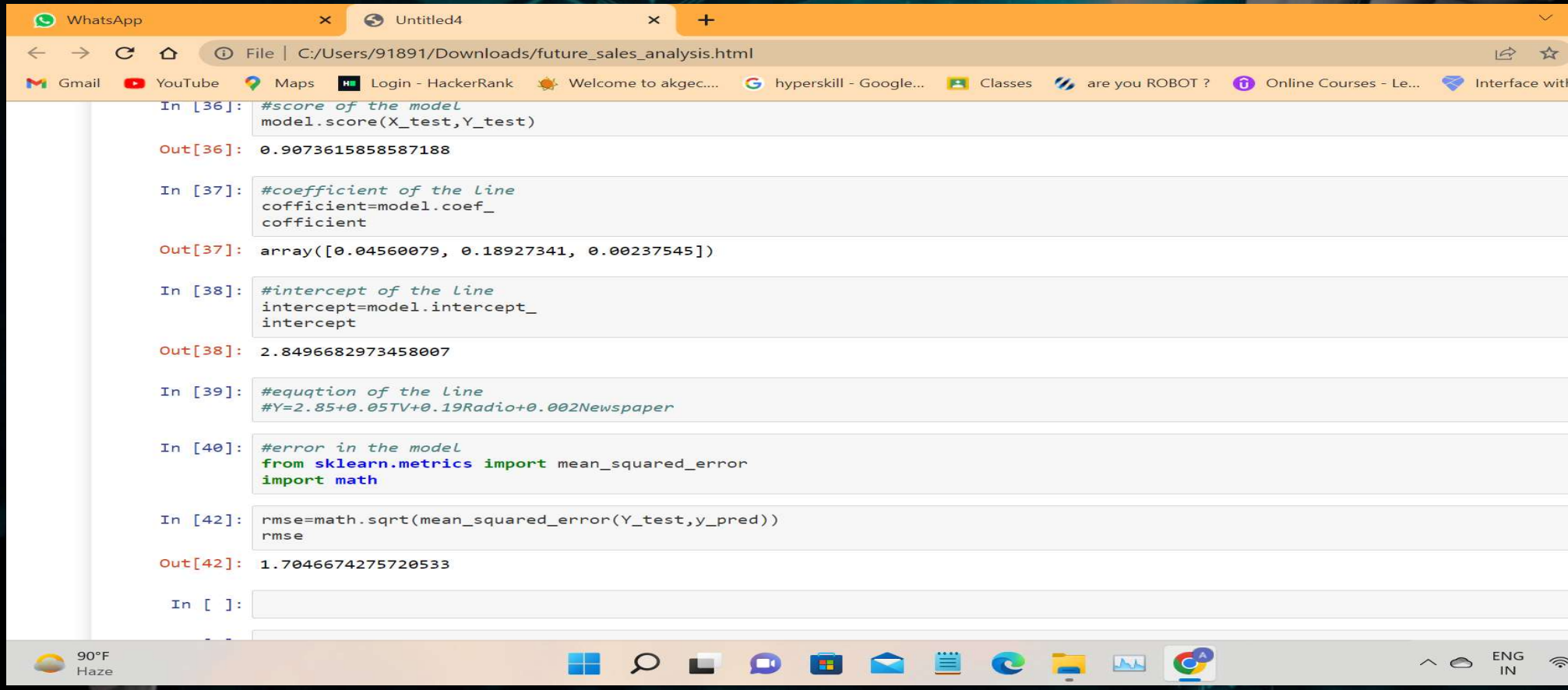


Applying Linear Regression

- **After having training and testing sets we apply linear regression algorithm two test sets by importing sklearn and its sub-module model_selection.**



Calculating Error Using mean Square



Conclusion

- We have seen the score of the model is 90% and Error is 1.7 which is good so our model is good and it predicts almost accurate value.