Faculty of
computers and
artificial intelligence

Cover sheet

AI330 Machine Learning Project

Team no. :

Name	ID
زينة محمود محمد كمال	20210391
رودينا يحيي محمد	20210344
جواهر ابراهيم على حسن	20210258
ثناء خيري سيد محمد	20210247
سارة على مرسي راجح	20210398
شهاب محمد احمد إبراهيم عربان	20210462

Numerical dataset

General information about dataset

Name	House price prediction
Total no. of samples	4600
No. of samples in training\validation	4140
No. of samples in testing	460

Linear Regression model

Feature scaled phase:-

1) The number of scaling feature(before) : 18
2) Feature Names: ['date', 'price', 'bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors', 'waterfront', 'view', 'condition', 'sqft_above', 'sqft_basement', 'yr_built', 'yr_renovated', 'street', 'city', 'statezip', 'country']
3) The number of scaling feature(after) : 12

Cross validation:-

Is cross-validation used in any of implemented models?
No, the Linear Regression doesn't use cross-validation.

Hyper parameters:-

1) copy_X: True

2) fit_intercept: True

3) n_jobs: None

4) positive: False

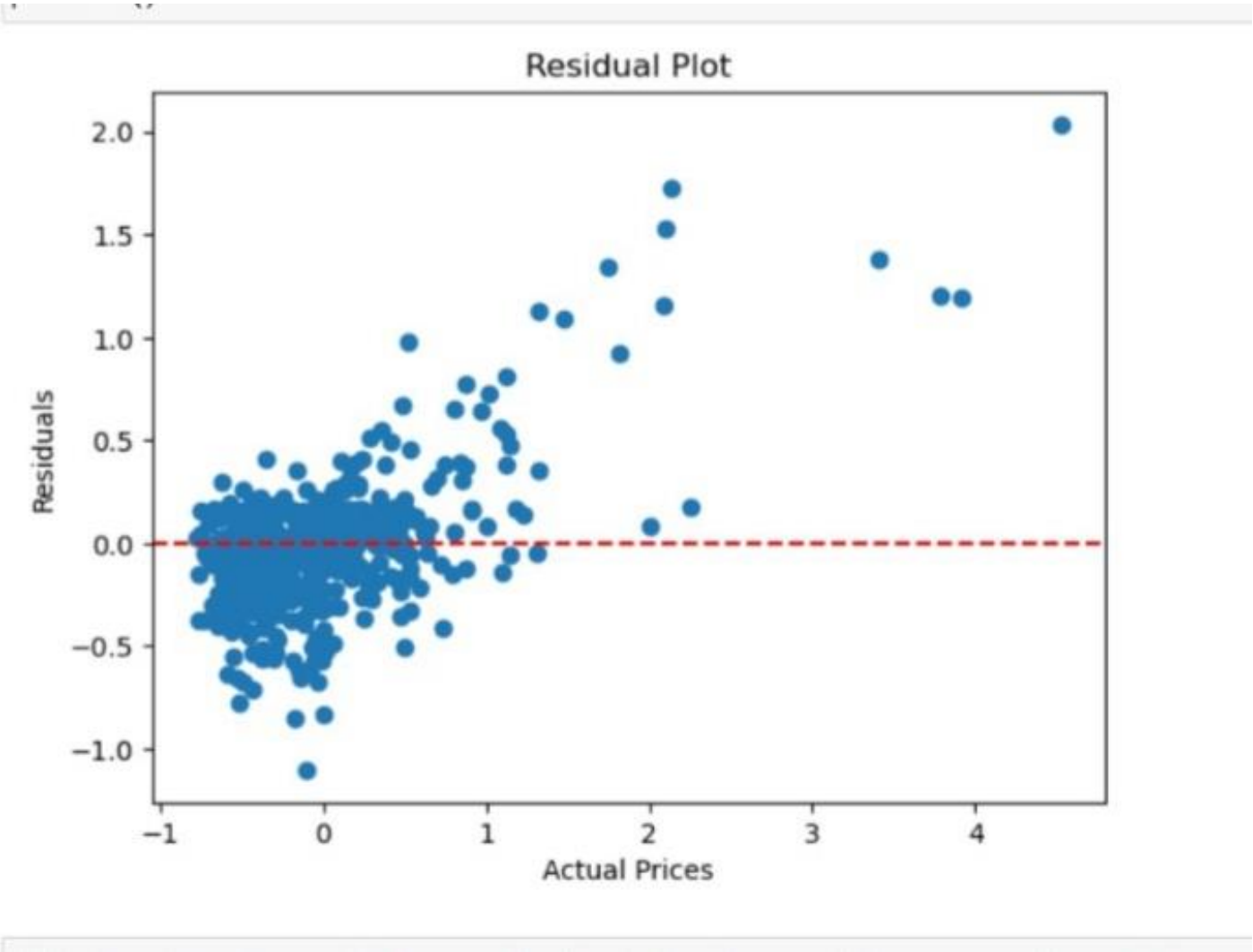
Result on testing data:-

The accuracy: 70%

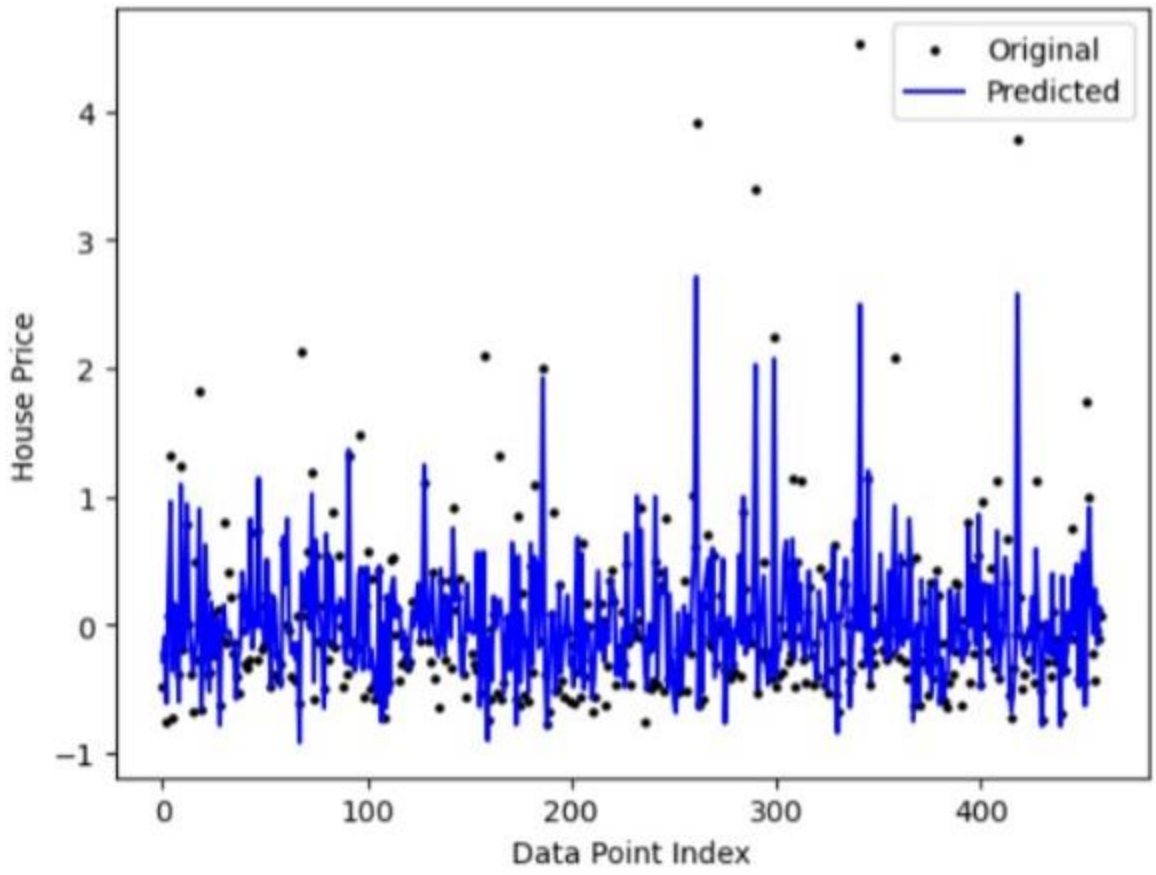
Visualization:-

1) Scatter plot

2) Residual plot



3) Line plot


Original
Predicted
House Price
Data Point Index
4
3
2
1
0
−1
0
100
200
300
400

KNN model

Feature scaled phase:-

3) **The number of scaling feature(before) : 18**
4) **Feature Names: ['date', 'price', 'bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors', 'waterfront', 'view', 'condition', 'sqft_above', 'sqft_basement', 'yr_built', 'yr_renovated', 'street', 'city', 'statezip', 'country']**
3) **The number of scaling feature(after) : 12**

Cross validation:-

Is cross-validation used in any of implemented models?
No, the KNN doesn't use cross-validation.

Hyper parameters:-

1) **n_neighbors: 9**

2) **weights: 'distance'**

3) `algorithm: 'auto'`

4) `n_jobs: None`

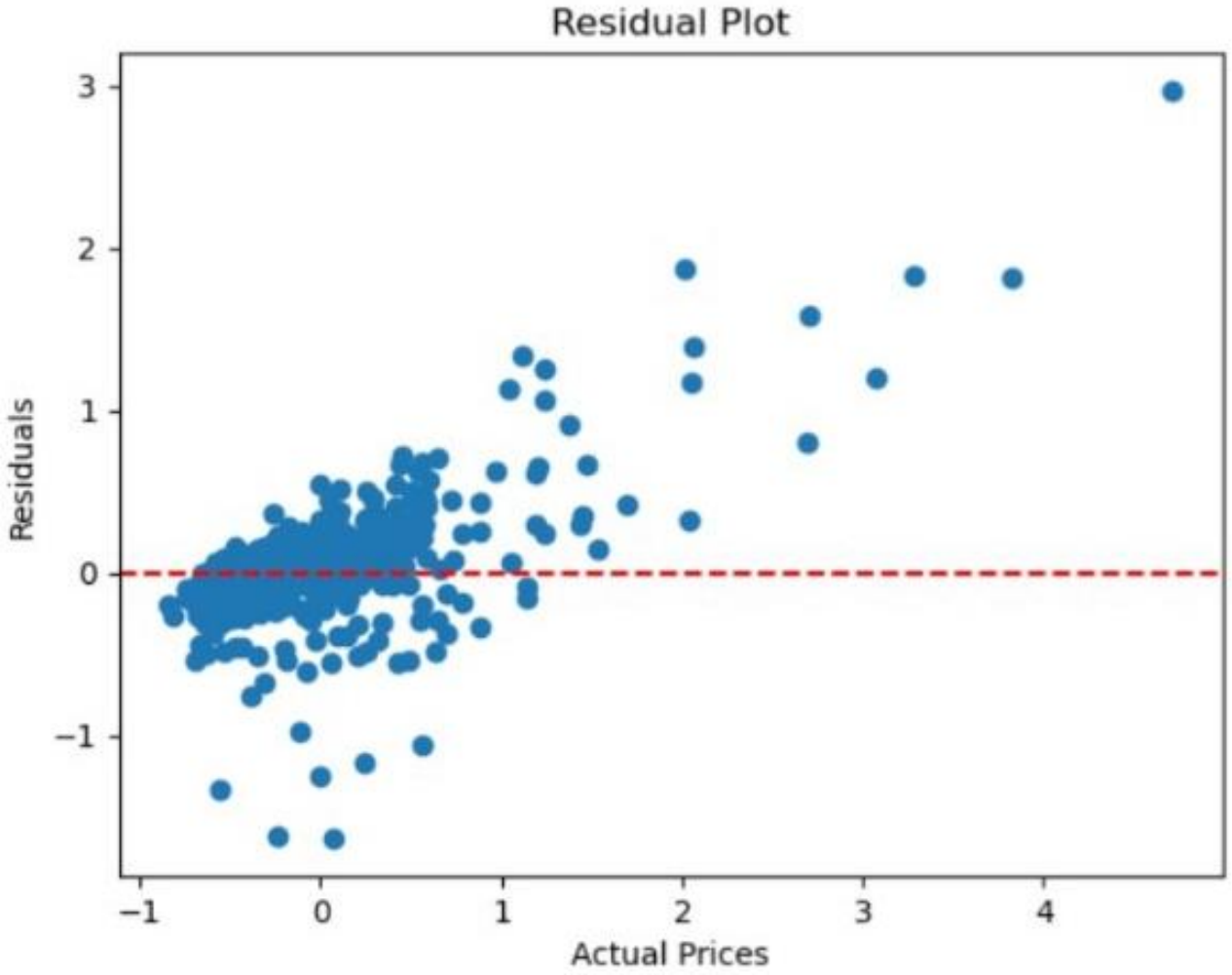
Result on testing data:-

```
The accuracy: 63%
```

Visualization:-

1) Scatter plot

2) Residual plot


Residual Plot
Residuals
Actual Prices
3
2
1
0
−1
−1
0
1
2
3
4

3) Line plot

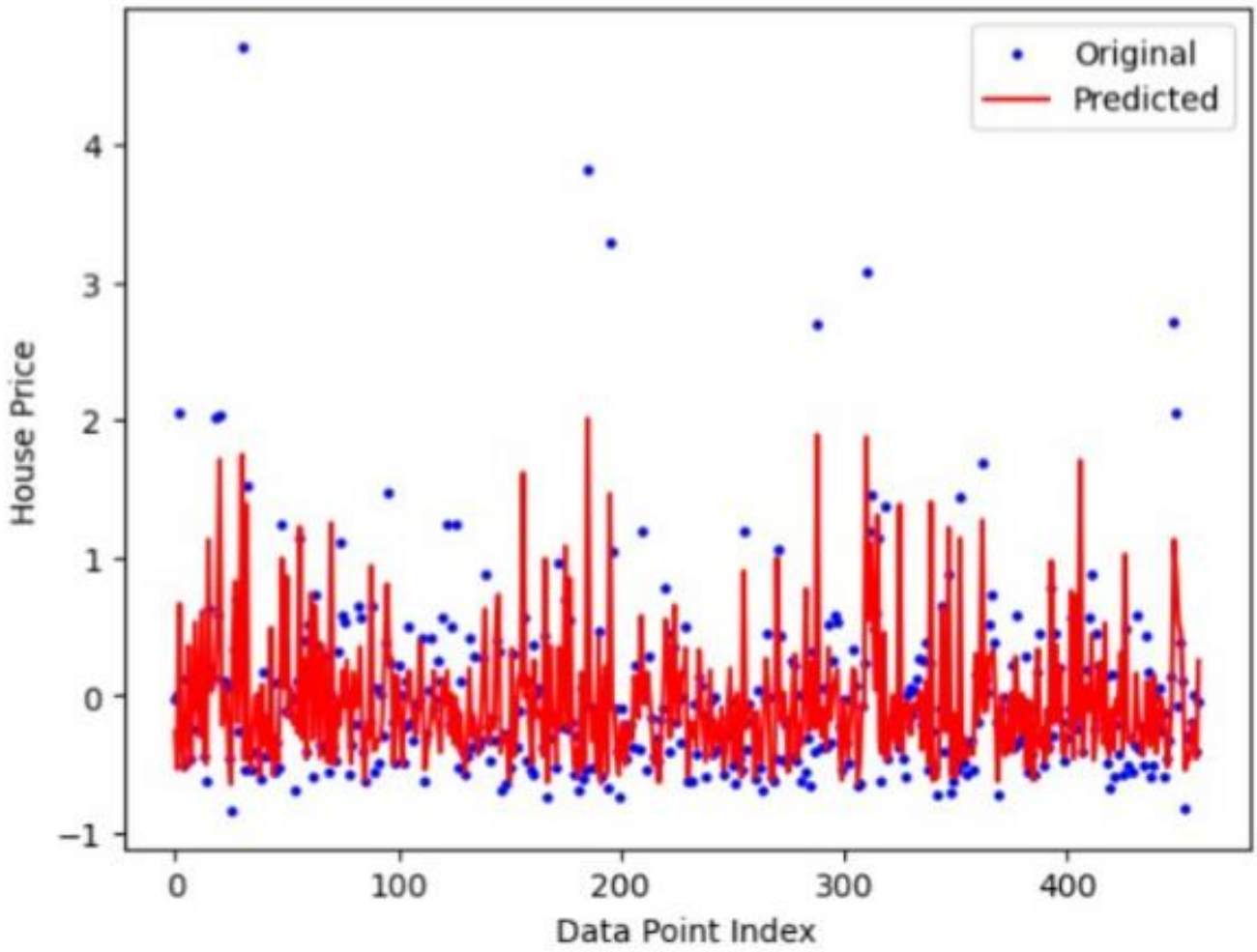

Original
Predicted
House Price
4
3
2
1
0
−1
0
100
200
300
400
Data Point Index

Image dataset

General information about dataset:

Name	STL-10 dataset
No. of classes	3
Total no. of samples	3229
Size of image	32x32 pixels
No. of samples in training\validation	2184
No. of samples in testing	1045

Feature Extraction Phase:

Number of HOG features extracted: 324

Names of HOG features: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323]

Dimension of resulted features: 324

Cross-Validation:

Is Cross-Validation Used? No, the code does not explicitly use cross-validation.

Hyperparameters of Logistic Regression Model:

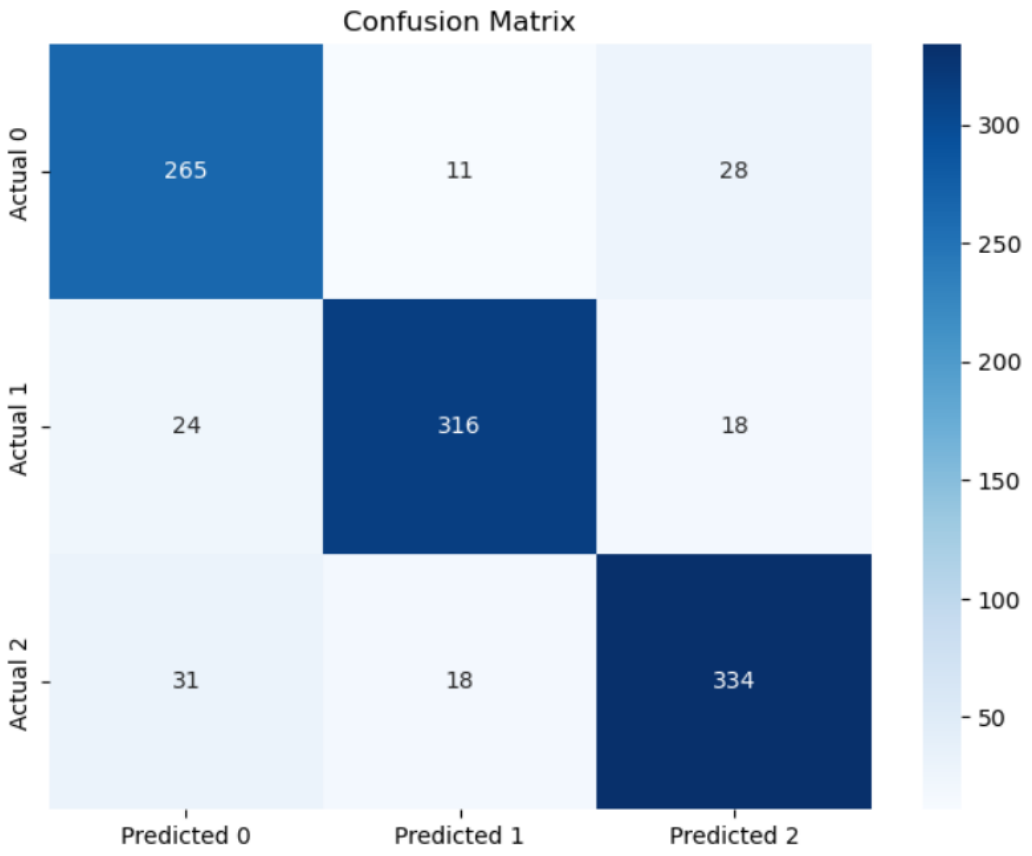
max_iter: 1000
penalty: l2

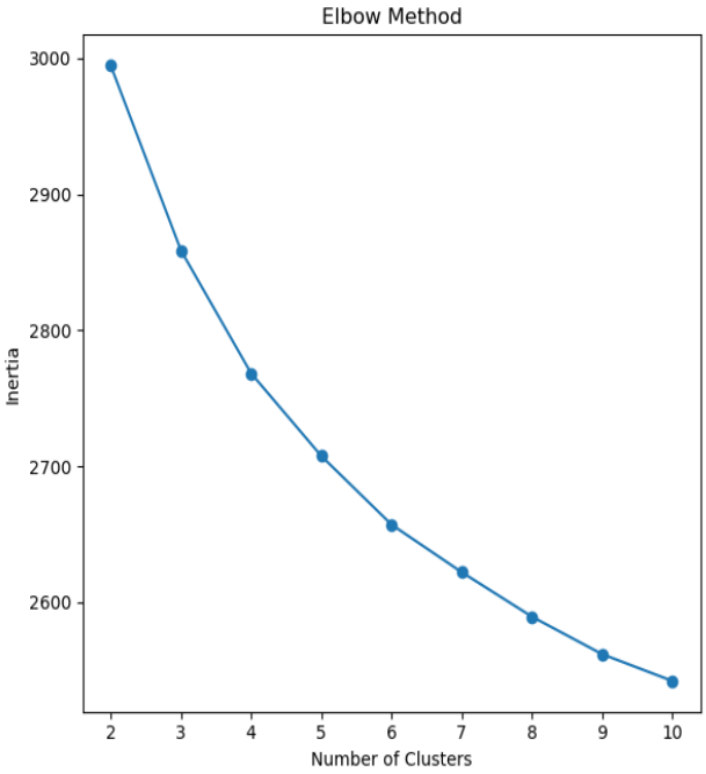
Results details:

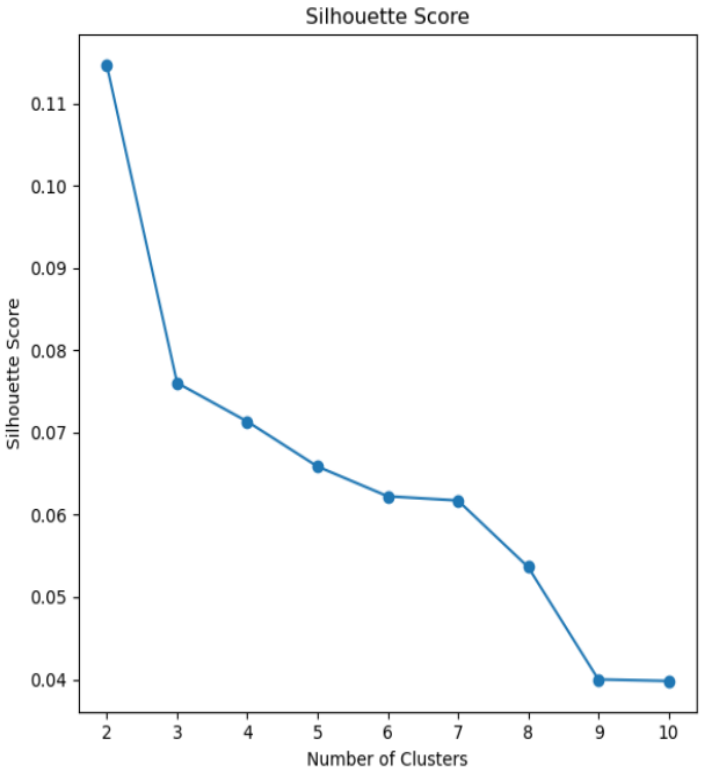
Accuracy of logistic regression with HOG features: 0.8755980861244019

Confusion Matrix:

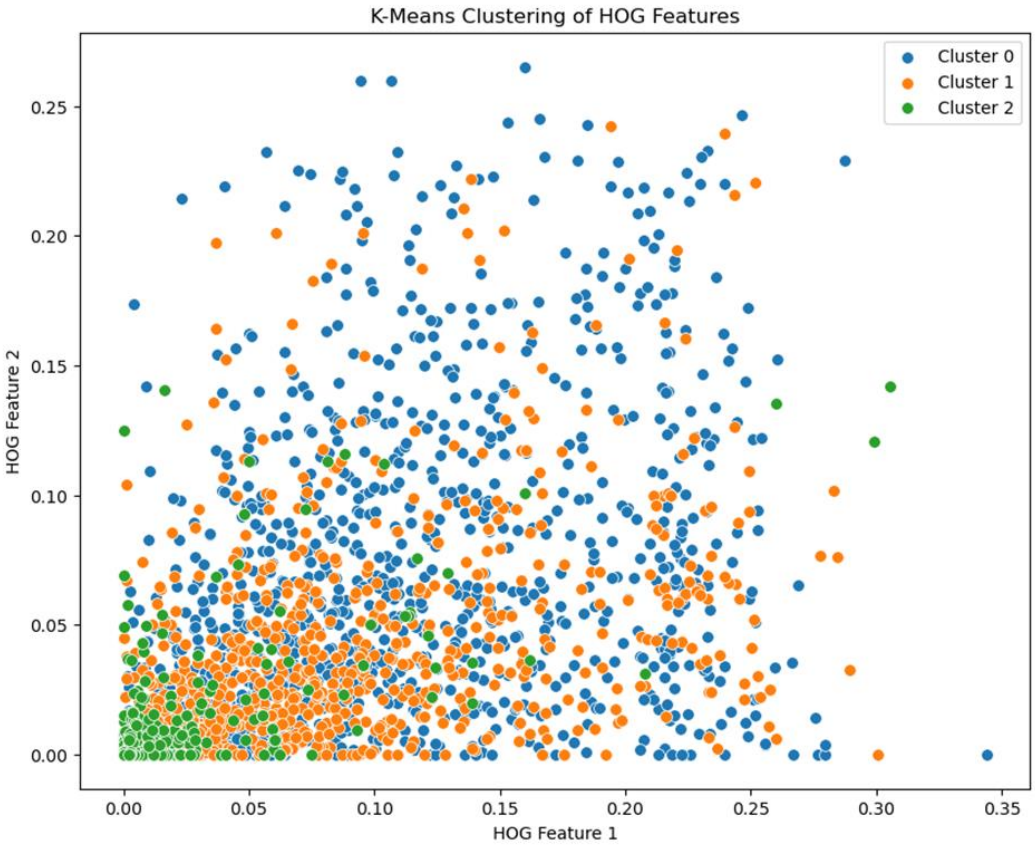
[[306 6]
 [14 298]]

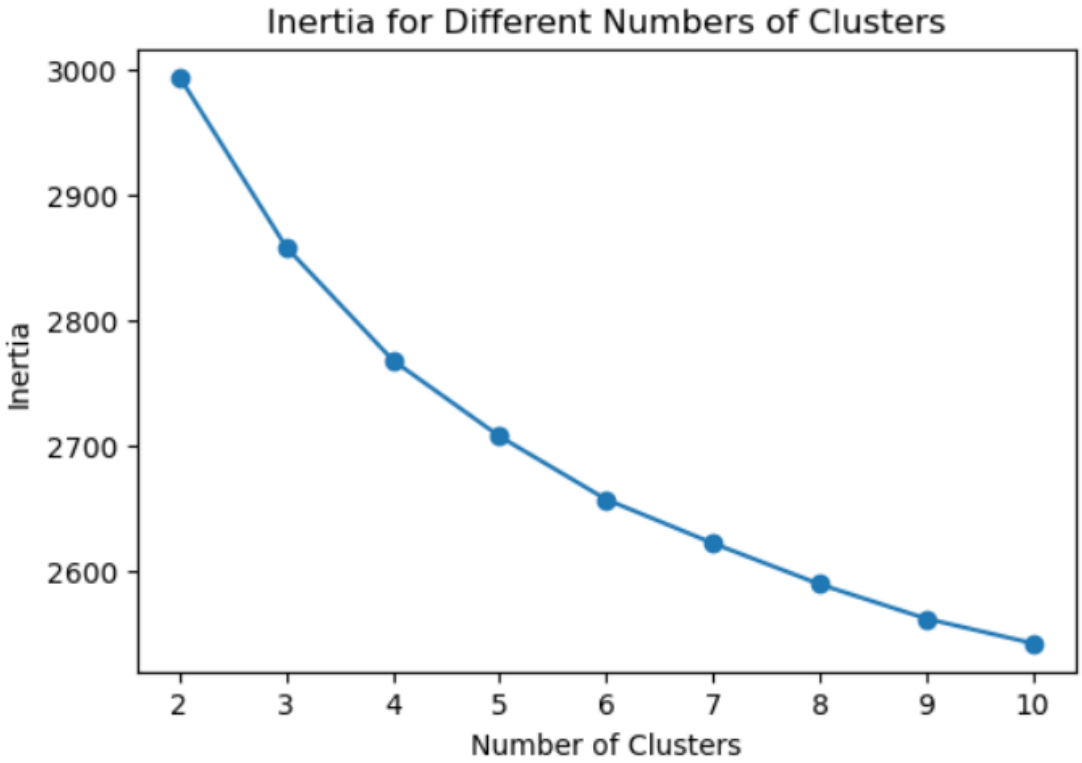


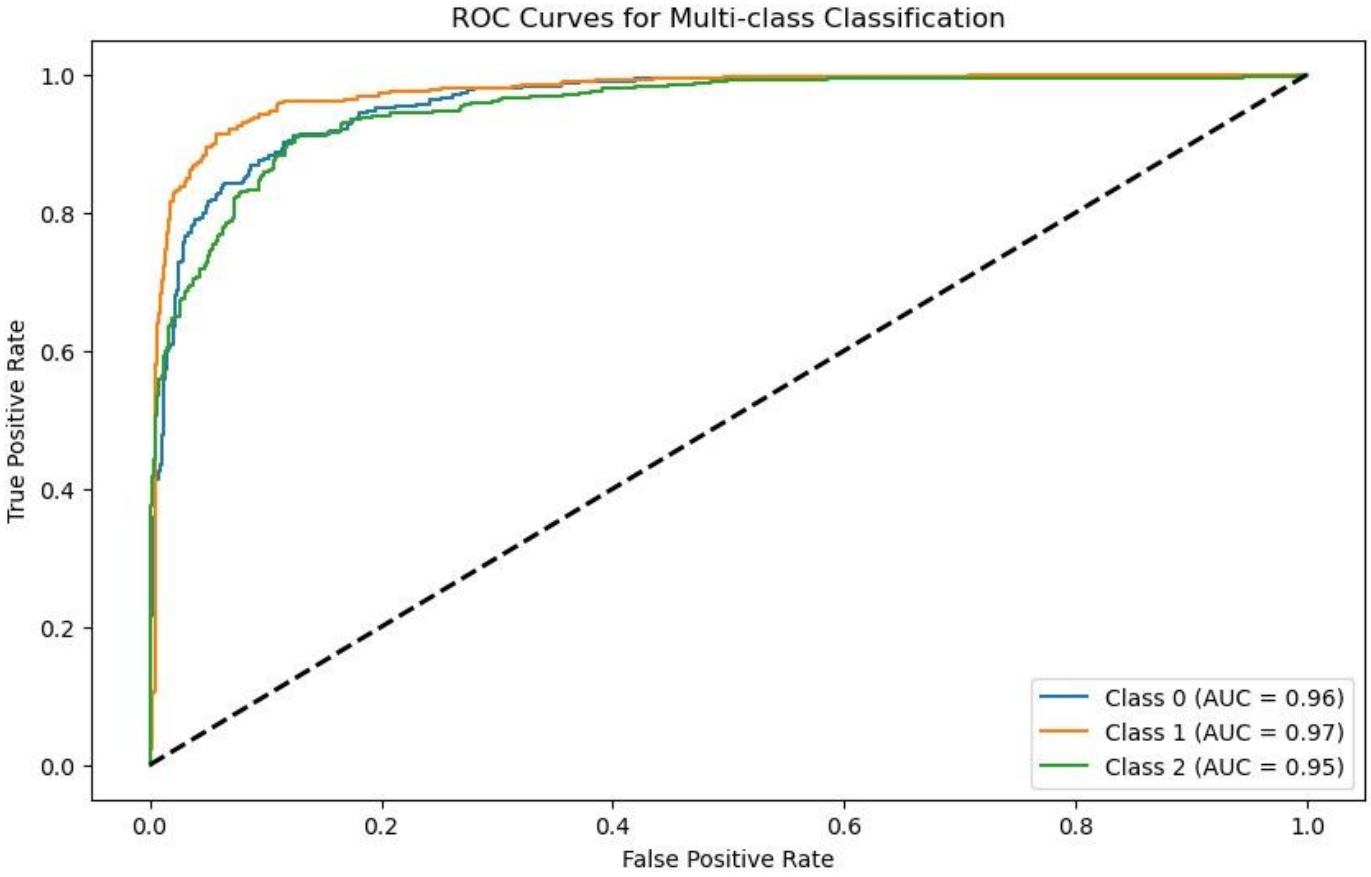




Silhouette Score: 0.0760431359206689


K-Means Clustering of HOG Features
Cluster 0
Cluster 1
Cluster 2
HOG Feature 2
HOG Feature 1
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35


Inertia for Different Numbers of Clusters
Inertia
Number of Clusters
3000
2900
2800
2700
2600
2
3
4
5
6
7
8
9
10


ROC Curves for Multi-class Classification
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Class 0 (AUC = 0.96)
Class 1 (AUC = 0.97)
Class 2 (AUC = 0.95)