Woxsen University

Machine Learning

Assignment-2

Presented by

K.Sagar Datta

M.Srikar Reddy

M.Deepak

K.Sai Teja

C.Mohith Reddy

Next Slide

Supervised Machine Learning

Task-1

The process of training a model by giving a labeled data where the target value is known
.labeled data is divided into two parts
1) training dataset
2) testing dataset

Advantages of using supervised ML

1)High accuracy when trained with sufficient labeled data
2)The presence of labeled data makes it easier to define a clear relationship between inputs and outputs, enabling the model to learn effectively.

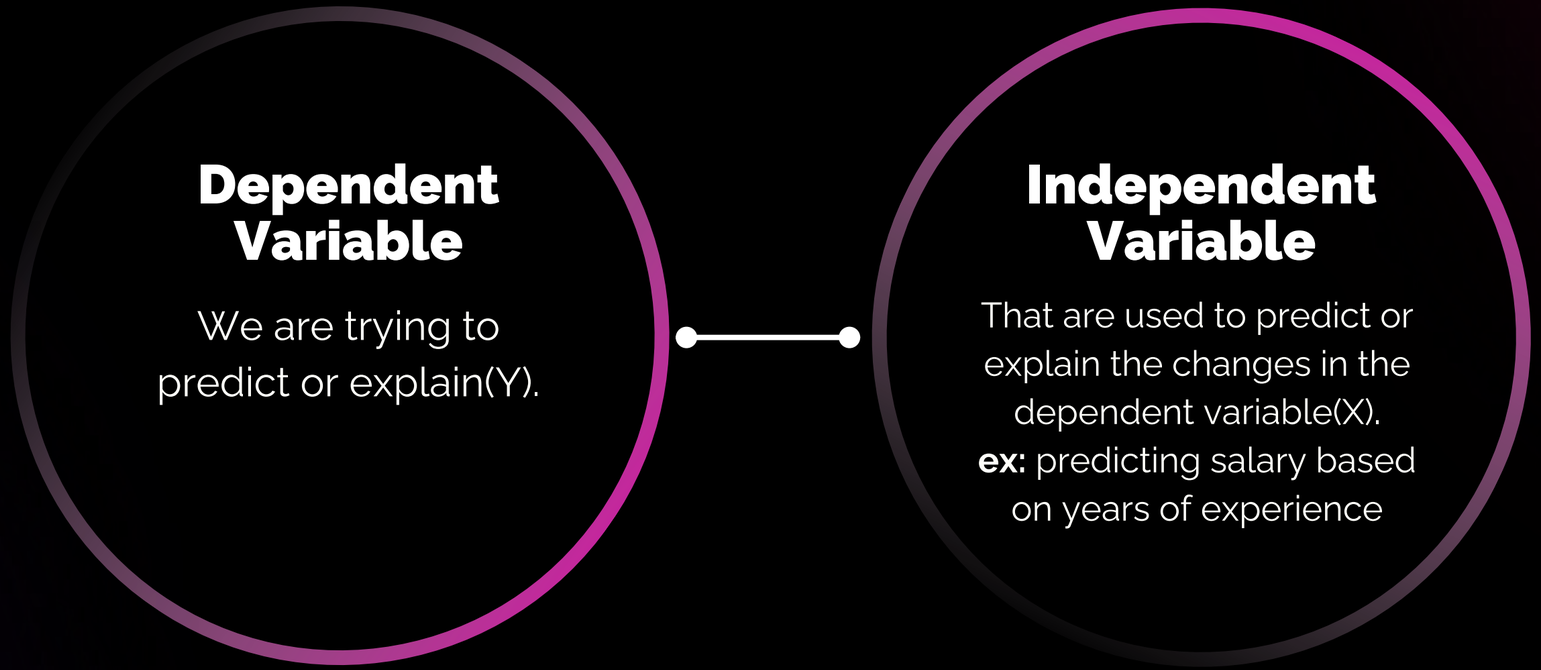
example: calculations of weight based on height

there are two types of supervised ml
1)Regression task and
2)Classification tasks

Next Slide

Classification task:
It is a supervised learning where the goal is to predict a category for a given input based on the training data.

Regression task:
It is the statistical method that help us to understand and predict the relation btw the varable.



Next Slide

Unsupervised Machine Learning

Unsupervised learning involves analyzing data without labeled outputs to find hidden patterns or structures.

Key Algorithms used are K-Means Clustering, Hierarchical Clustering, DBSCAN, PCA (Principal Component Analysis) and Autoencoders

Advantages

No need for labeled data, useful for exploratory data analysis.
Challenges: Difficult to evaluate, results are not always interpretable, and prone to overfitting

Next Slide

Supervised Learning

Classification	Regression
Logistic Regression	Linear Regression
Naive Bayes	Ridge Regression
Linear Discriminant Analysis (LDA)	Lasso Regression
Decision Trees	
Random Forest	
Support Vector Machines (SVM)	
k-Nearest Neighbors (KNN)	
Gradient Boosting algorithms	
Neural Networks	

Unsupervised Learning

PCA
K-mean Clustering
Hierarchical Clustering
DB Scan Clustering

REFERENCE

1..Nasteski, V. (2018). An overview of the supervised machine learning methods. Retrieved from ResearchGate

2. Author(s). (2023). Unsupervised machine learning for disease prediction: A comparative study. H and Technology, 13(1), 45–56

3. Lu, H., & Uddin, S. (2024). Unsupervised machine learning for disease prediction: a comparative performance analysis using multiple datasets. Health and Technology, 14(1), 141-154

Next Slide

Task-2

Python Code Workflow for Data Manipulation and Regression

Model Evaluation

Assessing the model's performance

Model Training

Applying algorithms to train the model

Data Visualization

Creating visual representations of data

Data Cleaning

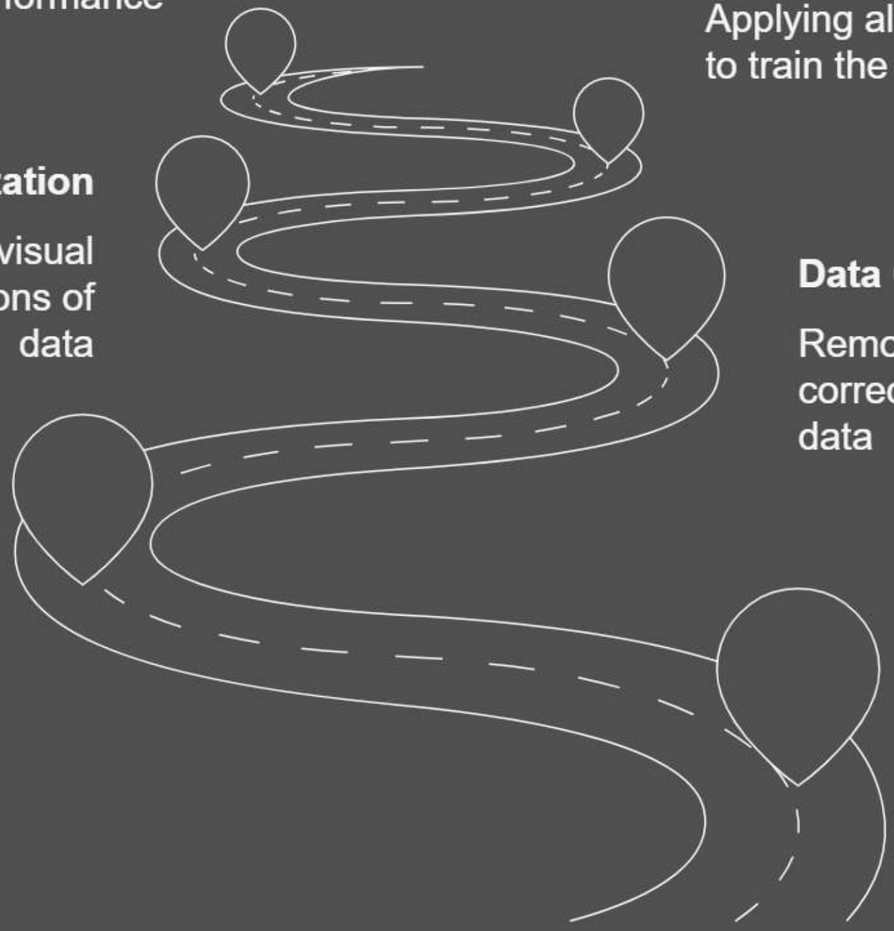
Removing or correcting erroneous data

Load Data

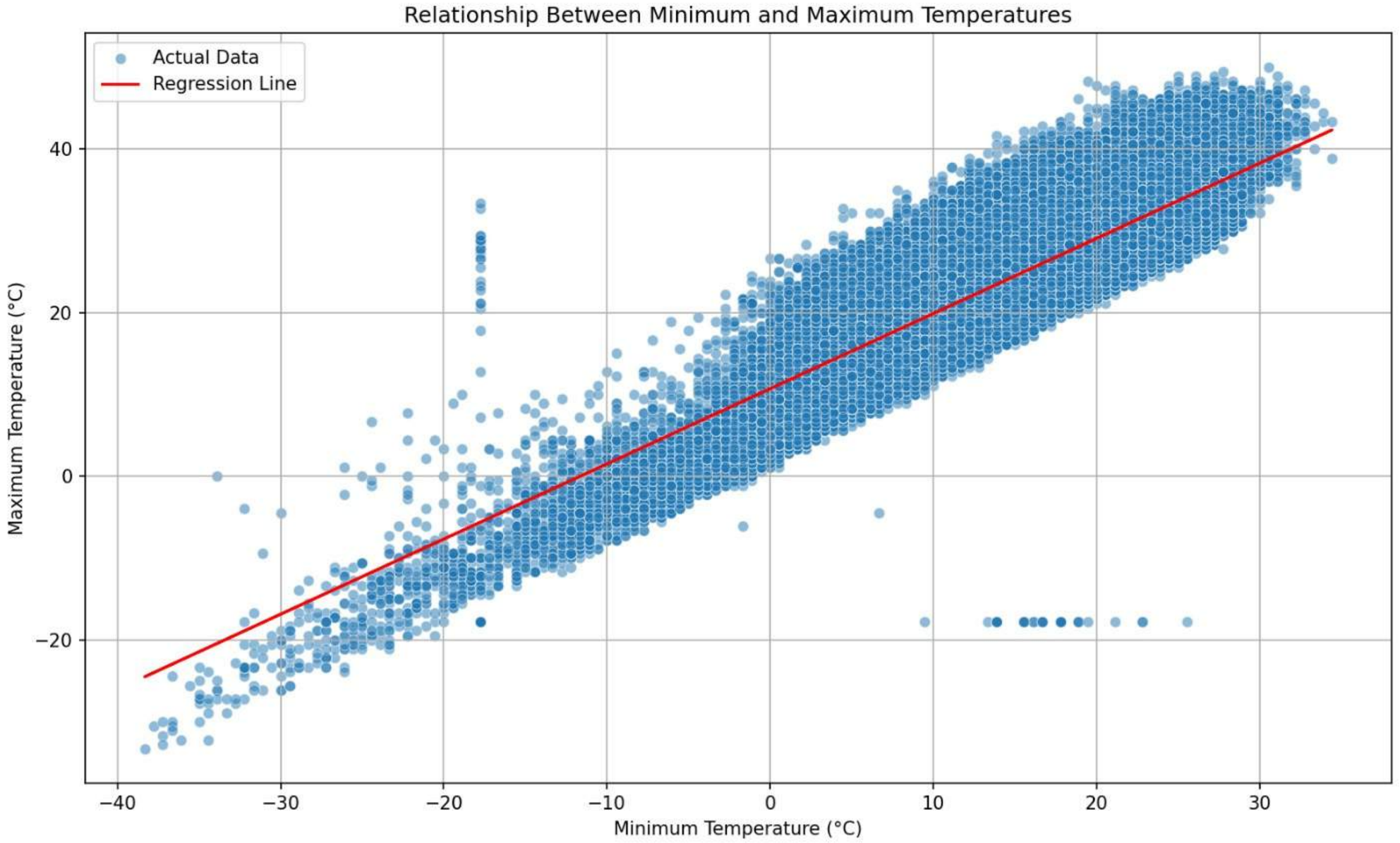
Reading data into the program for processing

Import Libraries

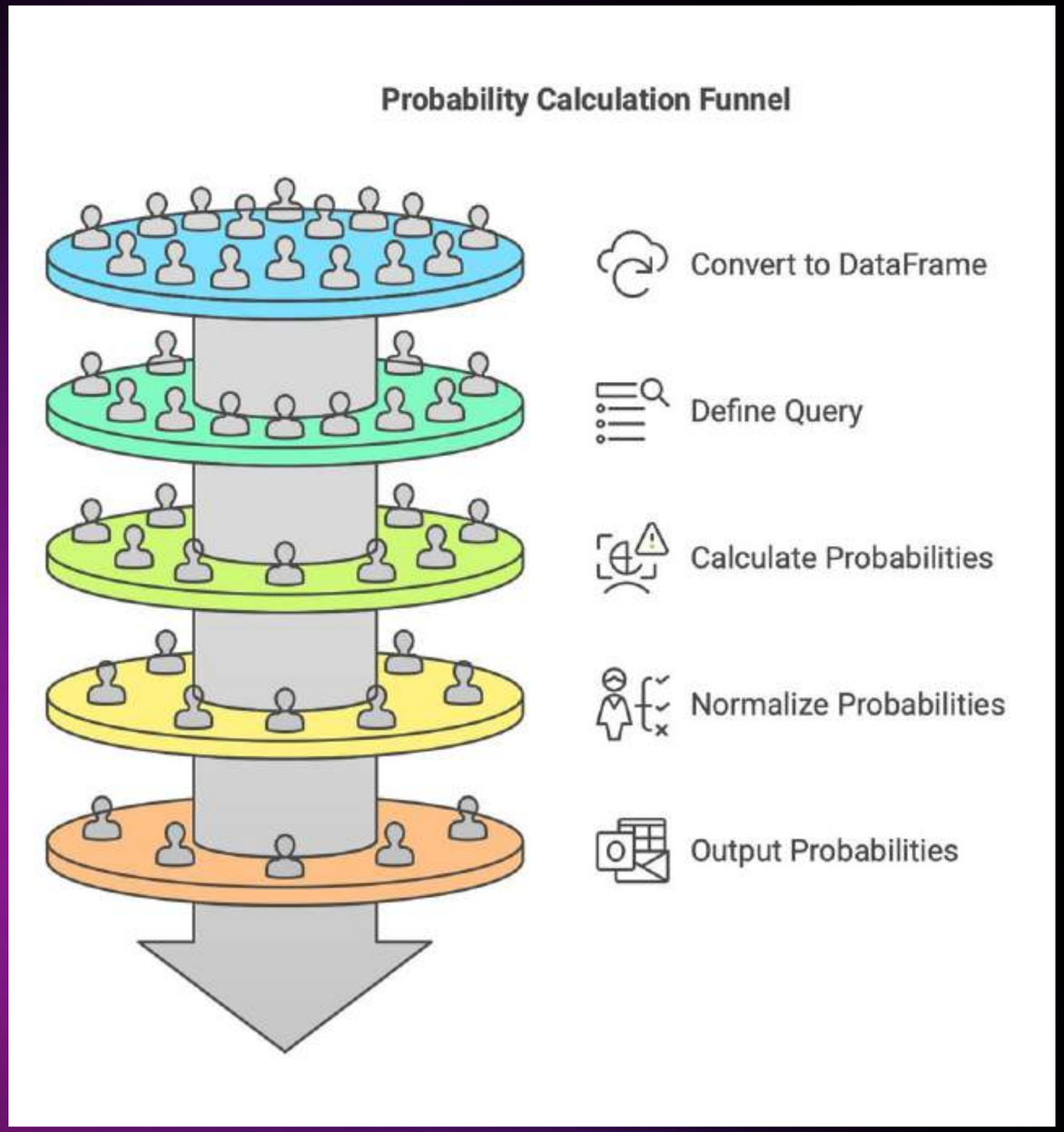
Loading necessary libraries for data manipulation and modeling



Output-1

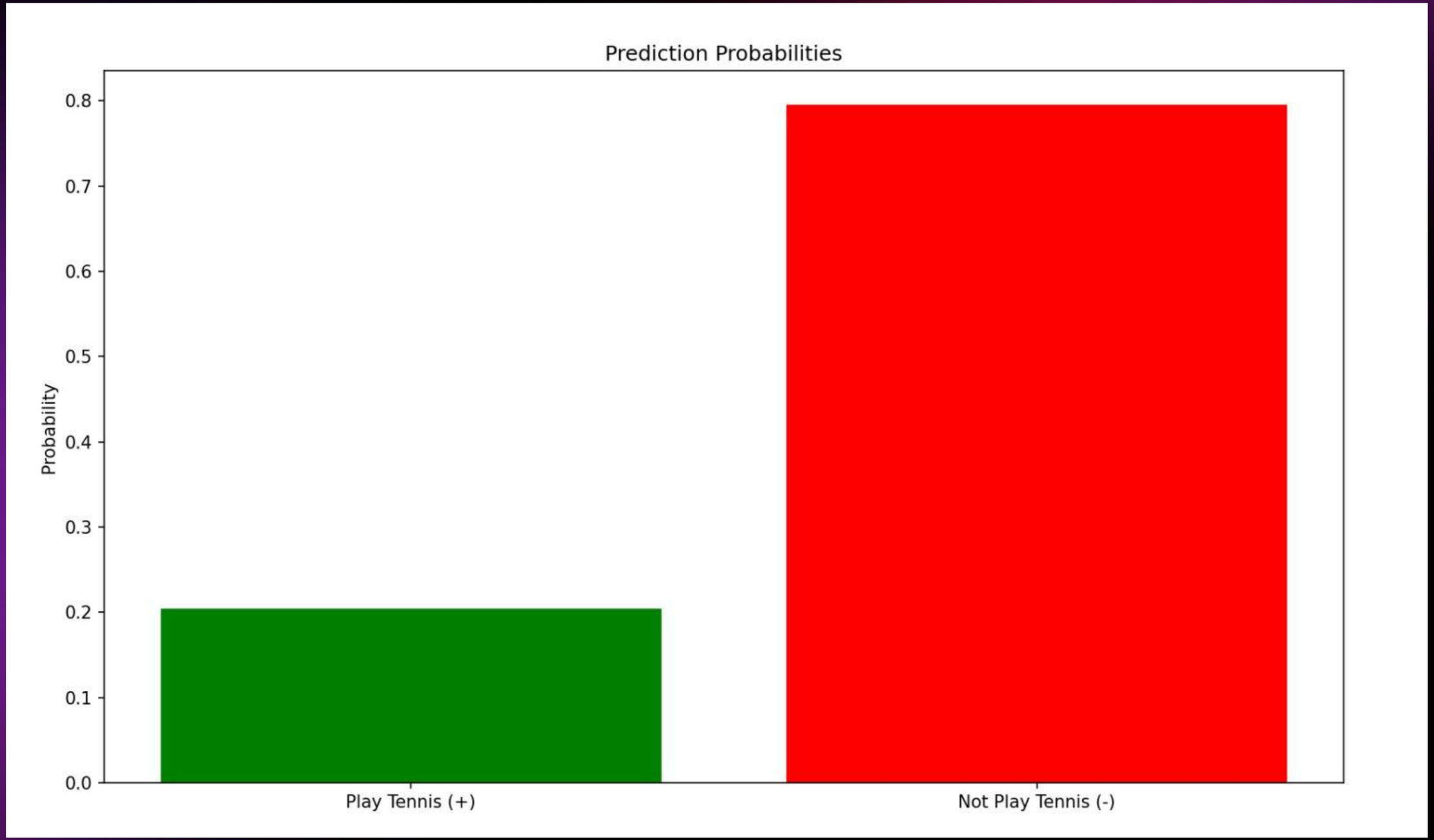

Relationship Between Minimum and Maximum Temperatures
Actual Data
Regression Line
Maximum Temperature (°C)
Minimum Temperature (°C)
40
20
0
−20
−40
−30
−20
−10
0
10
20
30

Task-3

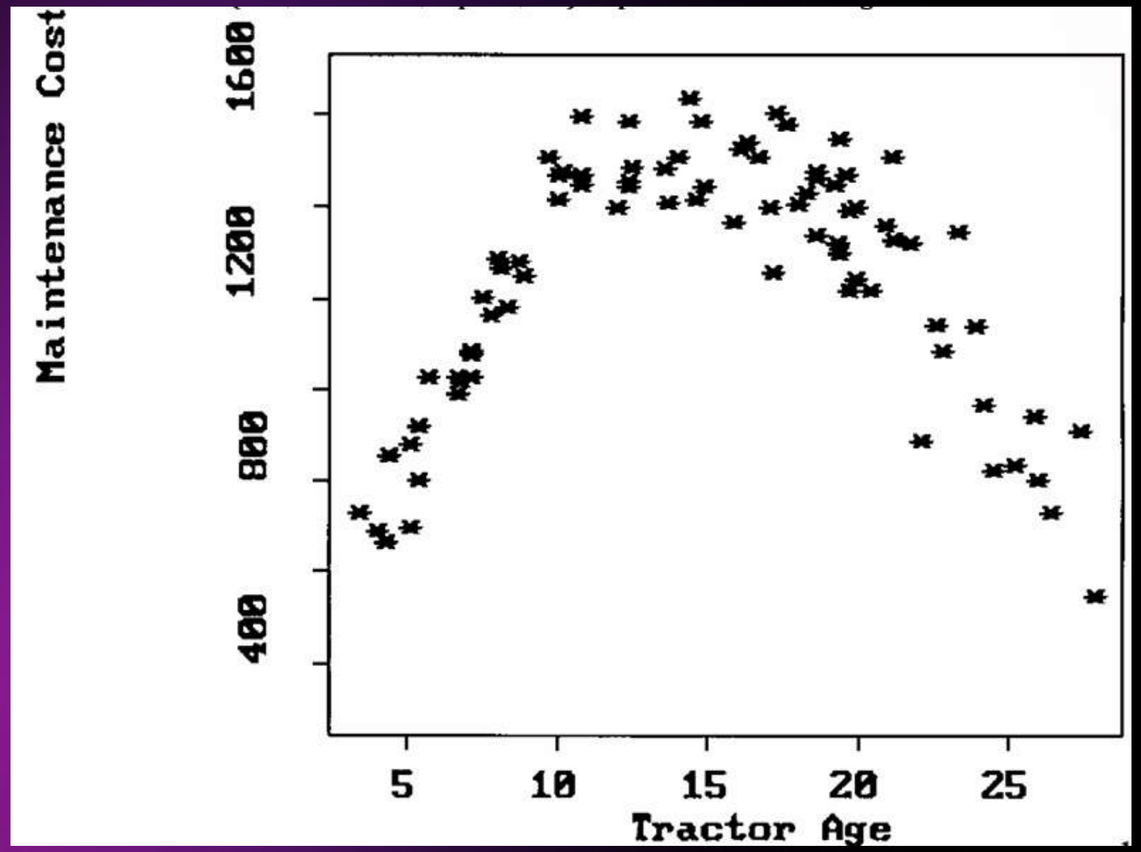


Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	−
sunny	hot	high	true	−
overcast	hot	high	false	+
rain	mild	high	false	+
rain	cool	normal	false	+
rain	cool	normal	true	−
overcast	cool	normal	true	+
sunny	mild	high	false	−
sunny	cool	normal	false	+
rain	mild	normal	false	+
sunny	mild	normal	true	+
overcast	mild	high	true	+
overcast	hot	normal	false	+
rain	mild	high	true	−

Output-2


Prediction Probabilities
Probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Play Tennis (+)
Not Play Tennis (-)

Task-4



Methodology:

1. Data Preparation

- The dataset consists of two numerical arrays:
- tractor_age: Represents the age of the tractors in years.
- maintenance_cost: Represents the corresponding maintenance cost in USD.

```
tractor_age = np.array([5, 8, 10, 12, 15, 18, 20, 23, 25])
maintenance_cost = np.array([800, 1100, 1400, 1550, 1600, 1450, 1300, 900, 700])
```

2. Polynomial Regression (Degree 2)

- Polynomial regression is used because the relationship between tractor age and maintenance cost appears to be non-linear.
- The Polynomial.fit() function is used to fit a second-degree polynomial (quadratic function) to the data.
- convert().coef extracts the coefficients of the polynomial in standard form.

```
coefficients = Polynomial.fit(tractor_age, maintenance_cost, deg=2).convert().coef
```

Task-4

3. Generate Regression Line Data

- We generate a smooth curve using 100 equally spaced points between 5 and 25.
- Using the polynomial equation, we compute the predicted maintenance cost (y_fit) for these points.

```
x_fit = np.linspace(5, 25, 100)
y_fit = coefficients[0] + coefficients[1] * x_fit + coefficients[2] * x_fit**2
```

4. Plotting the Results

Scatter Plot (Data Points)

- The original dataset is plotted as black dots to represent real-world data points.

```
plt.scatter(tractor_age, maintenance_cost, color='black', label='Data Points')
```

Plot Regression Curve

- The polynomial regression line is plotted in blue.

```
plt.plot(x_fit, y_fit, color='blue', label='Regression Line', linewidth=2)
```

Task-4

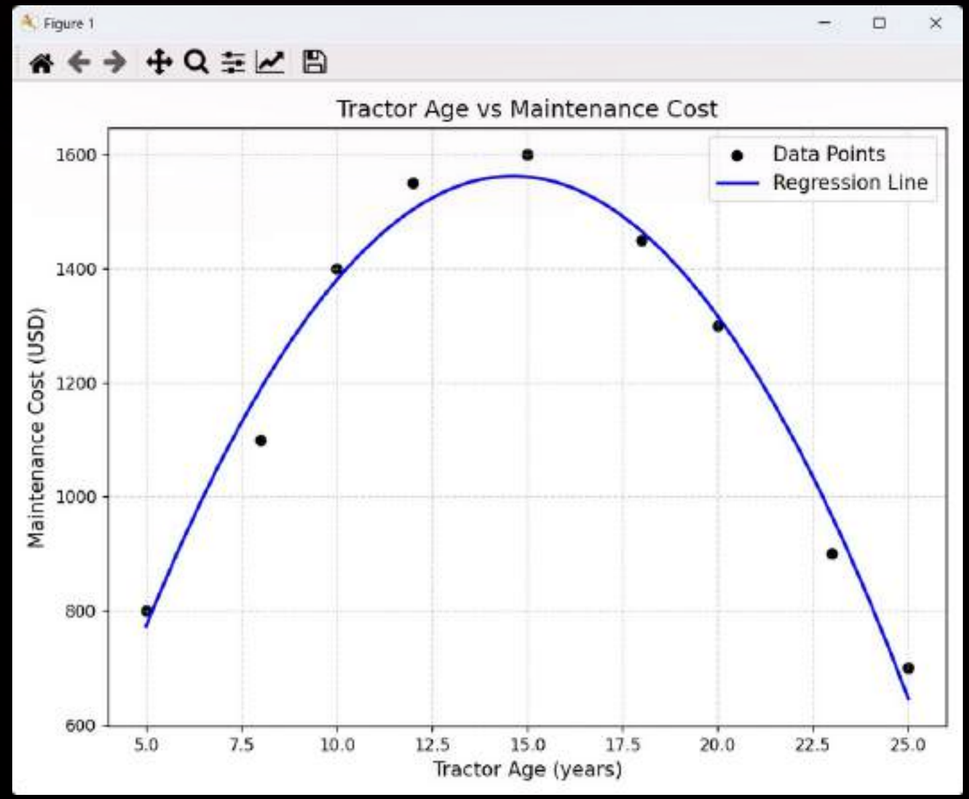
5. Customizing the Plot
   - Labels & Title: Improves readability.
   - Legend: Helps differentiate between data points and the regression curve.
   - Grid & Ticks: Enhances clarity.

```
# Add labels, title, and legend
plt.xlabel("Tractor Age (years)", fontsize=12)
plt.ylabel("Maintenance Cost (USD)", fontsize=12)
plt.title("Tractor Age vs Maintenance Cost", fontsize=14)
plt.legend(fontsize=12)

# Customize grid and ticks
plt.grid(True, linestyle='--', alpha=0.5)
plt.xticks(fontsize=10)
plt.yticks(fontsize=10)
```

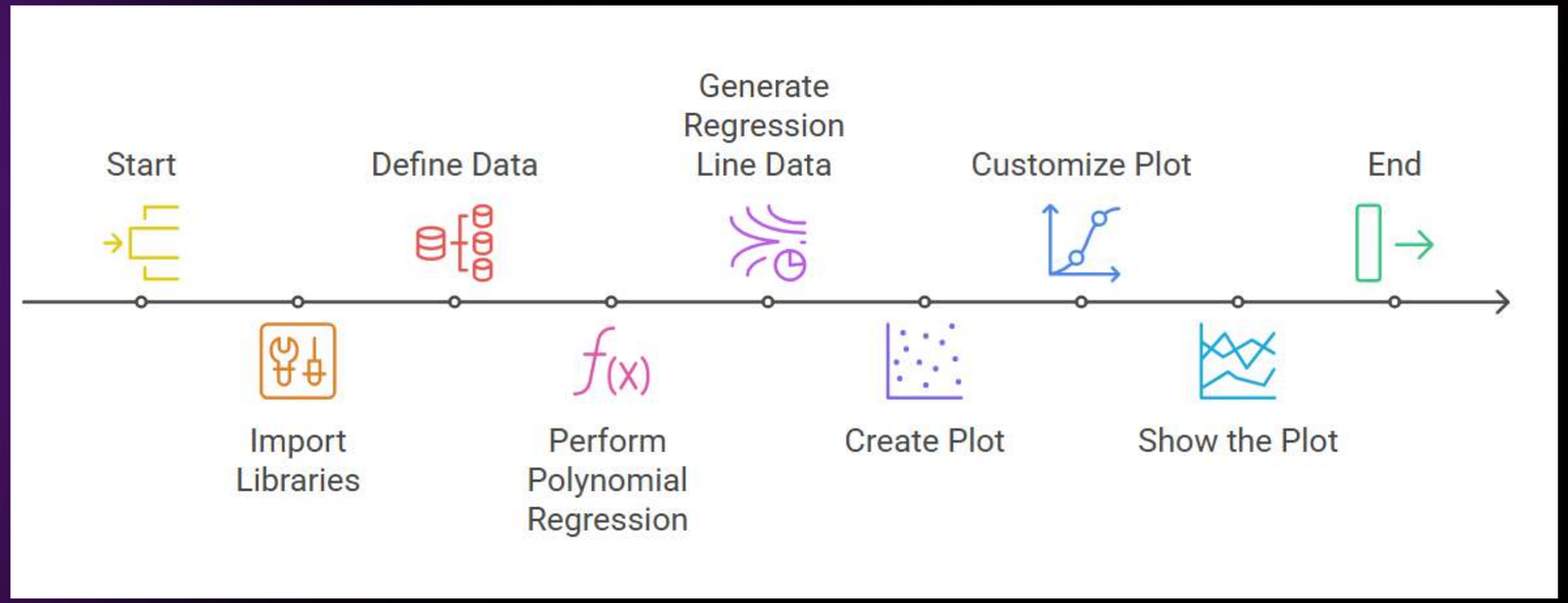
6. Display the Plot
   - Ensures proper layout and displays the final visualization.

```
plt.tight_layout()
plt.show()
```



Task-4

Flowchart



Woxsen University

Thank You!