Faculty of Computers and Artificial Intelligence

AI Department

2023/2024

Machine Learning Project

	ID	Name	Grade
1.	20210364	زياد ضياء عبدالله	
2.	20211053	يس ضياء الدين احمد نبيل الشرقاوى	
3.	20210195	اواب محمود الطرابيلي	
4.	20210239	تسنيم محسن محمود	
5.	20210470	شهد محمد عبد السلام	
6.	20210275	حبيبه انور شافعى محمد	

I. NUMERICAL DATASET

1. Project Introduction

a. Dataset Name

CO2 Emissions in Canada

b. Number of classes and their labels

N/A

c. Dataset Samples Numbers

Total number of samples: (6720x7)

d. Training, Validation and Testing

Number of samples used in training: (80% of the dataset)
Number of samples used in testing: (20% of the dataset)

2. Implementation Details

a. Extracted Features

'Year', 'Engine Size(L)', 'Cylinders', 'Fuel Consumption Comb (L/100 km)', 'CO2 Emissions(g/km)'

b. Cross-validation

Cross-validation used: NO

c. Linear Regression

℘ Hyper-parameters

Hyper-parameters:
Initial learning rate: N/A
Optimizer: N/A
Regularization: ridge
Batch size: N/A
Number of epochs: N/A

d. K-Nearest Neighbors (KNN)

℘ Hyper-parameters

Number of neighbors: (3)
Leve size 10
'algorithm': 'auto', 'leaf_size': 10, 'n_neighbors': 3, 'p': 1, 'weights': 'distance'

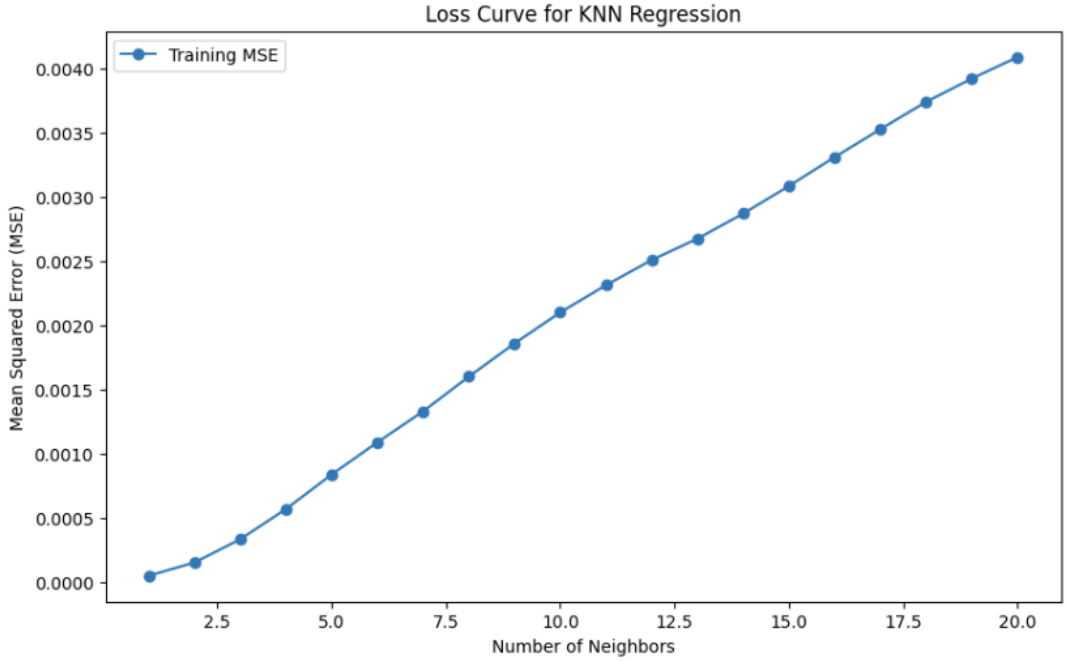
3. Models Results

a. Linear Regression Results:

- **Loss curve: N/A**
- **Accuracy: 99.3%**
- **Confusion matrix: N/A**
- **ROC curve: N/A**
- **Mean Squared error: 0.00022827852243677196**

b. KNN Results:

- **Loss curve:**



- **Accuracy: 97.59188589981756**
- **Confusion matrix: N/A**

- **ROC curve: N/A**
- **Mean Squared error: 0.00075274698553121 05**

II. IMAGE DATASET

1. Project Introduction

a. Dataset Name

Tomato Image Dataset

b. Number of classes and their labels

Number of classes: 2 (Reject, Unripe)

c. Dataset Images Numbers and size

- Total number of images in the dataset: 1600 (800 Reject, 800 Unripe)
- Size of each image: 128x128 pixels

d. Training, Validation and Testing

- Number of images used in training: 1120
- Number of images used in testing: 480

2. Implementation Details

a. Extracted Features

- HOG features were extracted, including their names and dimensions.

b. Cross-validation

Cross-validation is used with 5 folds for model evaluation.

c. Logistic Regression

℘ Hyper-parameters

- Initial learning rate: Not explicitly specified in the provided code.
- Optimizer: Liblinear (default for LogisticRegression)
- Regularization: L2 regularization (default for LogisticRegression)
- Batch size: Not applicable
- Number of epochs: Not applicable

d. K-Means

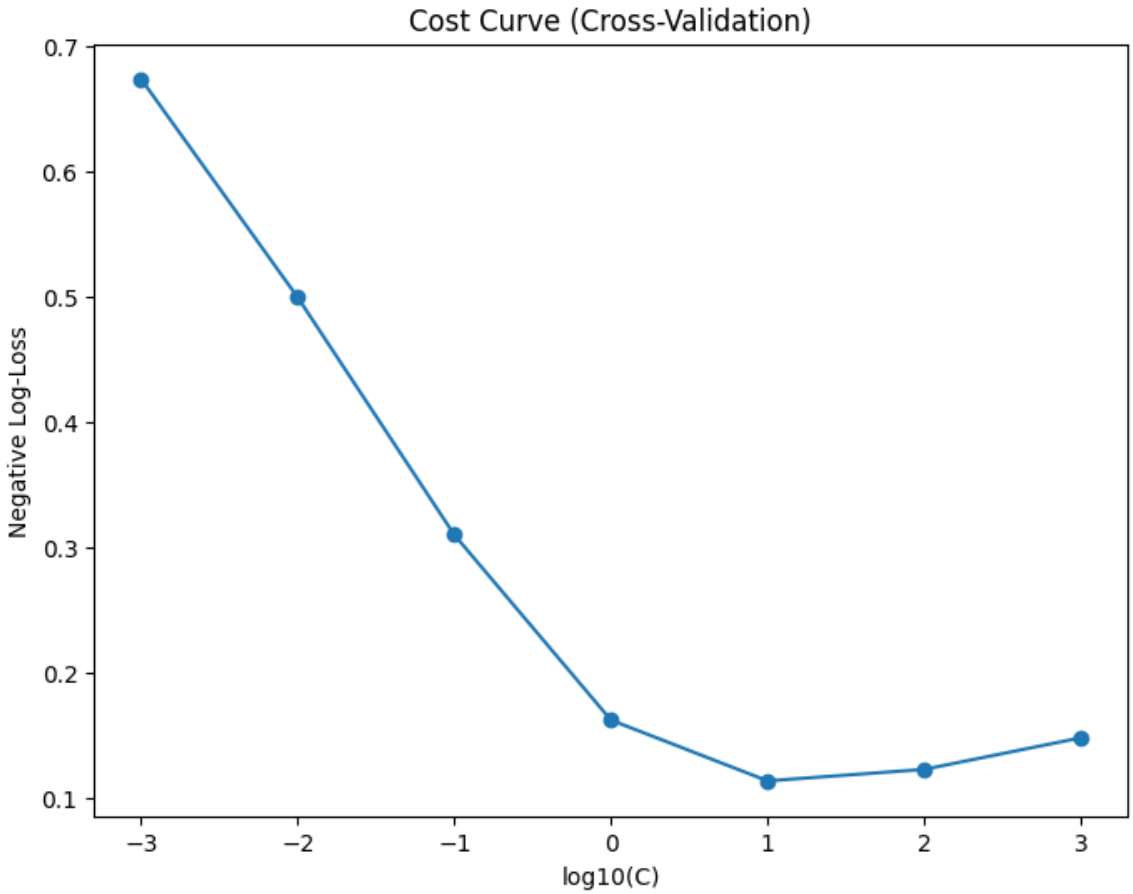
℘ Hyper-parameters

- Number of clusters (n_clusters): 2
- Initialization runs (n_init): 10
- Random state: 42

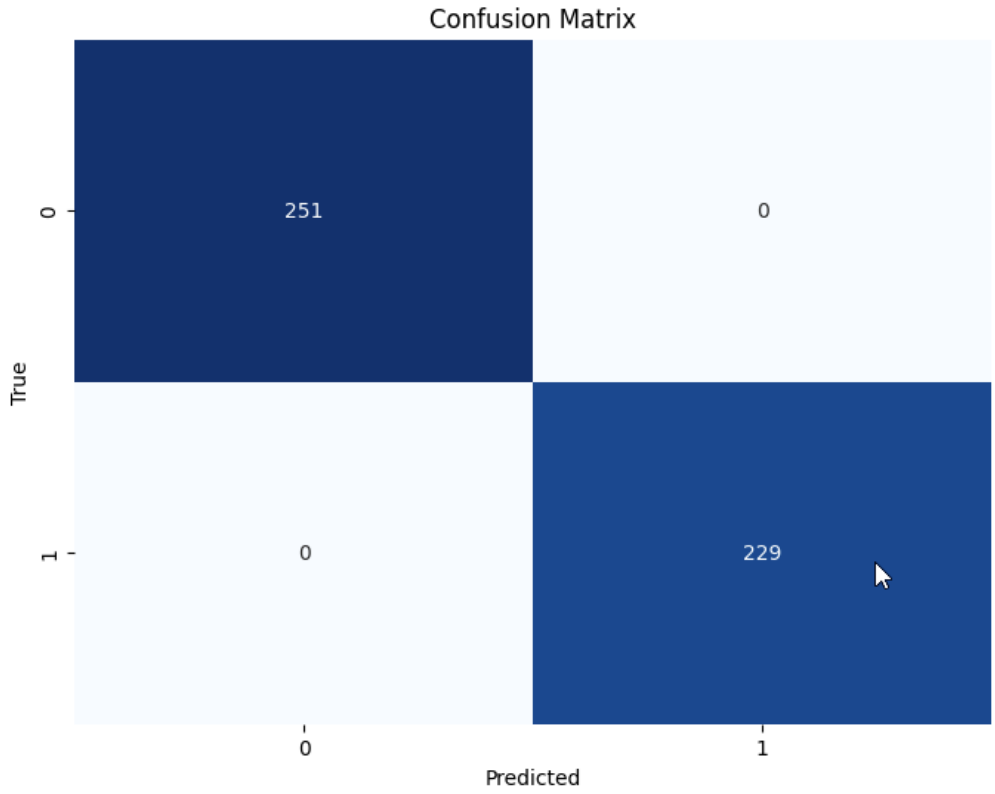
3. Models Results

a. Logistic Regression Results

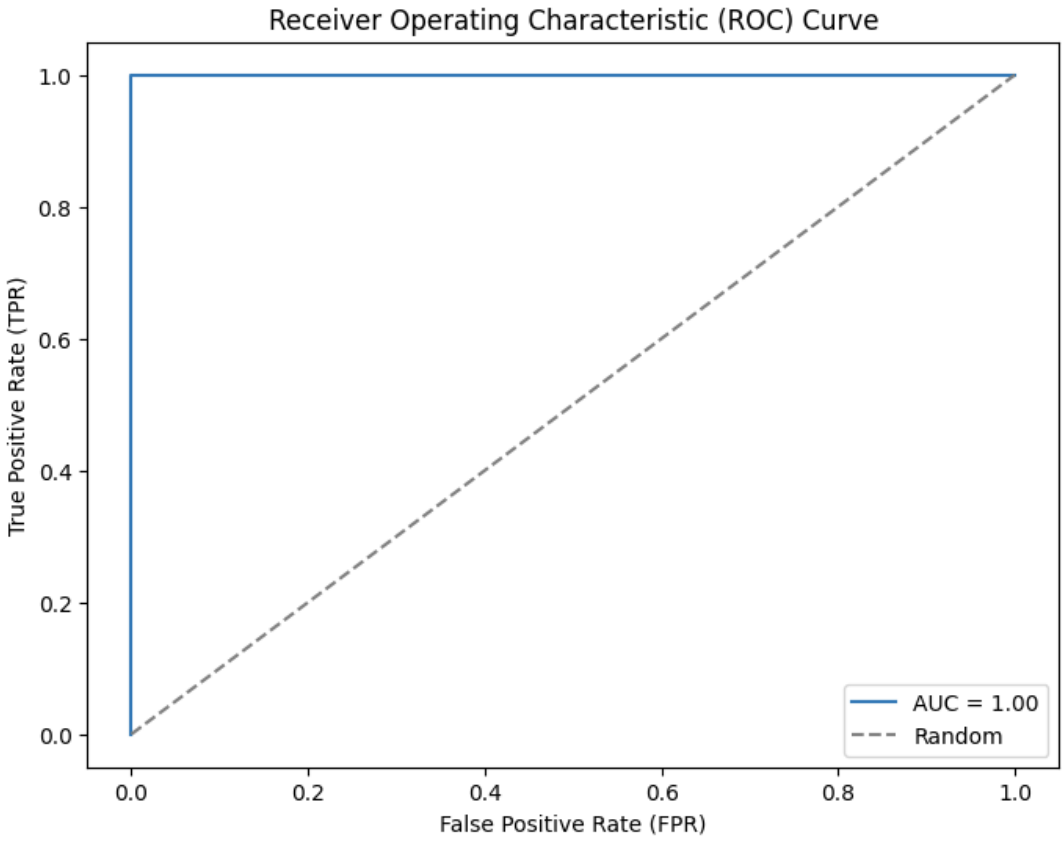
- **Loss curve:**



- **Accuracy: 97.92%**
- **Confusion matrix:**

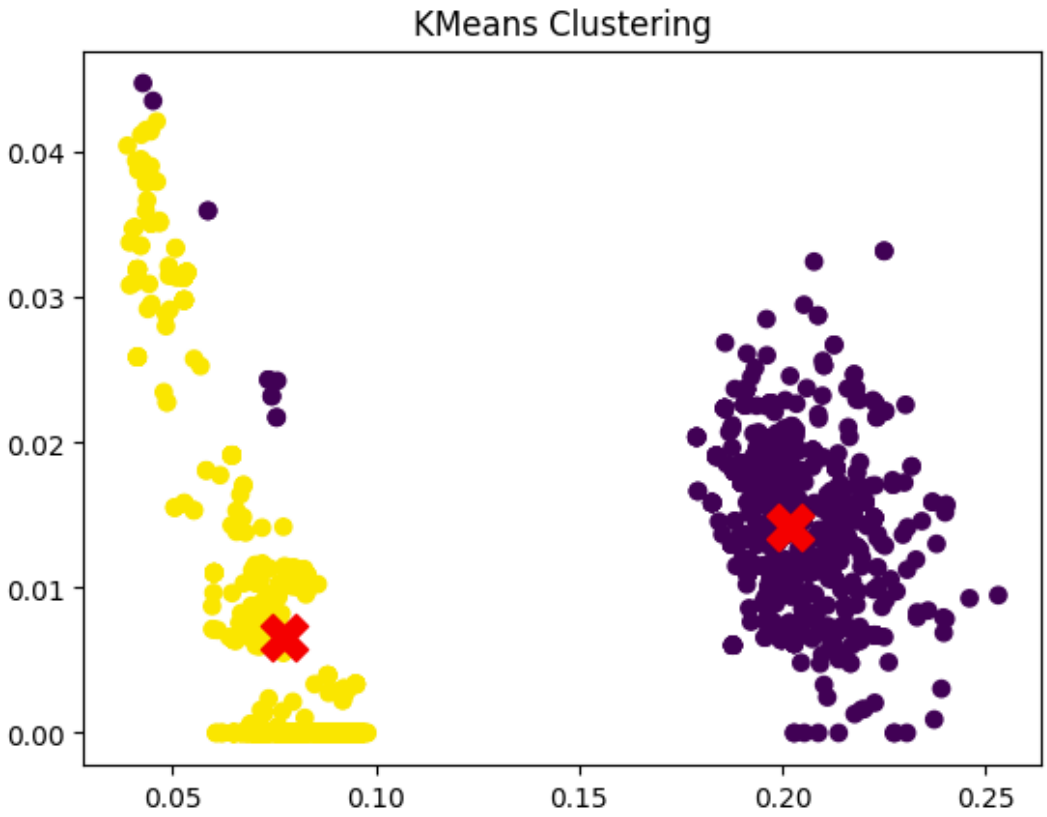

Confusion Matrix
251
0
0
229
True
0
1
0
1
Predicted

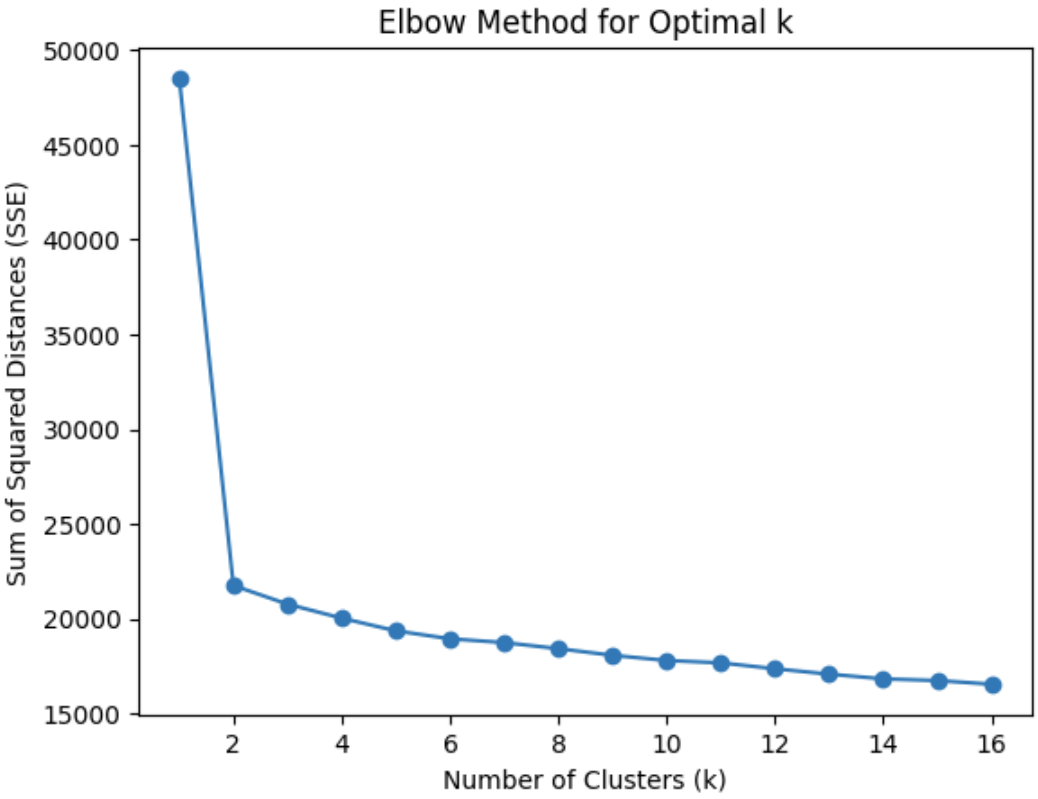
- **ROC curve:**



b.K-Means Results

- **Loss curve: N/A**
- **Confusion matrix: N/A**
- **ROC curve: N/A**
- **Silhouette Score: 0.470104236699423 97**


KMeans Clustering
0.04
0.03
0.02
0.01
0.00
0.05
0.10
0.15
0.20
0.25


Elbow Method for Optimal k
Sum of Squared Distances (SSE)
Number of Clusters (k)
50000
45000
40000
35000
30000
25000
20000
15000
2
4
6
8
10
12
14
16