Machine learning

Assignment-2

Presented by:

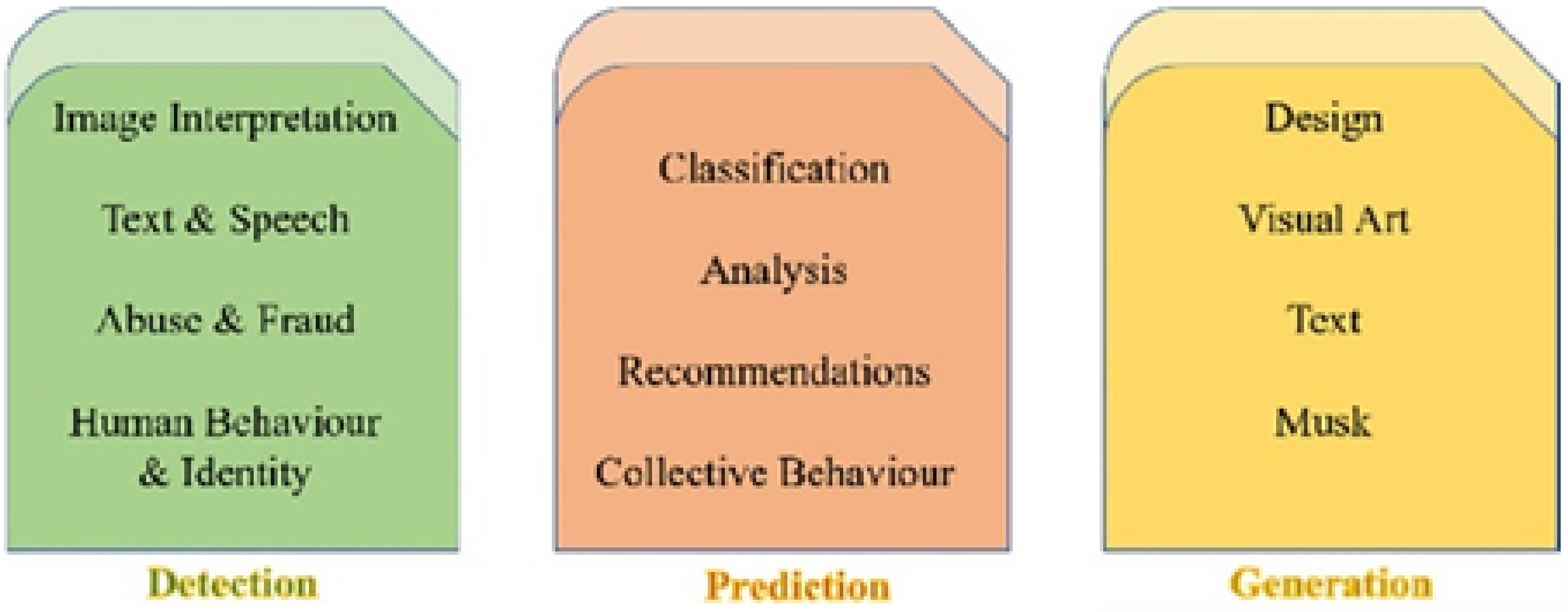
N. Prajnika 23WU0102139
Indra Teja 23WU0102095
Krishi Valleru 23WU0102099
Nikhitha Yerra 23WU0102134
Kundan M. 23WU0102101

Next Slide

TASK-1

Example use cases for ML

ML has a variety of applications, ranging from automation, detection and personalization to things like optimization.



Sub-fields in machine learning

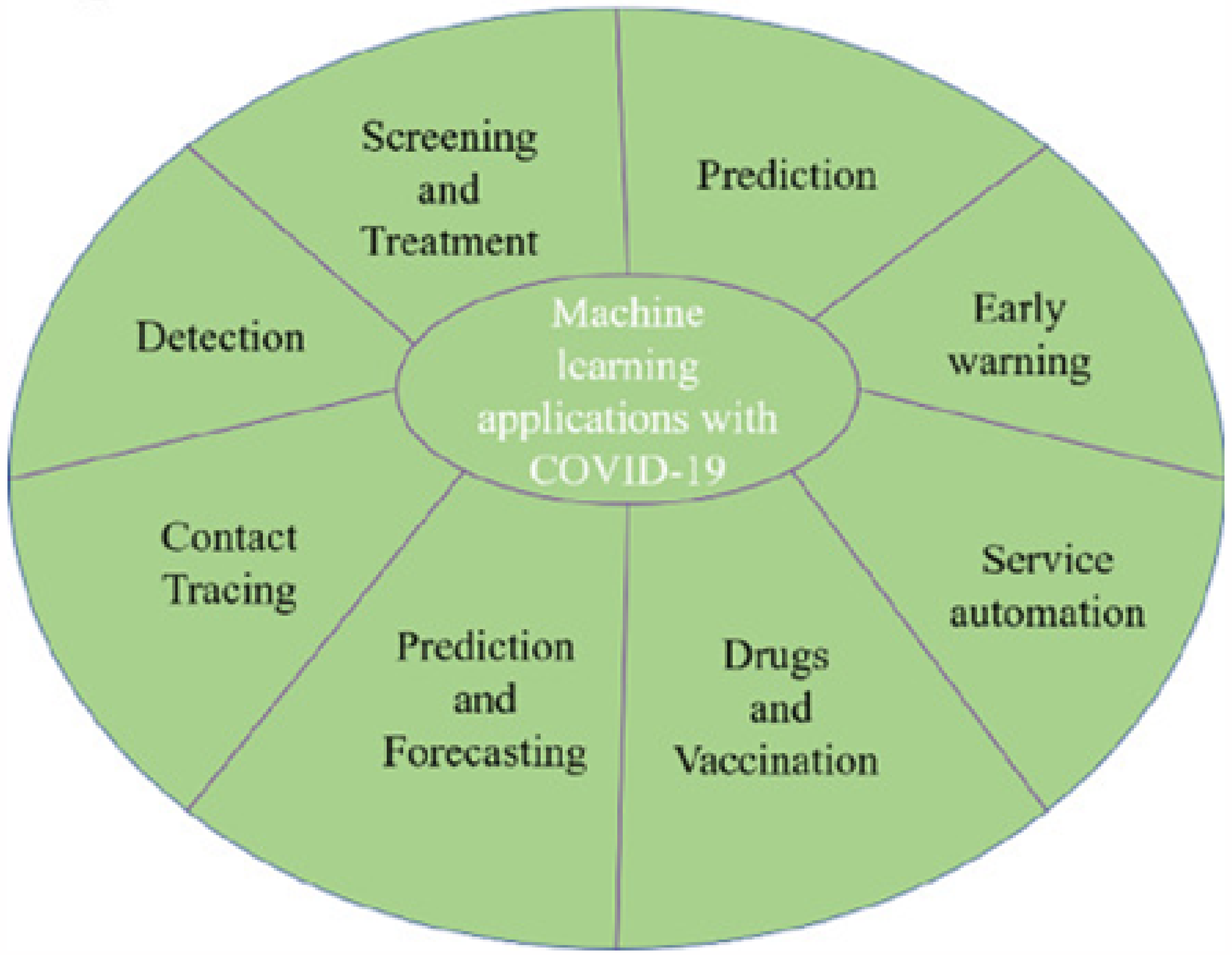


Image Source: https://www.sciencedirect.com/science/article/pii/S2666285X21000042#fig0003

Next Slide

TASK-1

Identifying Good Problems for ML

We can identify applications for Machine Learning by looking for ideal traits like:

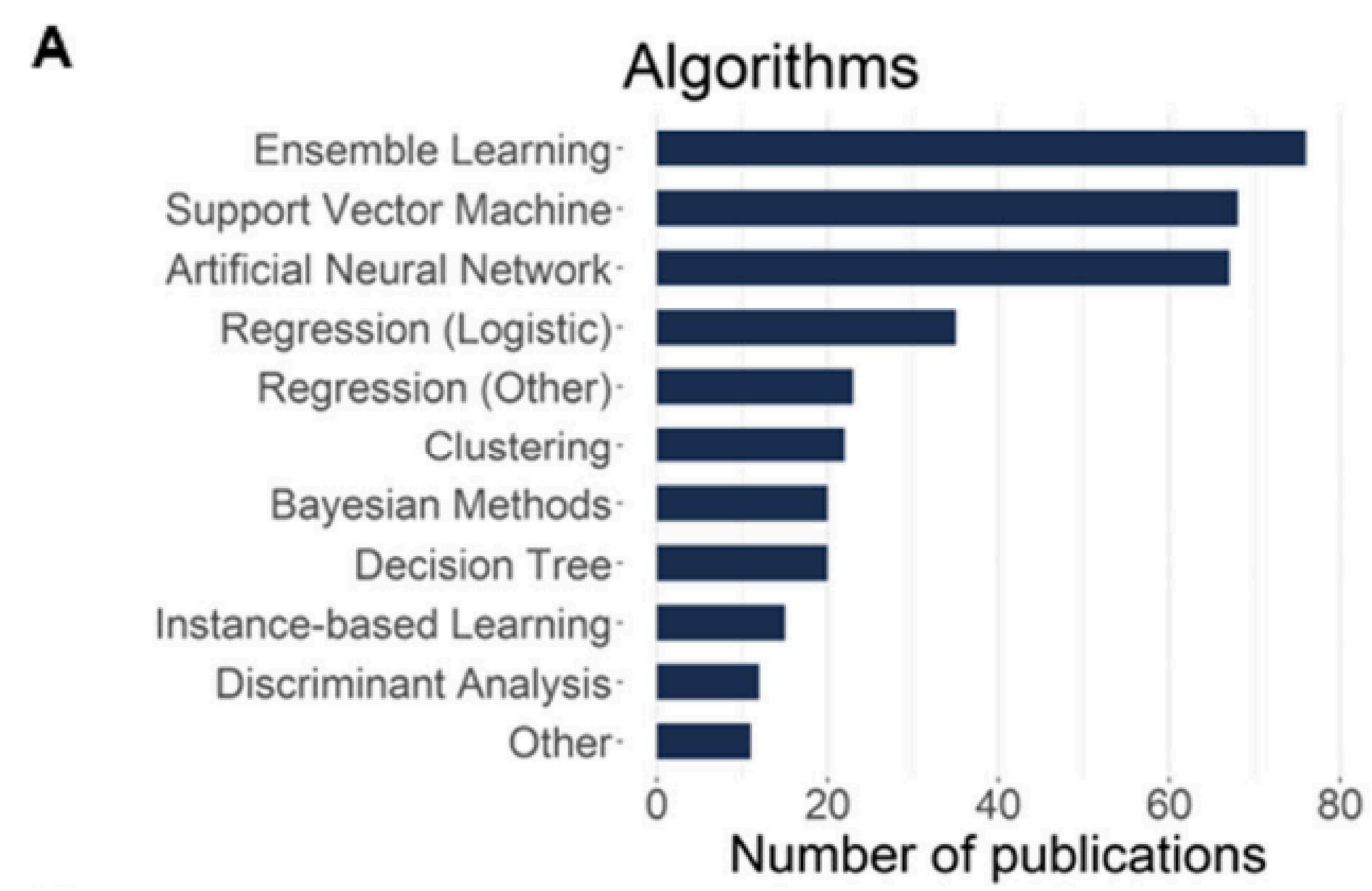
- Large size, high quality datasets are available.
- Clear objectives exist (e.g., predicting, classifying, or clustering).
- Data patterns or correlations can be extracted from the given data.

Next Slide

TASK-1

Hard ML Problems

- Datasets may have less or incorrect data.
- Can degrade model performance.
- Ethical concerns such as privacy issues or bias in data.
- Ensuring models perform well on unseen, real-world data beyond training datasets.



Schaefer, J., Lehne, M., Schepers, J. et al. The use of machine learning in rare diseases: a scoping review. Orphanet J Rare Dis 15, 145 (2020). https://doi.org/10.1186/s13023-020-01424-6

Next Slide

TASK-1

References

Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x

Khan, M. A., Bahadur, A., Khan, M. S., et al. (2021). Machine learning: Algorithms, real-world applications, and research directions. SN Computer Science, 2(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Schaefer, J., Lehne, M., Schepers, J. et al. The use of machine learning in rare diseases: a scoping review. Orphanet J Rare Dis 15, 145 (2020). https://doi.org/10.1186/s13023-020-01424-6

Next Slide

TASK-2

Problem 3: Is there a relationship between the daily minimum and maximum temperature? Can you predict the maximum temperature given the minimum temperature?

To create the flowchart, we used the Graphviz library to visualize the workflow of our temperature analysis process.

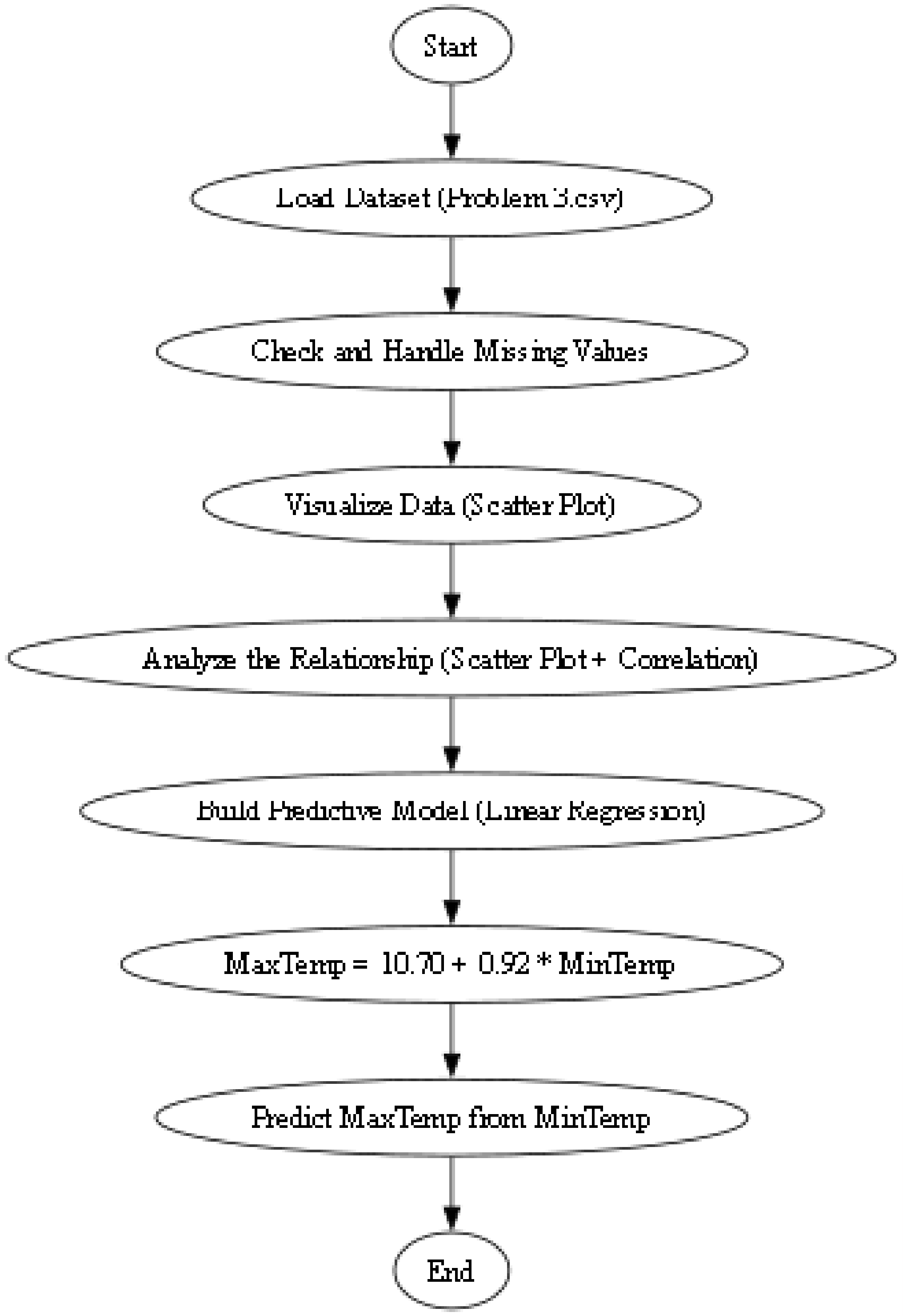
1. Nodes: Represented each key step in the process (e.g., loading data, cleaning data, visualization, regression modeling, and prediction).
2. Edges: Connected the steps in logical order to show the flow of actions.
3. Regression Equation: Included the linear equation from the model as part of the flow.
4. Output: Rendered the flowchart as a PNG image for use in the presentation.

This flowchart visually summarizes the entire analytical process from data input to prediction.

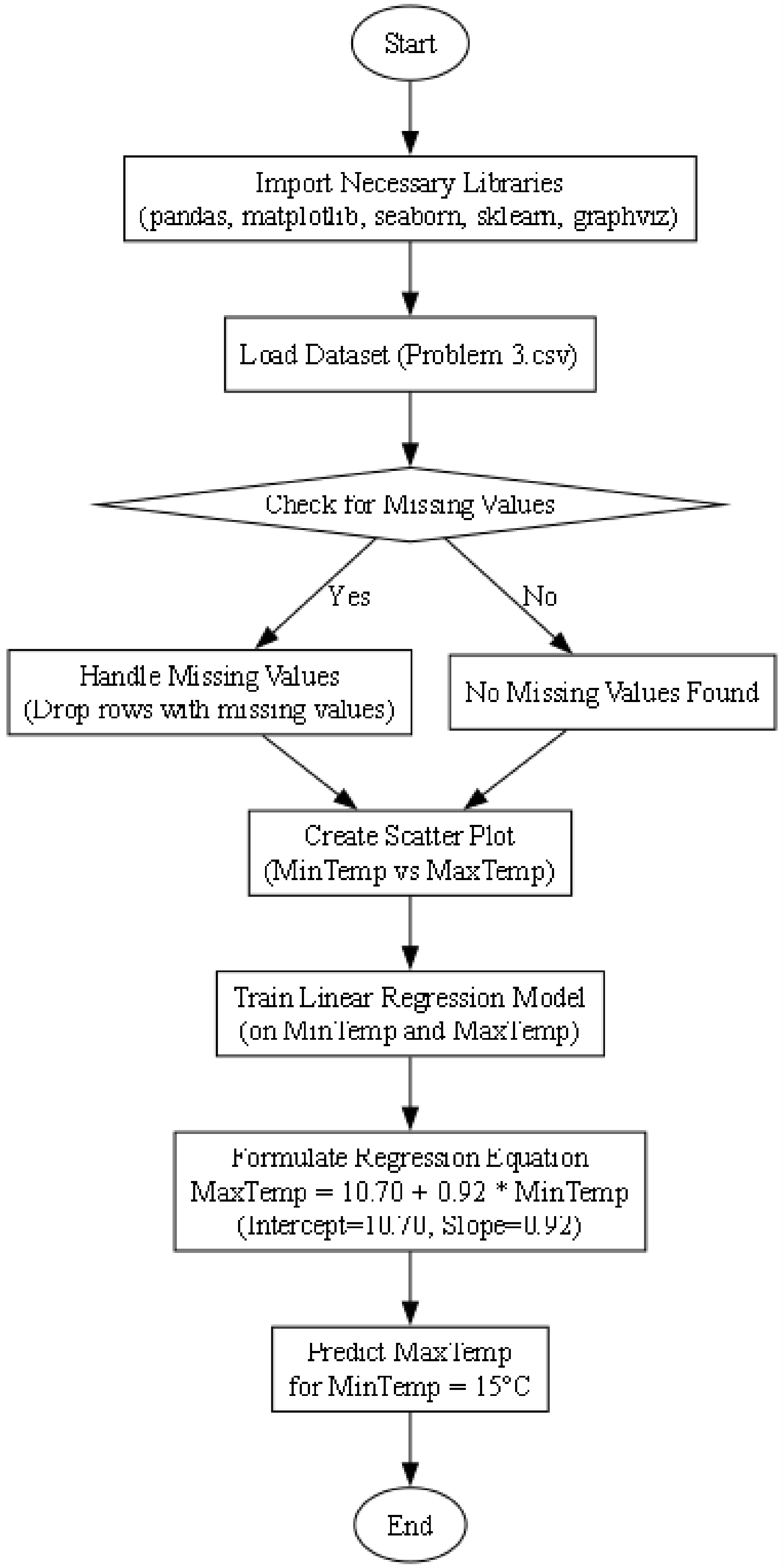
Next Slide

TASK-2

Initial Result:



Final Result:

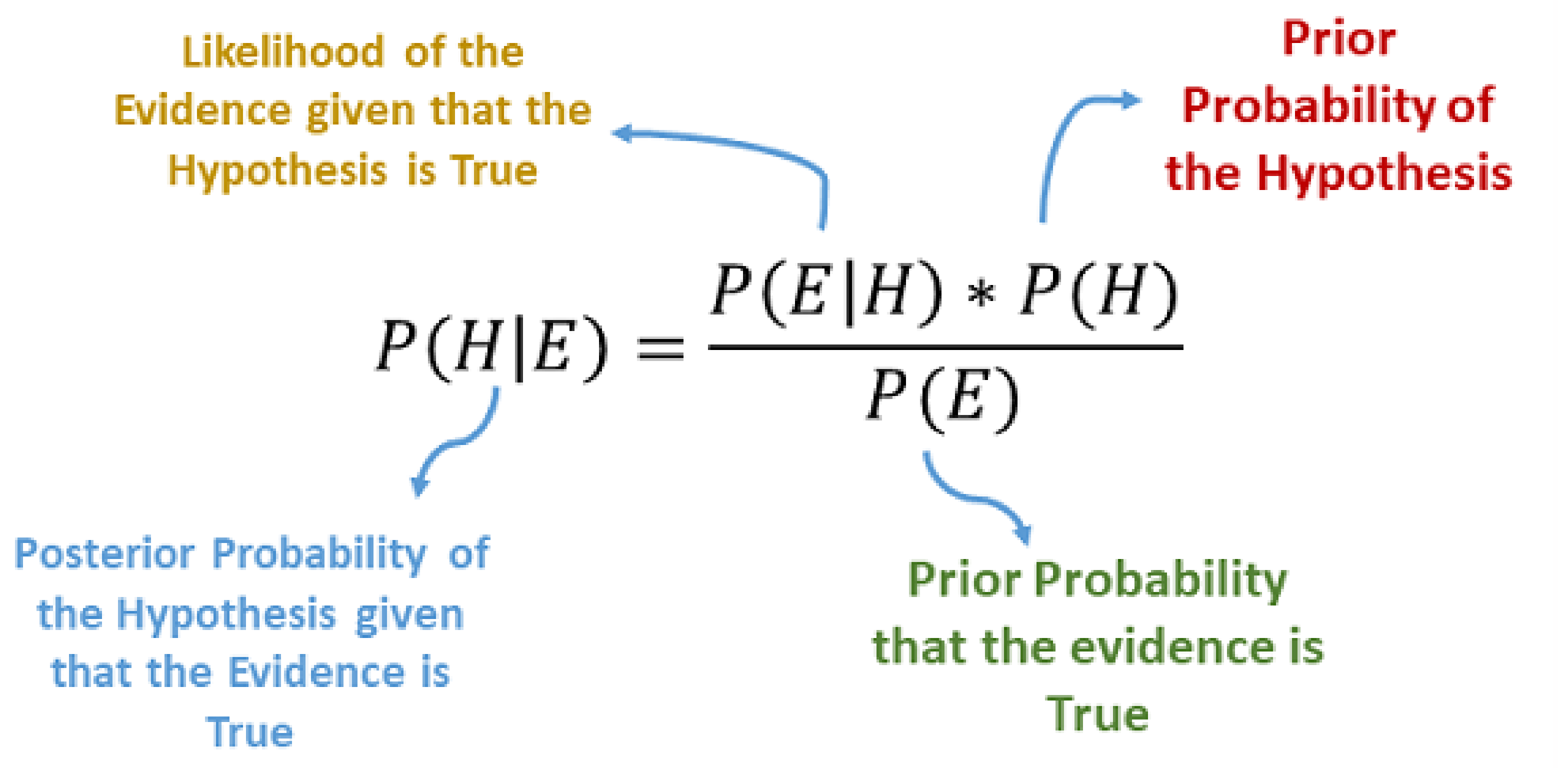


Next Slide

TASK-3

“Predict playing tennis when <sunny, cool, high, strong> What probability should be used to make the prediction? How to compute the probability?”

Naive Bayes Classification:
P(Play | sunny, cool, high, strong)
This represents the likelihood of playing tennis given specific weather conditions.



Next Slide

TASK-3

```python
import pandas as pd
from sklearn.naive_bayes import MultinomialNB
from sklearn.preprocessing import LabelEncoder

# Dataset
data = {
    "Outlook": ["sunny", "sunny", "overcast", "rain", "rain", "rain", "overcast", "sunny",
                "sunny", "rain", "sunny", "overcast", "overcast", "rain"],
    "Temperature": ["hot", "hot", "hot", "mild", "cool", "cool", "cool", "mild", "cool",
                    "mild", "mild", "mild", "hot", "mild"],
    "Humidity": ["high", "high", "high", "high", "normal", "normal", "normal", "high",
                 "normal", "normal", "normal", "high", "normal", "high"],
    "Windy": ["false", "true", "false", "false", "false", "true", "true", "false", "false",
              "false", "true", "true", "false", "true"],
    "Class": ["-", "-", "+", "+", "+", "-", "+", "-", "+", "+", "+", "+", "+", "-"]
}

# Create a DataFrame
df = pd.DataFrame(data)

# Create separate LabelEncoders for each feature column
encodes = {}
for col in ["Outlook", "Temperature", "Humidity", "Windy"]:
    encodes[col] = LabelEncoder()
    df[col] = encodes[col].fit_transform(df[col])

# Encode the target variable with a separate encoder
encode_Class = LabelEncoder()
df["Class"] = encode_Class.fit_transform(df["Class"])

# Split features (X) and target (y)
X = df[["Outlook", "Temperature", "Humidity", "Windy"]]
```

```python
X = df[["Outlook", "Temperature", "Humidity", "Windy"]]
y = df["Class"]

# Model
model = MultinomialNB()
model.fit(X, y)

# Predict for <sunny, cool, high, strong>
# Use the correct encoders for each feature to transform the sample data
sample = pd.DataFrame({
    "Outlook": encodes["Outlook"].transform(["sunny"]),
    "Temperature": encodes["Temperature"].transform(["cool"]),
    "Humidity": encodes["Humidity"].transform(["high"]),
    "Windy": encodes["Windy"].transform(["true"])
})

# Predict and decode the result
prediction = model.predict(sample)
decoded_prediction = encode_Class.inverse_transform(prediction)

# Print Prediction
print("Prediction:", decoded_prediction[0])

# Inspect Probabilities
prob = model.predict_proba(sample)
print("Probability of +:", prob[0][1])
print("Probability of -:", prob[0][0])
```

```
Prediction: -
Probability of +: 0.6406632947915922
Probability of -: 0.35933670520840805
```

Next Slide

TASK-4

Investigate the relationship between the age of farm tractors and their maintenance costs using data collected from Putnam County, FL.

Data Points:

- Tractor Age (5,10,15,20,25)
- Maintenance Costs (600,1400,1400,1200,800)

Data Characteristics:

- Maintenance costs peak at mid-life (15 years).
- Costs reduce after the mid-life stage.

Next Slide

TASK-4

Model Used:

Polynomial Regression (2nd Degree),4th degree and spline

Observation points:

1. Maintenance costs are highest at around 15 years of tractor age.
2. Younger tractors have relatively lower costs.
3. Older tractors beyond 20 years also show declining maintenance costs.

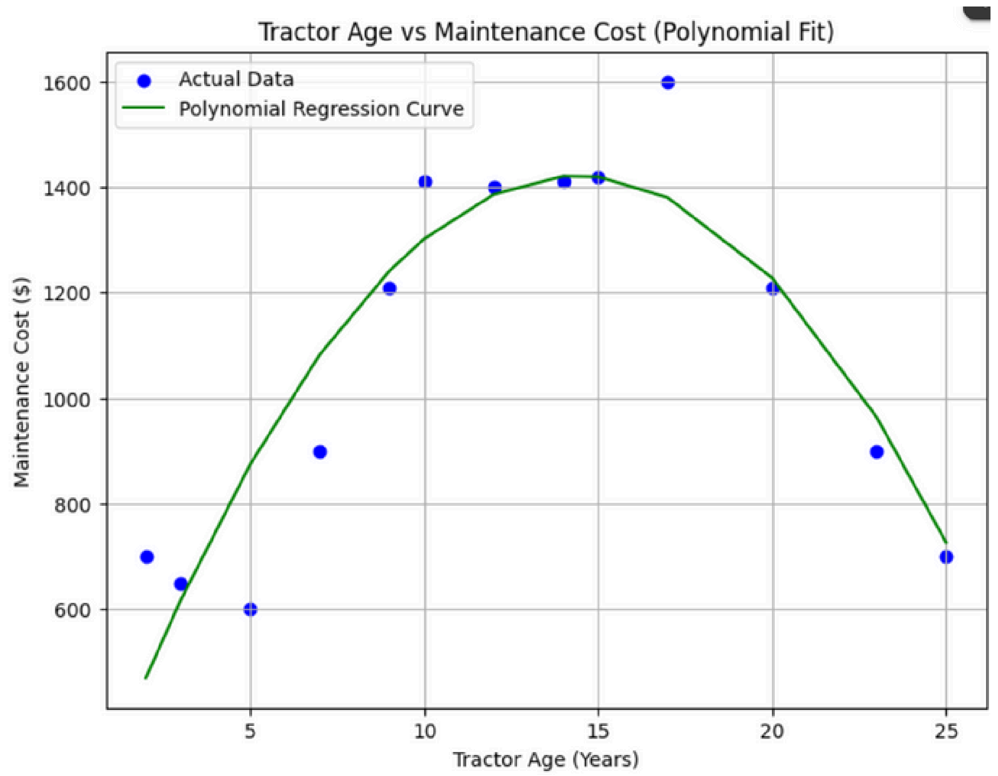
Next Slide

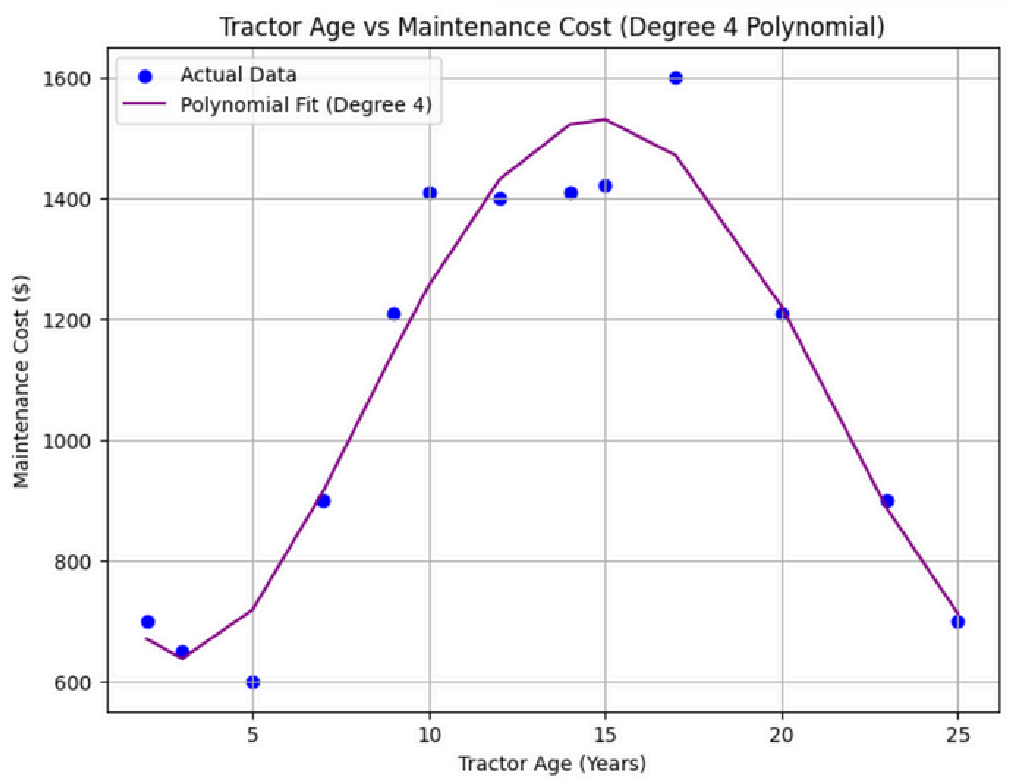
TASK-4

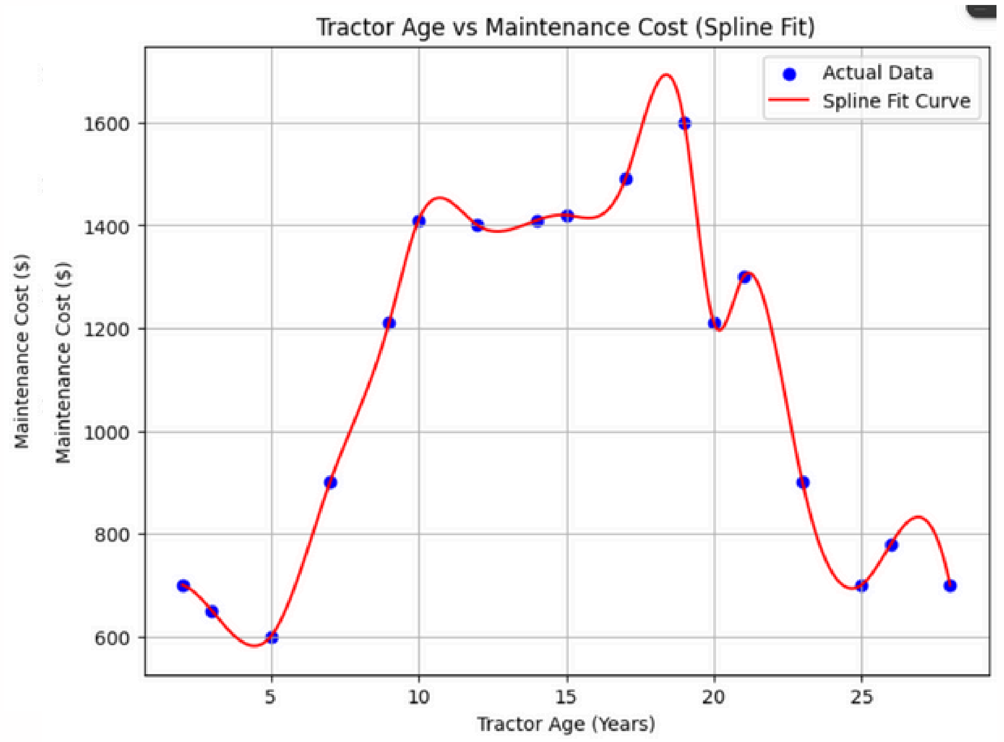
Practical Implications for Farmers:

Helps estimate maintenance budgets based on tractor age.

Guides decision-making for tractor replacement or servicing.







Next Slide

THANK YOU