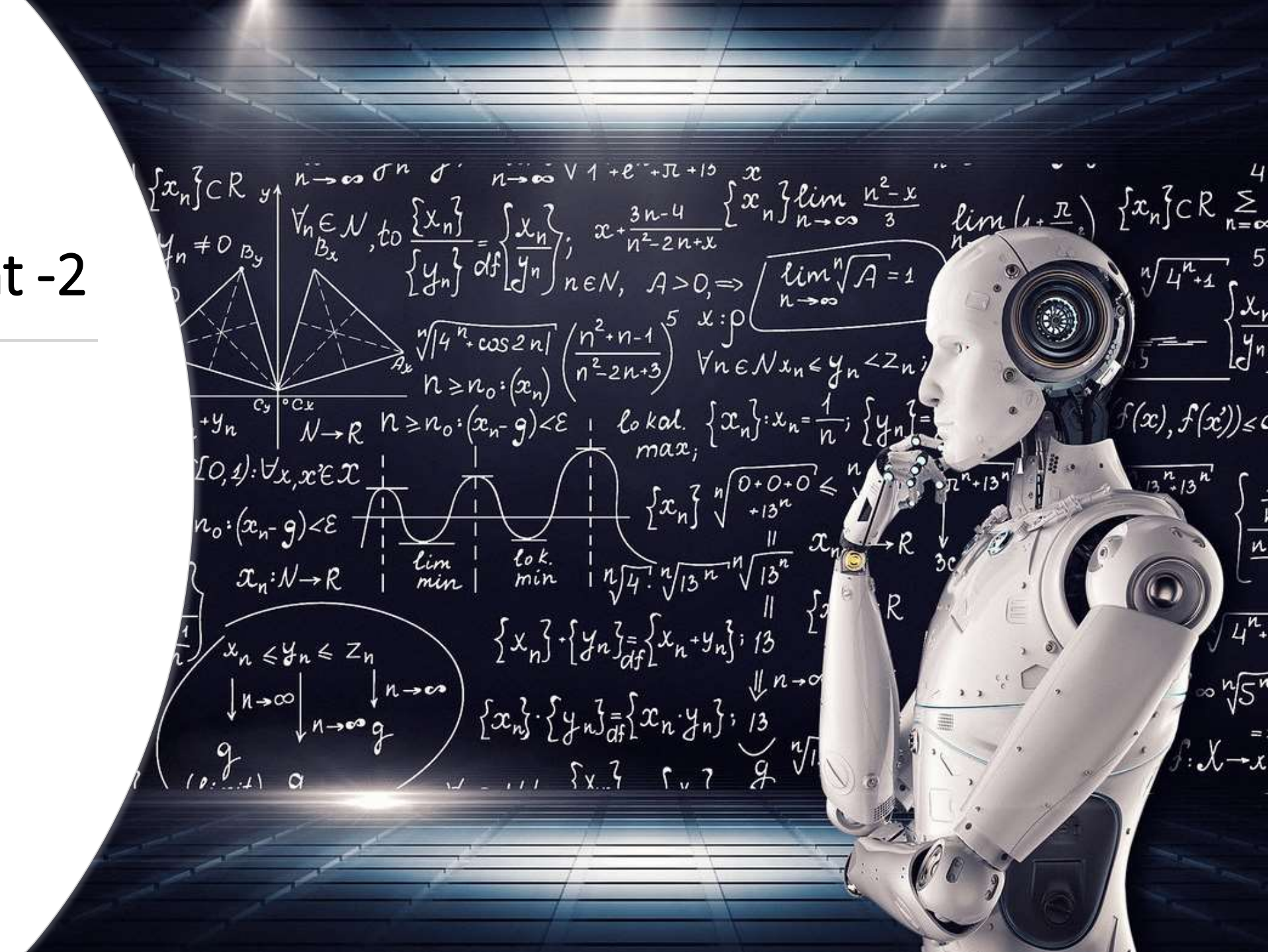
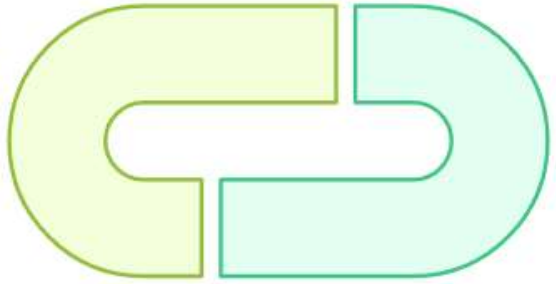
ML Assignment -2

Team members:

- Kesava Datta
- Sam Dheeraj
- Lakshmi Shivani
- Kritika Reddy



Task 1

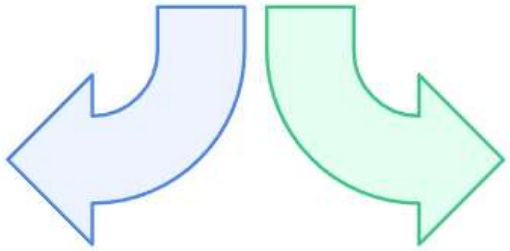


Inductive Bias

The assumptions guiding model selection from the hypothesis space.

Hypothesis Space

The set of all possible models a learning algorithm can consider.



Classification

Used for predicting discrete labels using models like Logistic Regression and Decision Trees.

Regression

Used for predicting continuous values using models like Random Forest and LSTM.

Reference:

- Kumari, S., & Singh, S. K. (2023). Machine learning-based time series models for effective CO2 emission prediction in India. *Environmental Science and Pollution Research*, *30*(55), 116601-116616.
- Baxter, J. (2000). A model of inductive bias learning. *Journal of artificial intelligence research*, *12*, 149-198.
- Loh, W. Y. (2011). Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, *1*(1), 14-23.

Problem Statement:

Now a days, Co2 emission is increasing rapidly due to increase of population. For the predicting the Co2 emission using machine learning models of different types of vehicles.

Input Features:

- **Engine Size**: 2.0L
- **Cylinders**: 4
- **Vehicle Class**: Compact

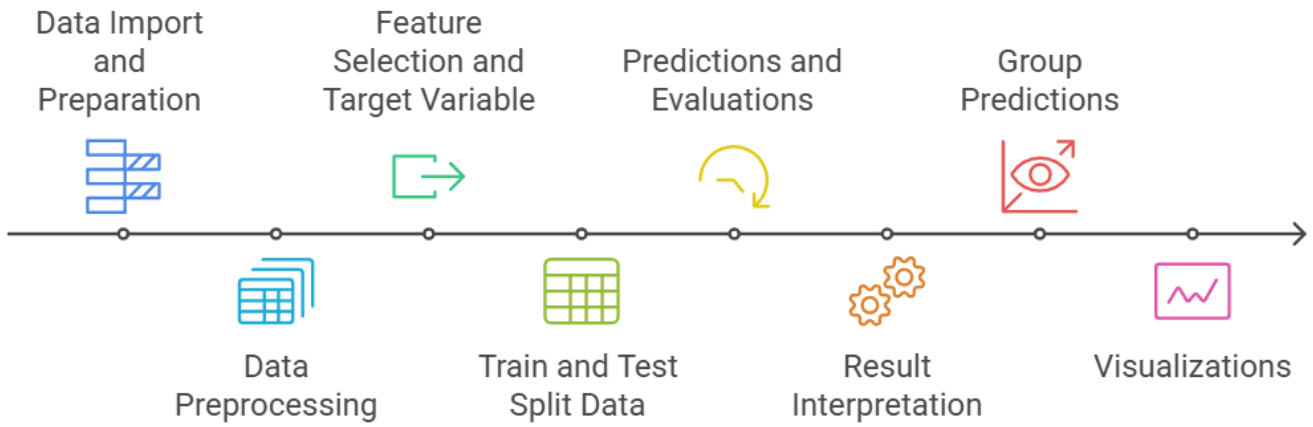
Random Forest Prediction:

- **100 trees** predict CO2 emissions:
- Tree 1: 180 g/km
- Tree 2: 190 g/km
- Tree 3: 185 g/km
- ...
- Tree 100: 195 g/km

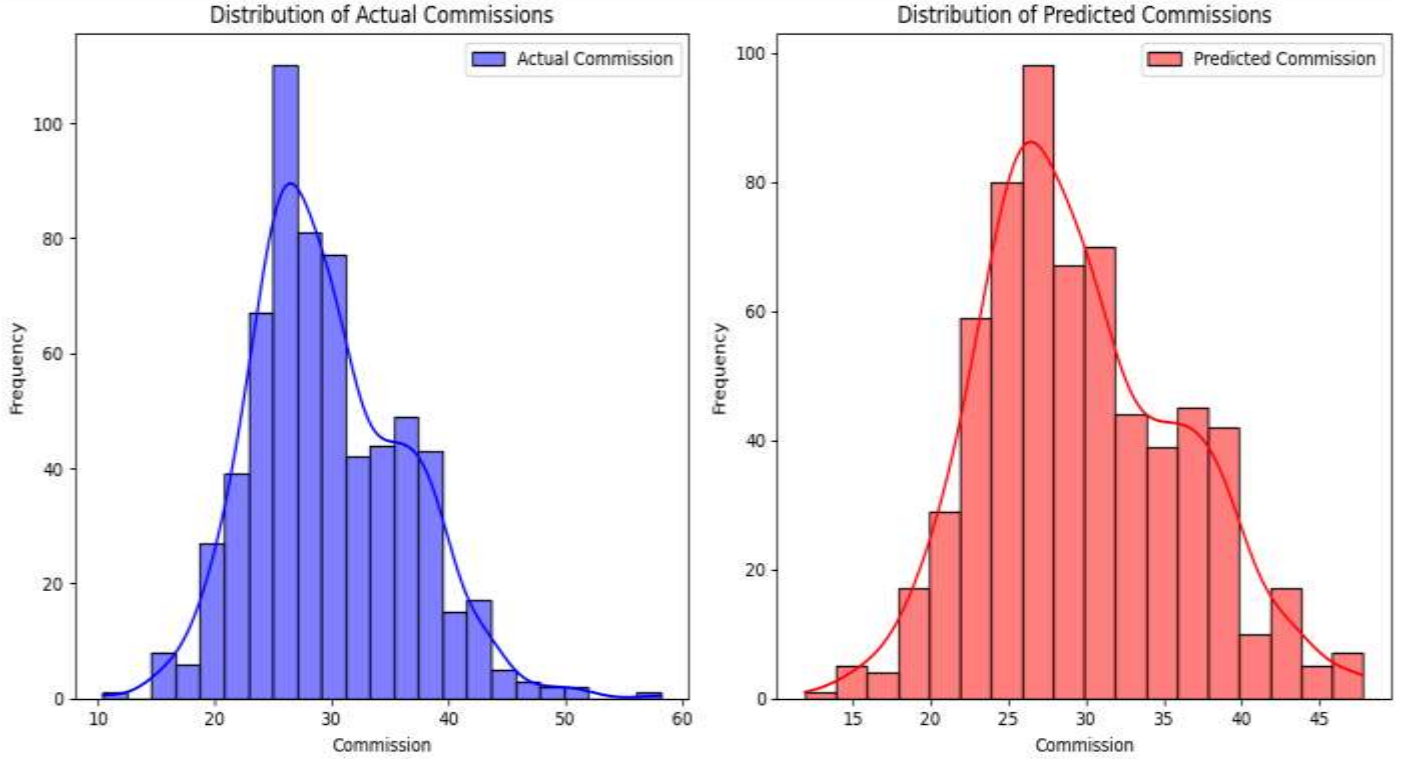
Final Prediction:

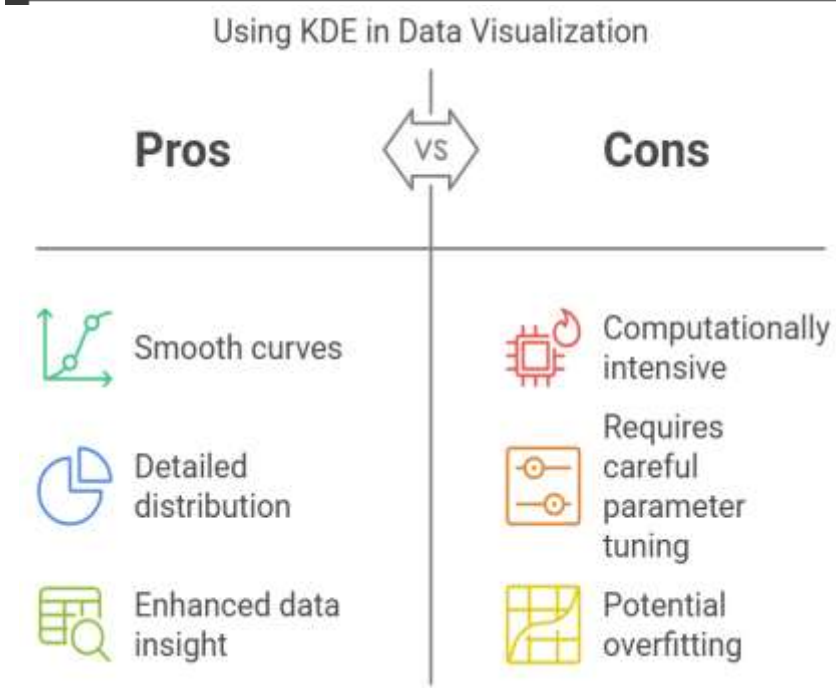
- **Average of all predictions**:188g/km
- **Predicted CO2 emissions**: 188 g/km

Task 2




```
Number of rows before dropping missing values: 640
Number of rows after dropping missing values: 639

Mean Absolute Error: 0.1708046875000062

Predicted Commissions for the first few vehicles:
    MAKE    MODEL  Predicted Commission
1  ACURA    1.6EL                21.596
2  ACURA    1.6EL                20.506
3  ACURA    3.2TL                26.500
4  ACURA    3.5RL                30.092
5  ACURA  INTEGRA                23.000

Predicted CO2 Emissions (Commissions) by Make and Model for All Cars:
      MAKE              MODEL  Predicted Commission
0    ACURA             1.6EL                21.051
1    ACURA             3.2TL                26.500
2    ACURA             3.5RL                30.092
3    ACURA           INTEGRA                22.455
4    ACURA  INTEGRA GSR/TYPE R                22.300
..     ...               ...                   ...
324  VOLVO               V70                25.650
325  VOLVO       V70 AWD TURBO                28.800
326  VOLVO       V70 GLT TURBO                27.400
327  VOLVO        V70 T5 TURBO                26.700
328  VOLVO      V70R AWD TURBO                29.918

[329 rows x 3 columns]
```

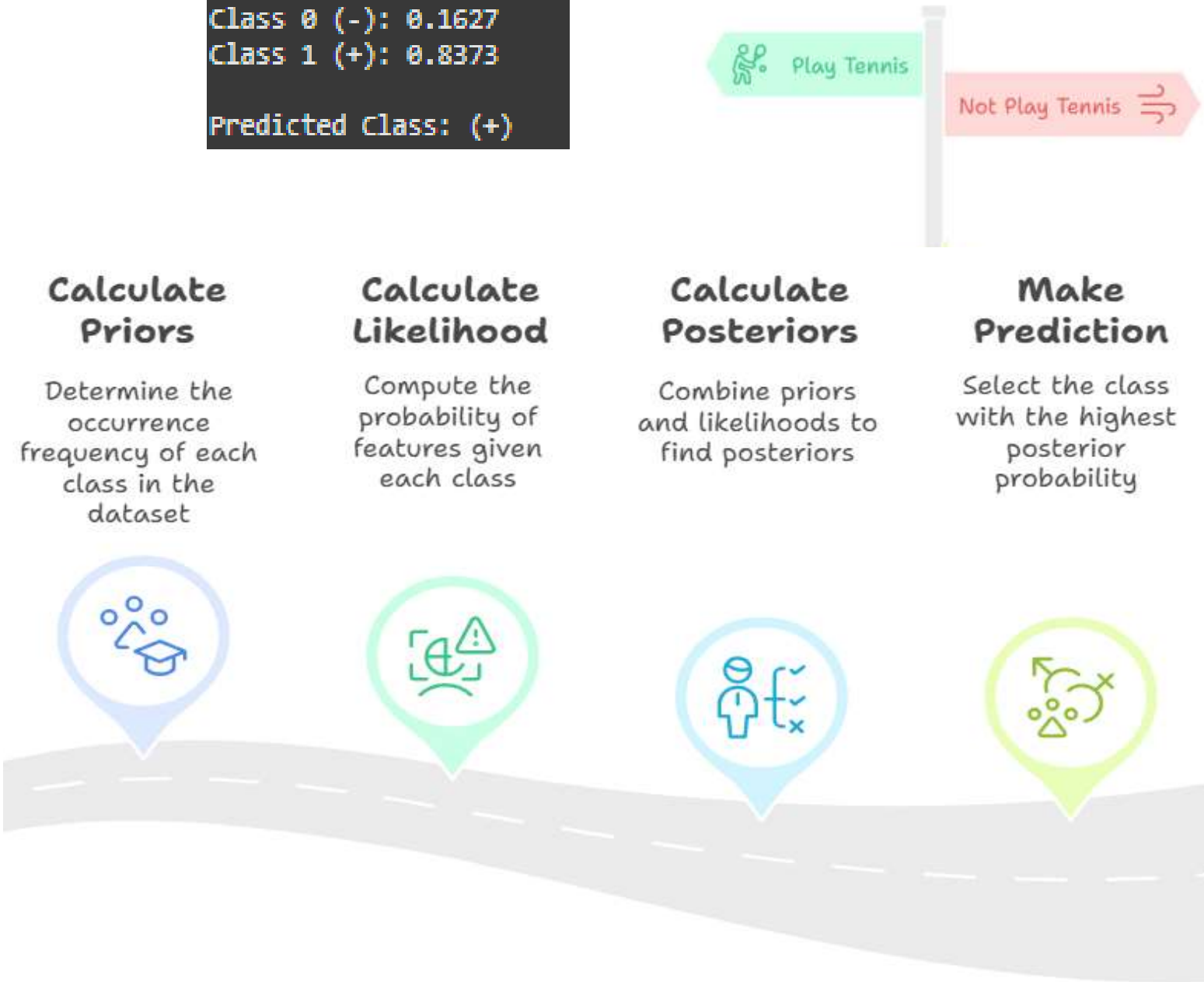




Task 3

Result:

```
Class Probabilities:
Class 0 (-): 0.1627
Class 1 (+): 0.8373

Predicted Class: (+)
```



Task 4

R-squared (R^2)

Value: 1.0

Perfect explanation of relationship

Based on simulated data

Model Evaluation

Mean Squared Error (MSE)

Value: 6.540756190228377e-24

Indicates near-identical predictions

Identify Relationship

Analyzing tractor age and maintenance cost

Select Polynomial Regression

Choosing model for curved data trends

Split Data

Dividing data into training and testing sets

Train Model

Using training data to build the model

Test Model

Evaluating model accuracy on unseen data