Team Members

	Team Member ID	Team member name (**in Arabic**)	Department	level
1	20210251	جمال سيد احمد	cs	3
2	20210269	حازم عماد حسين	cs	3
3	20211101	يوسف محمد فتحي	cs	3
4	20210276	حبيبة شريف السيد	cs	3
5	20210307	داليا محمود السيد	cs	3
7	20211078	يوسف طارق ابو الفتوح	cs	3

Project Description Document

Numerical dataset

General Information on Dataset

- **Dataset Name:** **housing.csv**
- **Number of Classes:** NONE
- **Labels:** NONE
- **Total Number of Samples:** **20640**
- **Sample Size (if applicable):** [numerical]
- **Samples Used:**
  - Training: 16512
  - Validation: [none]
  - Testing: 4128

Implementation Details

- **Feature Extraction Phase:**
  - **Number of Features Extracted:** 13

Feature Names: **['longitude', 'latitude', 'housing_median_age', 'total_rooms', 'total_bedrooms', 'population', 'households', 'median_income', 'ocean_proximity_<1H OCEAN', 'ocean_proximity_INLAND', 'ocean_proximity_ISLAND', 'ocean_proximity_NEAR BAY', 'ocean_proximity_NEAR OCEAN']**

-

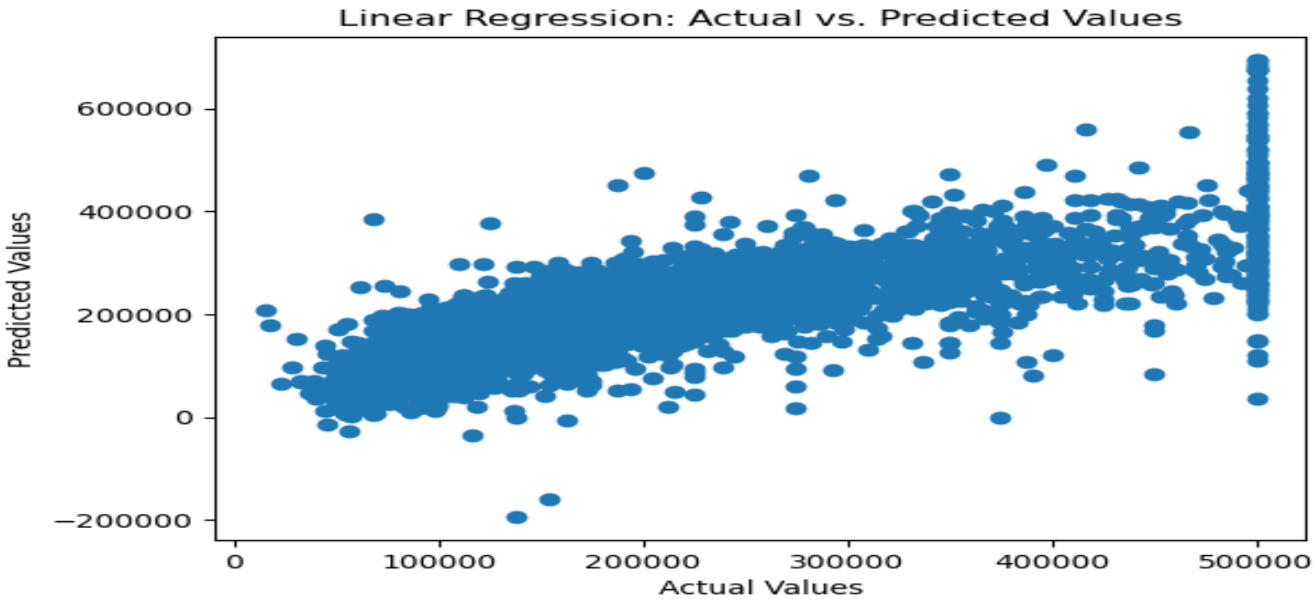
Dimension of Resulted Features: (20640, 13)

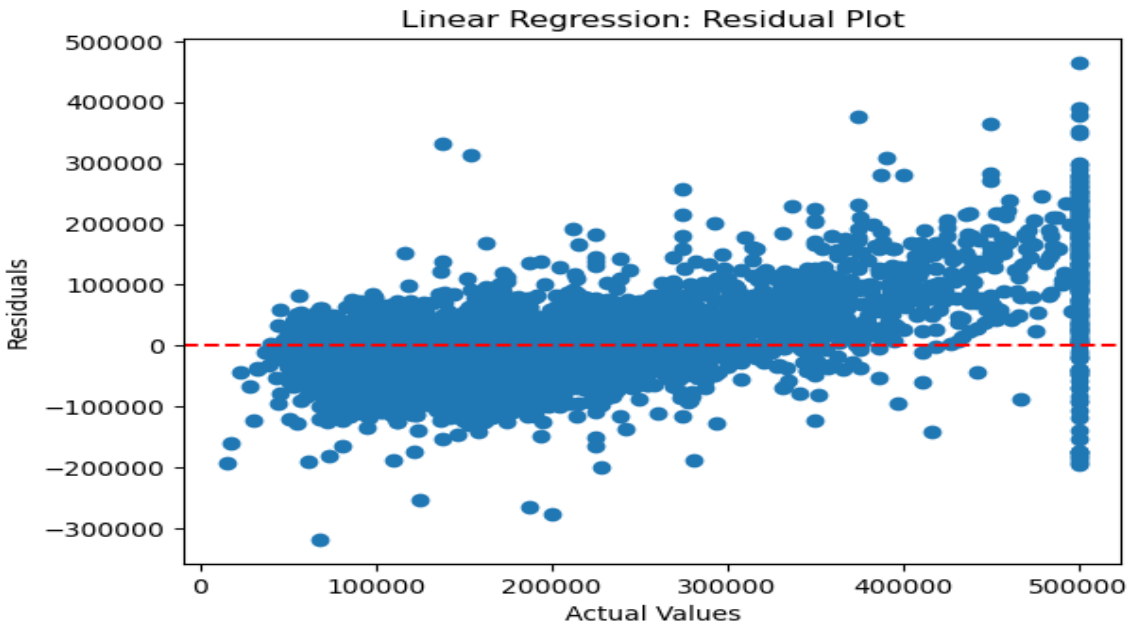
- **Cross-Validation:**
  - **Used?:** [**Yes**/No]
  - **Number of Folds:** [**5**]
  - **Training/Validation Ratio: 4:1**
- **Hyperparameters Used:**
  - **Initial Learning Rate:** [Rate]
  - **Optimizer:** [Optimizer]
  - **Regularization:** [Type and strength]
  - **Batch Size:** [Size]
  - **Number of Epochs:** [Number]

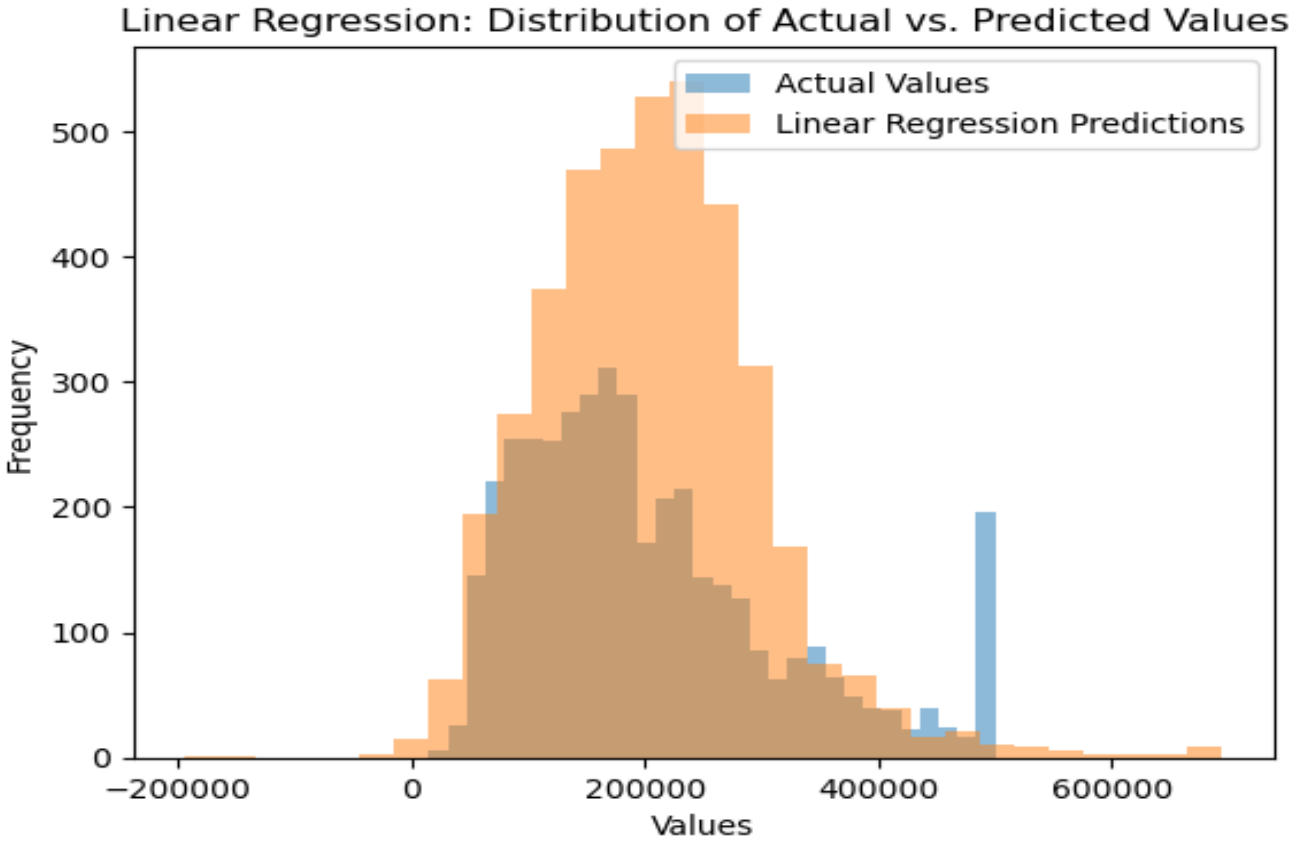
Results Details

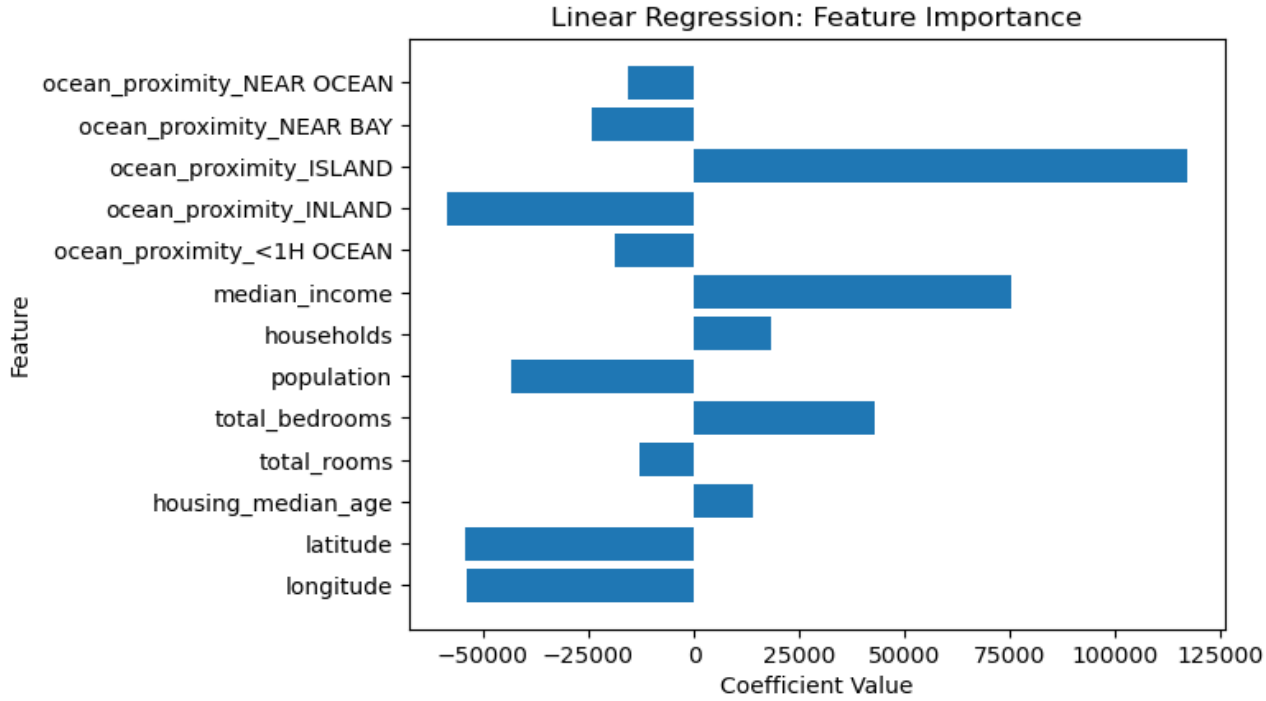
- ## Performance on Testing Data:

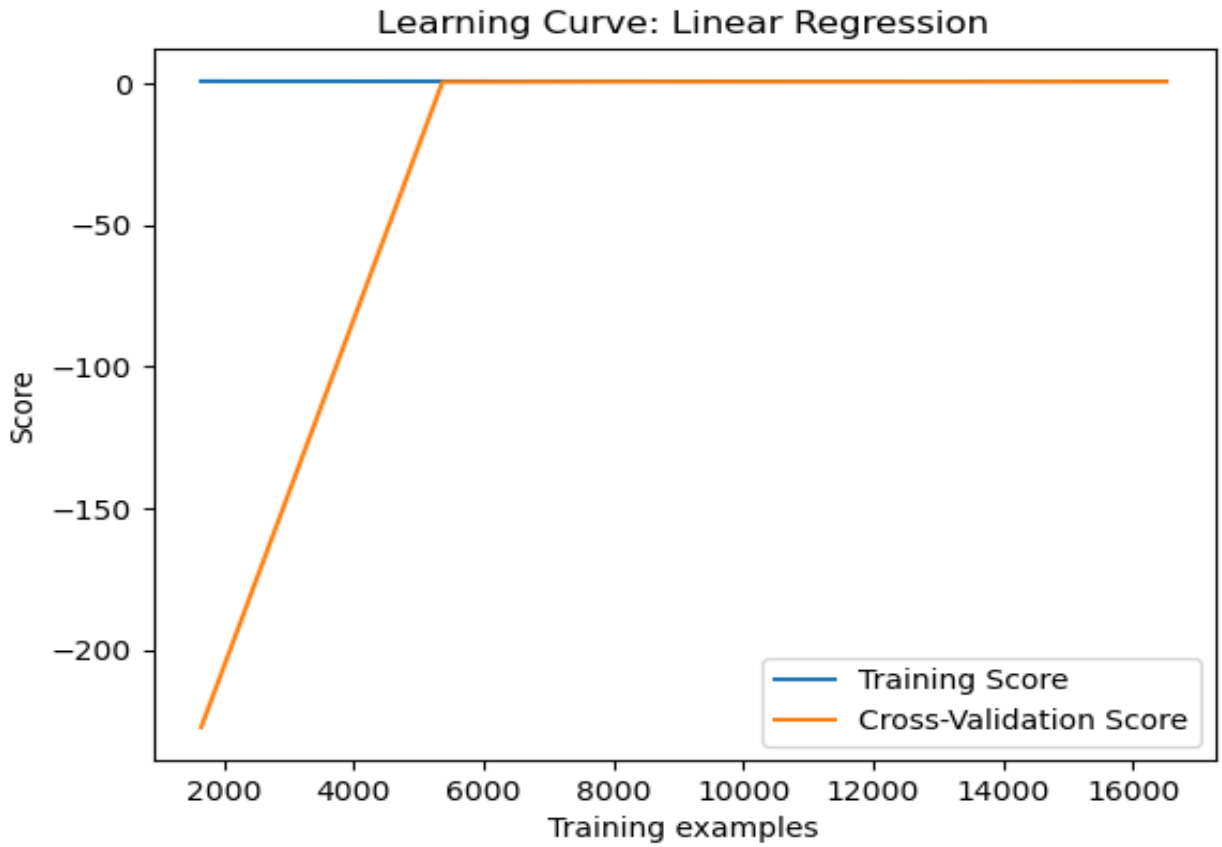
- **Linear Regression Score: 62.54**
- **Linear Regression MSE on Test Data: 4908476721.156613**
- **Linear Regression R-squared on Test Data: 0.6254240620553608**

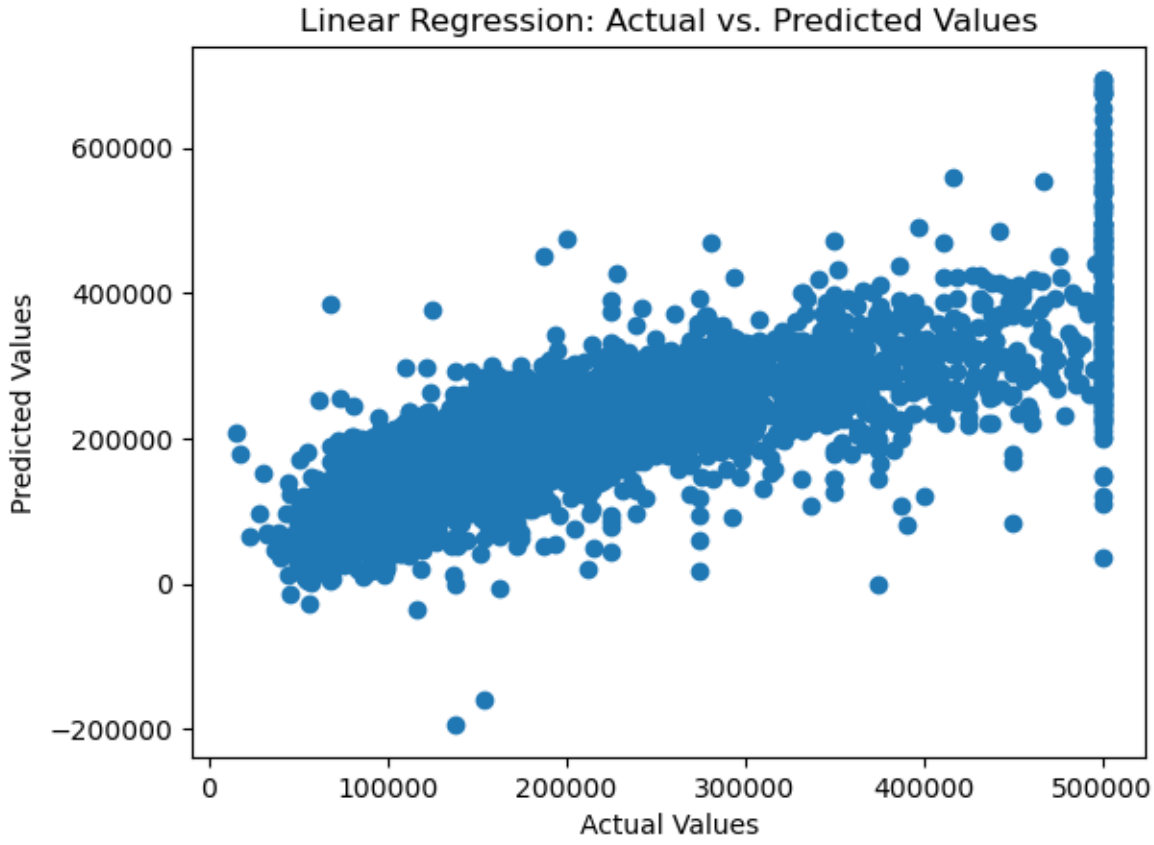


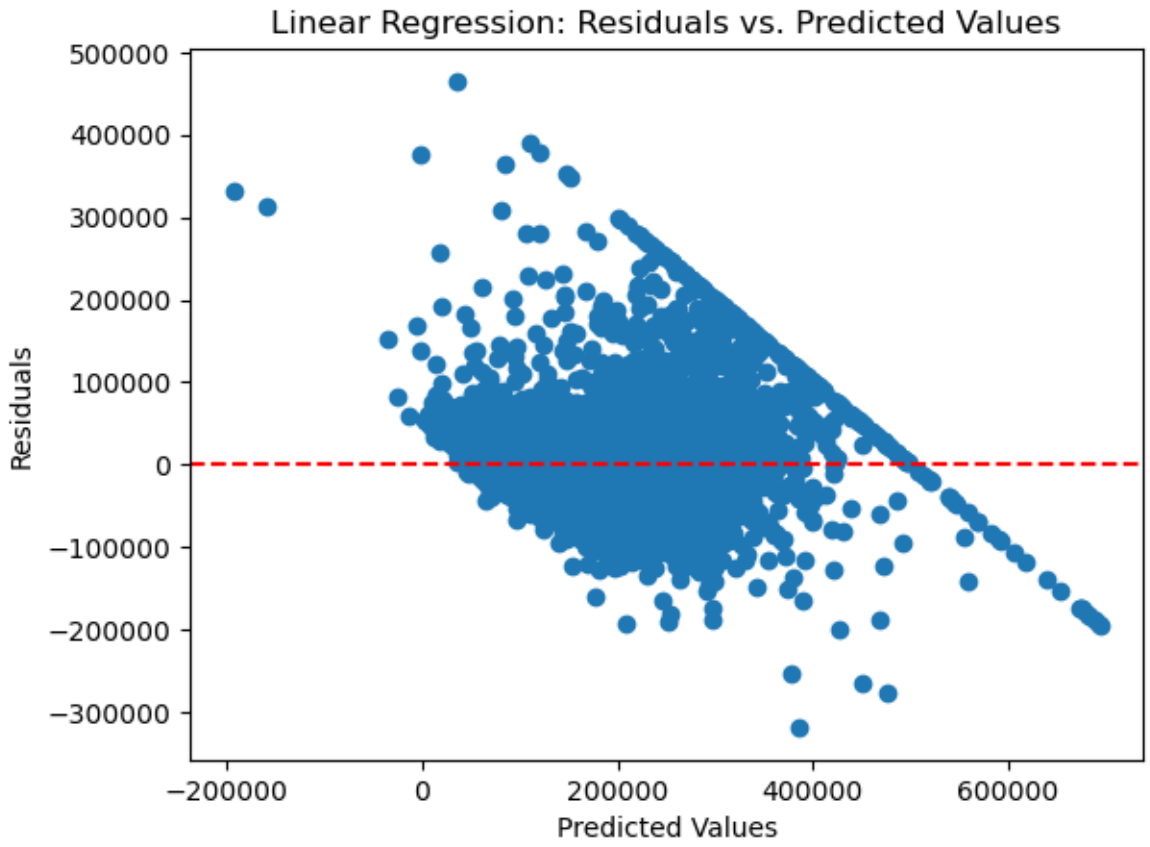

Linear Regression: Residual Plot
Residuals
Actual Values
500000
400000
300000
200000
100000
0
−100000
−200000
−300000
0
100000
200000
300000
400000
500000


Linear Regression: Distribution of Actual vs. Predicted Values
Actual Values
Linear Regression Predictions
Frequency
Values
500
400
300
200
100
0
−200000
0
200000
400000
600000


Linear Regression: Feature Importance
ocean_proximity_NEAR OCEAN
ocean_proximity_NEAR BAY
ocean_proximity_ISLAND
ocean_proximity_INLAND
ocean_proximity_<1H OCEAN
median_income
households
population
total_bedrooms
total_rooms
housing_median_age
latitude
longitude
Feature
−50000 −25000 0 25000 50000 75000 100000 125000
Coefficient Value


Learning Curve: Linear Regression
0
−50
−100
−150
−200
Score
2000 4000 6000 8000 10000 12000 14000 16000
Training examples
Training Score
Cross-Validation Score


Linear Regression: Actual vs. Predicted Values
Predicted Values
600000
400000
200000
0
−200000
0
100000
200000
300000
400000
500000
Actual Values

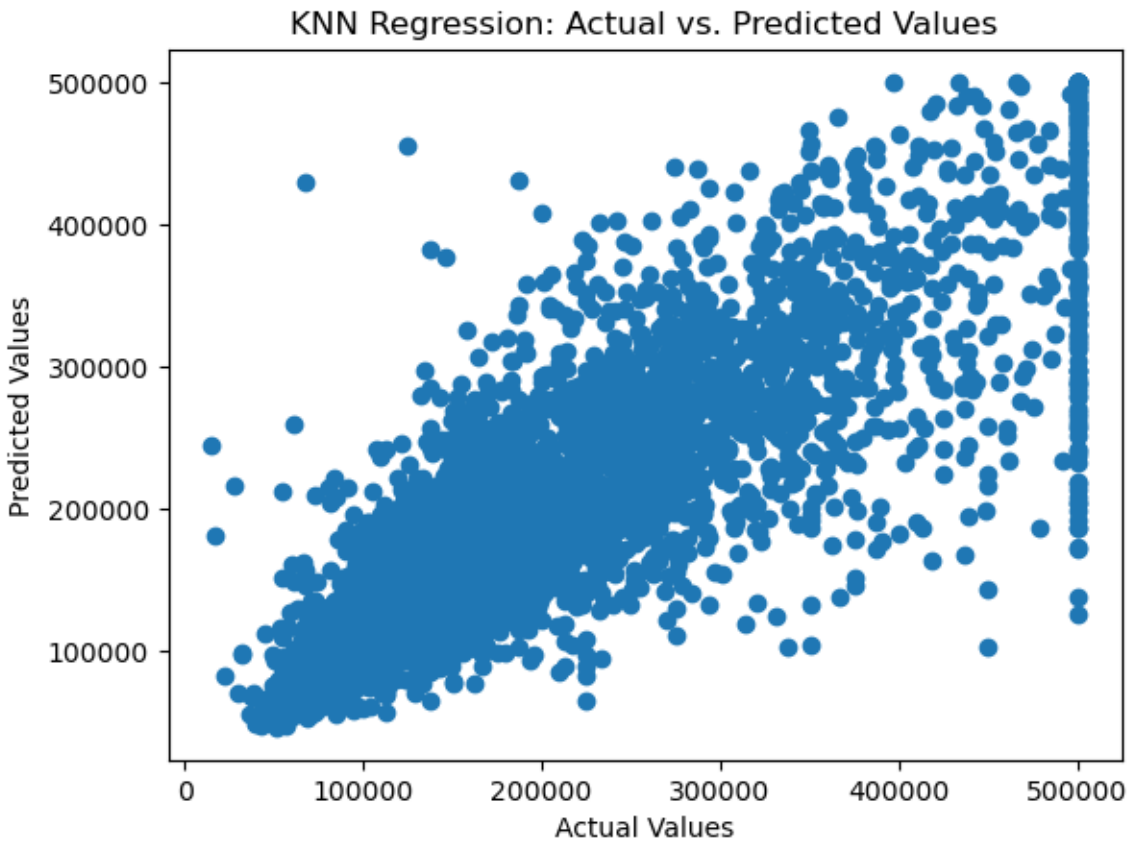

Linear Regression: Residuals vs. Predicted Values
Residuals
Predicted Values
500000
400000
300000
200000
100000
0
−100000
−200000
−300000
−200000
0
200000
400000
600000

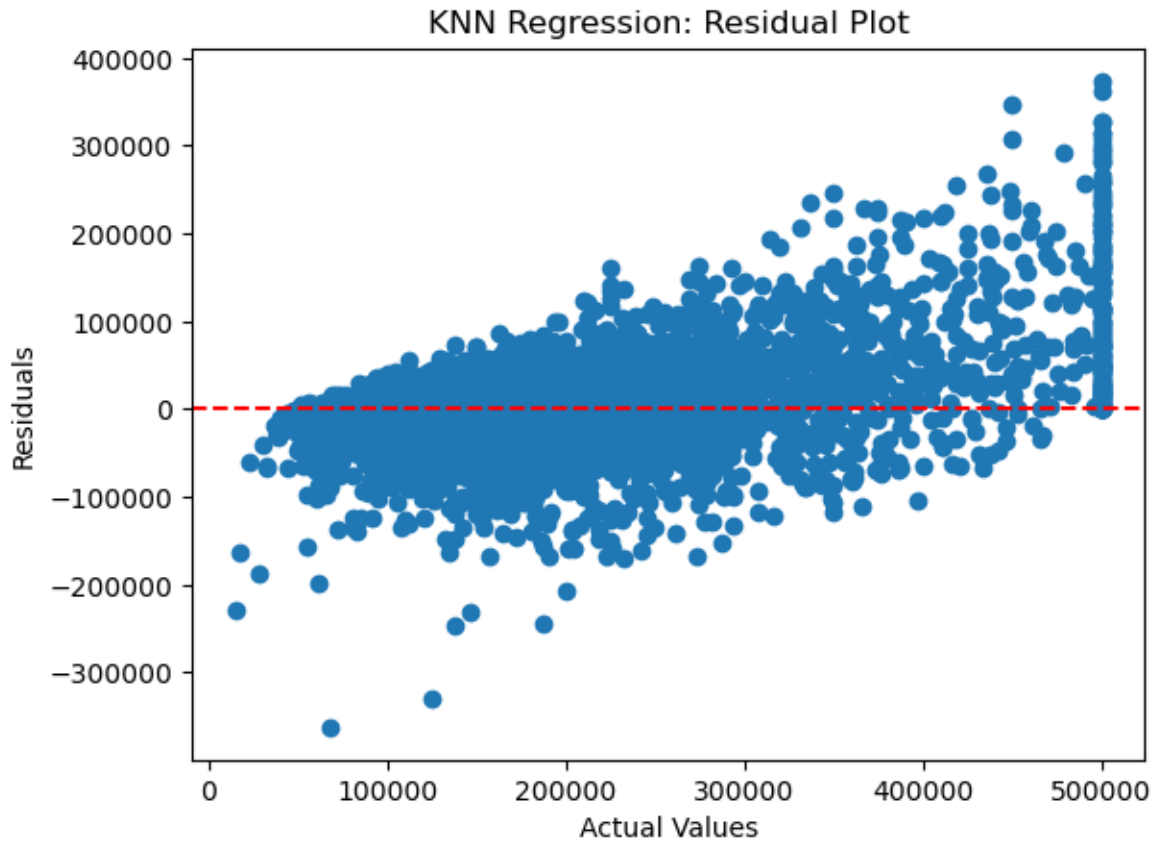
KNN

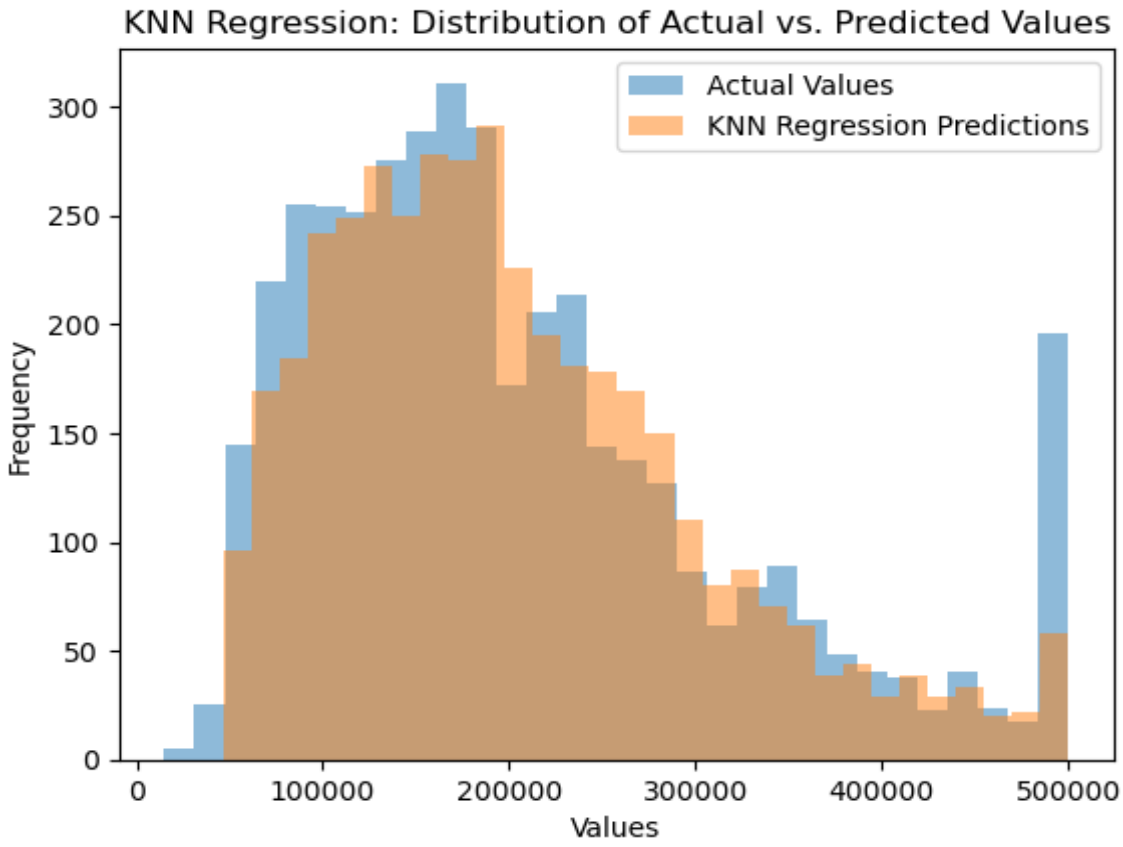
- **Hyperparameters Used: K=5**

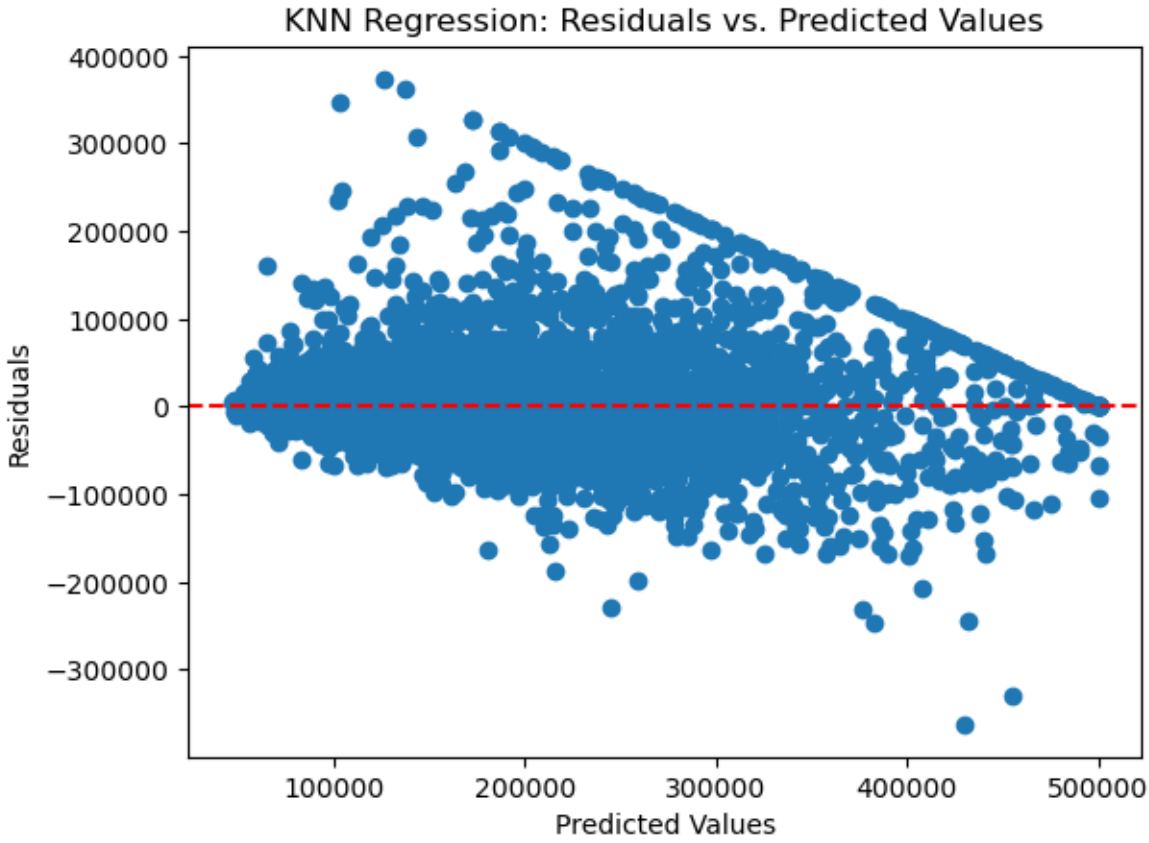
Results Details

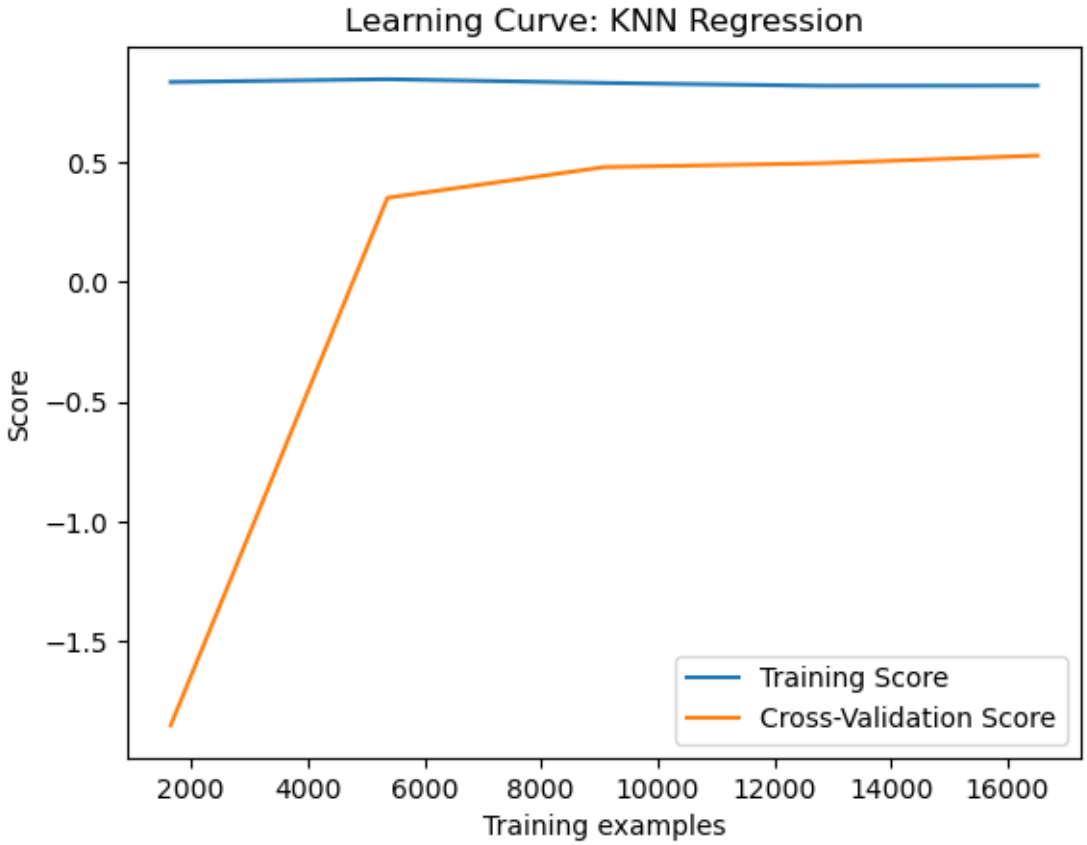
- **Performance on Testing Data:**
- **KNN Regression Score: 71.3**
- **KNN Regression MSE on Test Data: 3760982284.460552**
- **KNN Regression R-squared on Test Data: 0.7129917188518262**

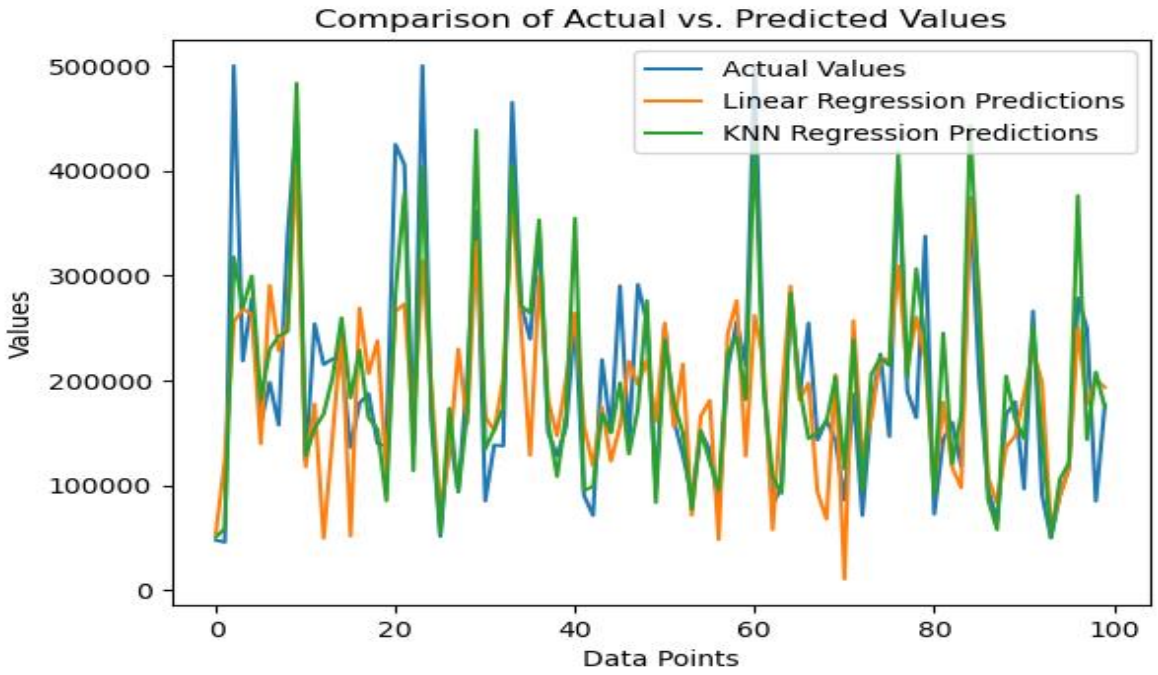


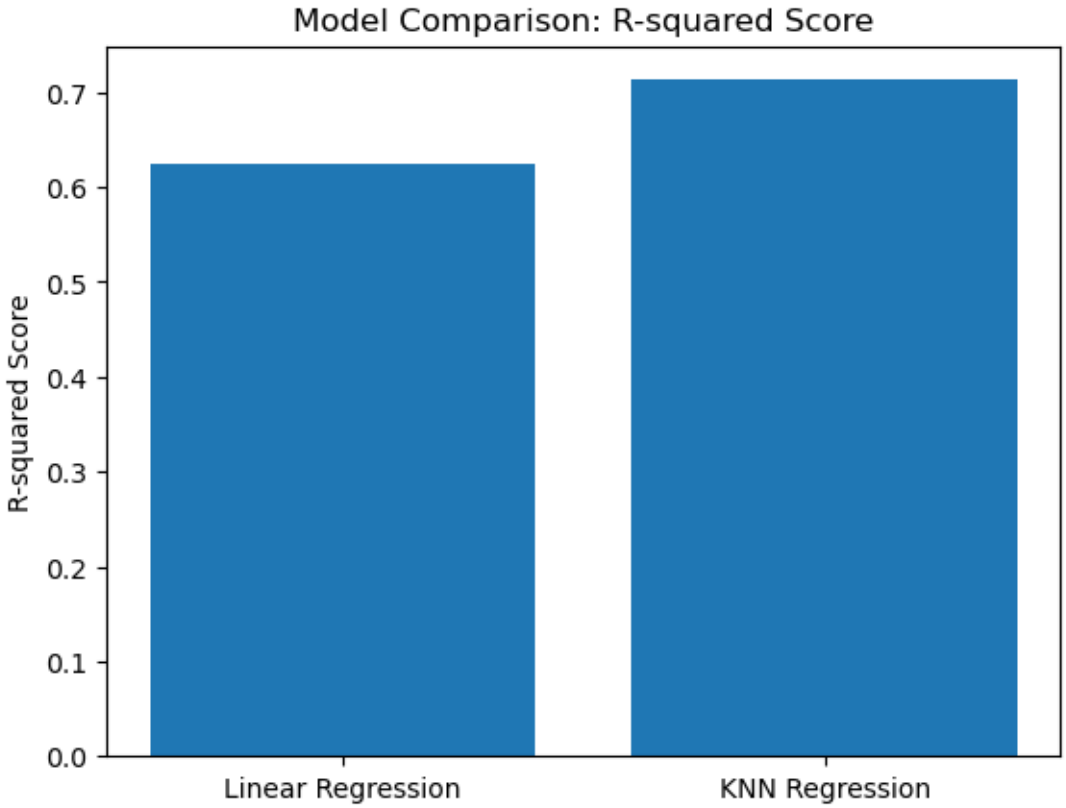

KNN Regression: Residual Plot
Residuals
Actual Values
400000
300000
200000
100000
0
−100000
−200000
−300000
0
100000
200000
300000
400000
500000

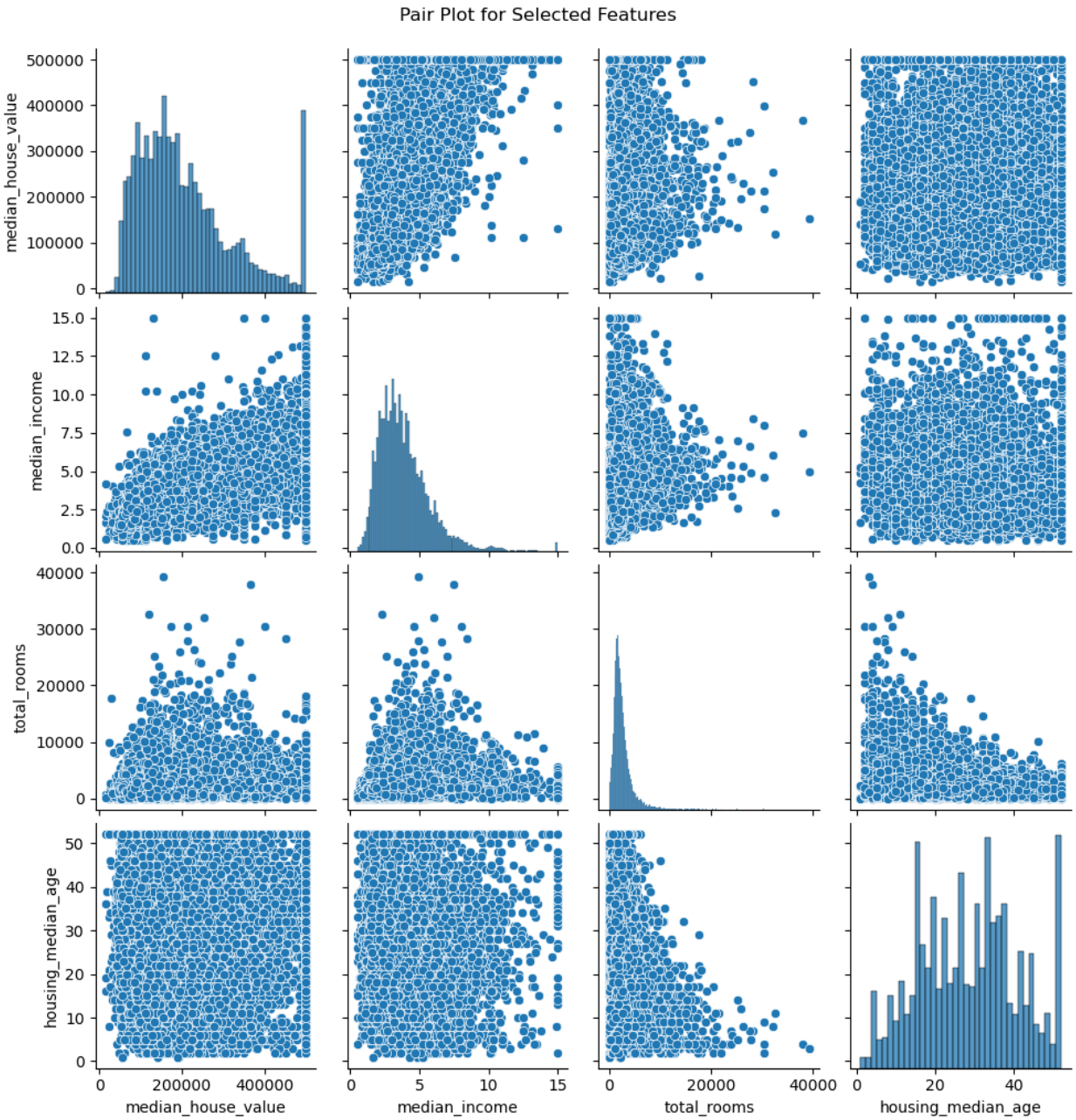

KNN Regression: Distribution of Actual vs. Predicted Values
Actual Values
KNN Regression Predictions
Frequency
Values
0
50
100
150
200
250
300
0
100000
200000
300000
400000
500000


KNN Regression: Residuals vs. Predicted Values
Residuals
Predicted Values
400000
300000
200000
100000
0
−100000
−200000
−300000
100000
200000
300000
400000
500000


Learning Curve: KNN Regression
Score
Training examples
0.5
0.0
−0.5
−1.0
−1.5
2000
4000
6000
8000
10000
12000
14000
16000
Training Score
Cross-Validation Score


Comparison of Actual vs. Predicted Values
Actual Values
Linear Regression Predictions
KNN Regression Predictions
Values
Data Points
0
100000
200000
300000
400000
500000
0
20
40
60
80
100


Model Comparison: R-squared Score
R-squared Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Linear Regression
KNN Regression


Pair Plot for Selected Features
median_house_value
median_income
total_rooms
housing_median_age

Logistic Regression Model

Image dataset

General Information on Dataset

- Dataset Name: dataset.csv
- Number of Classes: 3
- Labels: [0 2 3]
- Total Number of Samples: 35887
- Sample Size (if applicable): (48,48)

Samples Used:

Training: 13092

Validation: 3756

Testing: 1898

Samples Used per Class:

Training:

emotion count_train

0 3995

2	4097
3	5000

Validation:

emotion	count _val
0	958
2	1024
3	1774

Testing:

emotion	count_ test
0	491
2	528
3	879

Implementation Details

- **Feature Extraction Phase:**
- **Number of Features Extracted:**

Number of features extracted per image: 800

- **Feature Names:** **HoG**
- **Dimension of Resulted Features:** **(800,)**

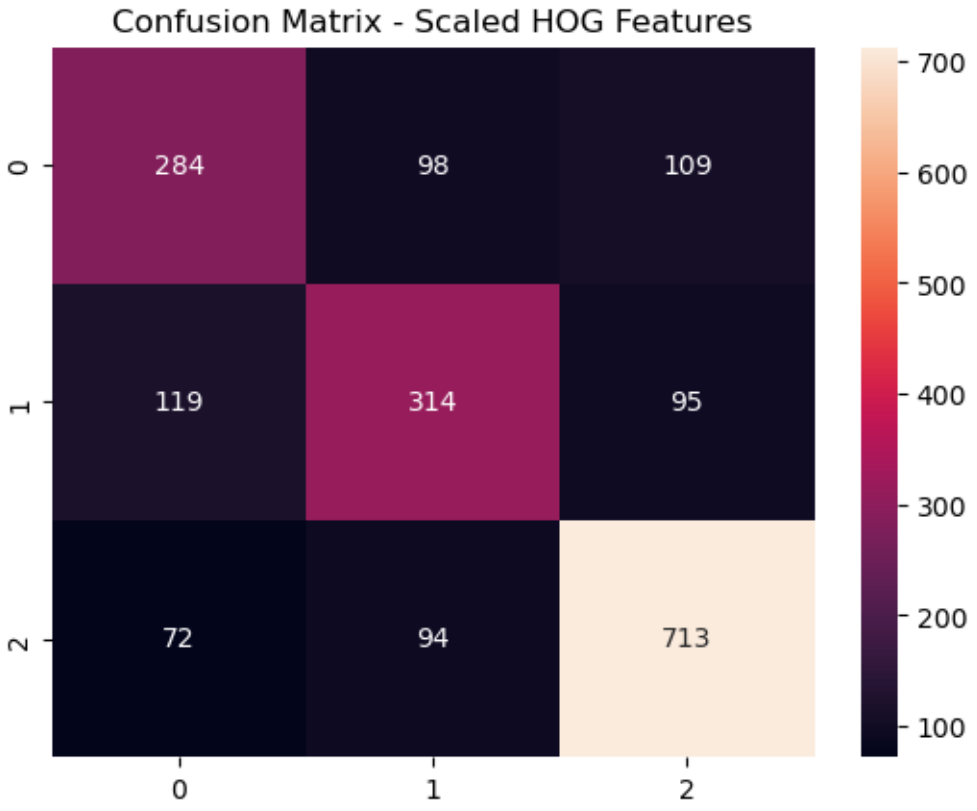
- **Cross-Validation:**
  - **Used?:** [Yes/**No**]
  - **Number of Folds:** [Number]
  - **Training/Validation Ratio:** [Ratio]
  -

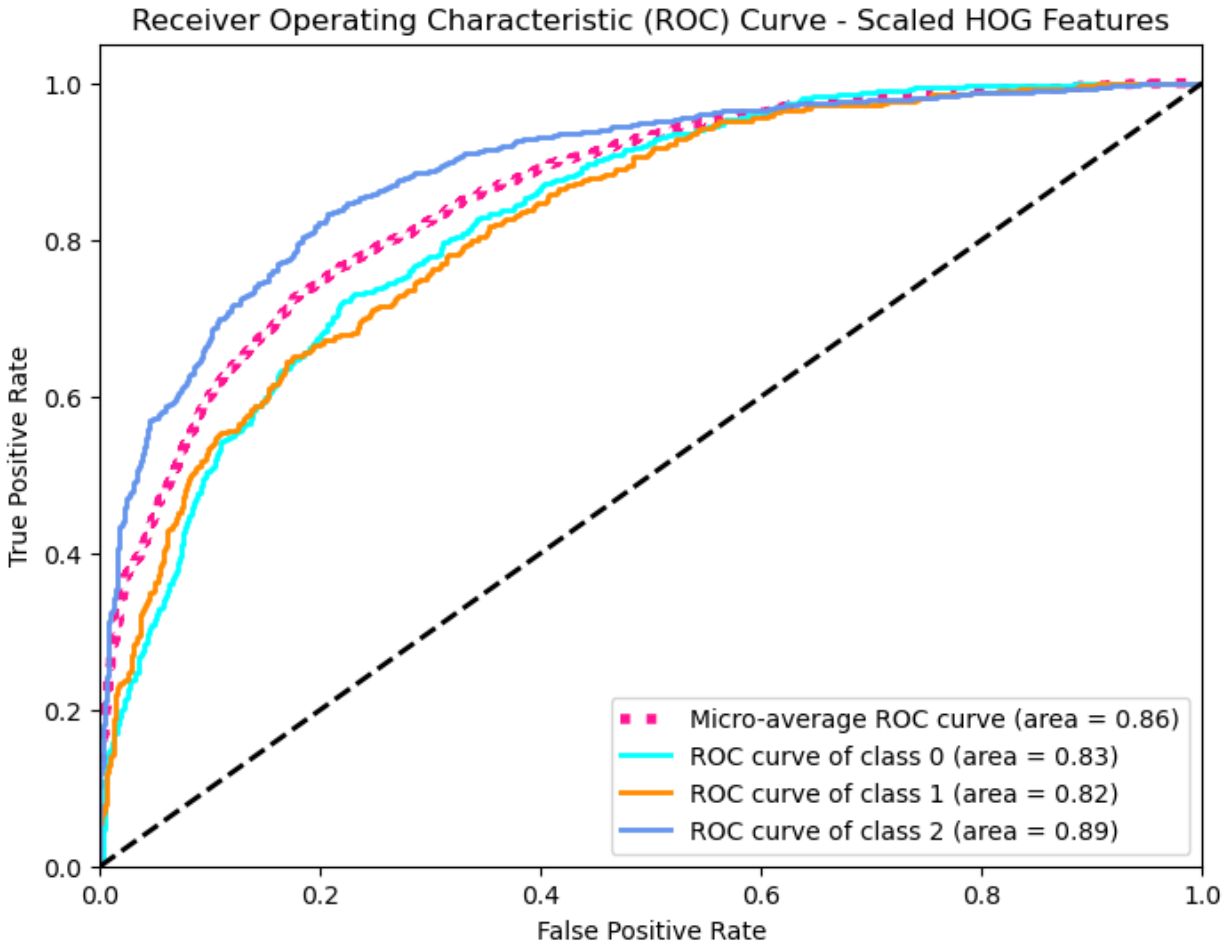
- **Hyperparameters Used:**

1. **'penalty': 'l2'], # regularization type**
2. **'C': 0.1, # regularization strength**
3. **'solver': ['saga']# solvers**

Results Details

- ## Performance on Testing Data:
- **Accuracy of Logistic Regression on HOG Features: 70%**




Receiver Operating Characteristic (ROC) Curve - Scaled HOG Features
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Micro-average ROC curve (area = 0.86)
ROC curve of class 0 (area = 0.83)
ROC curve of class 1 (area = 0.82)
ROC curve of class 2 (area = 0.89)

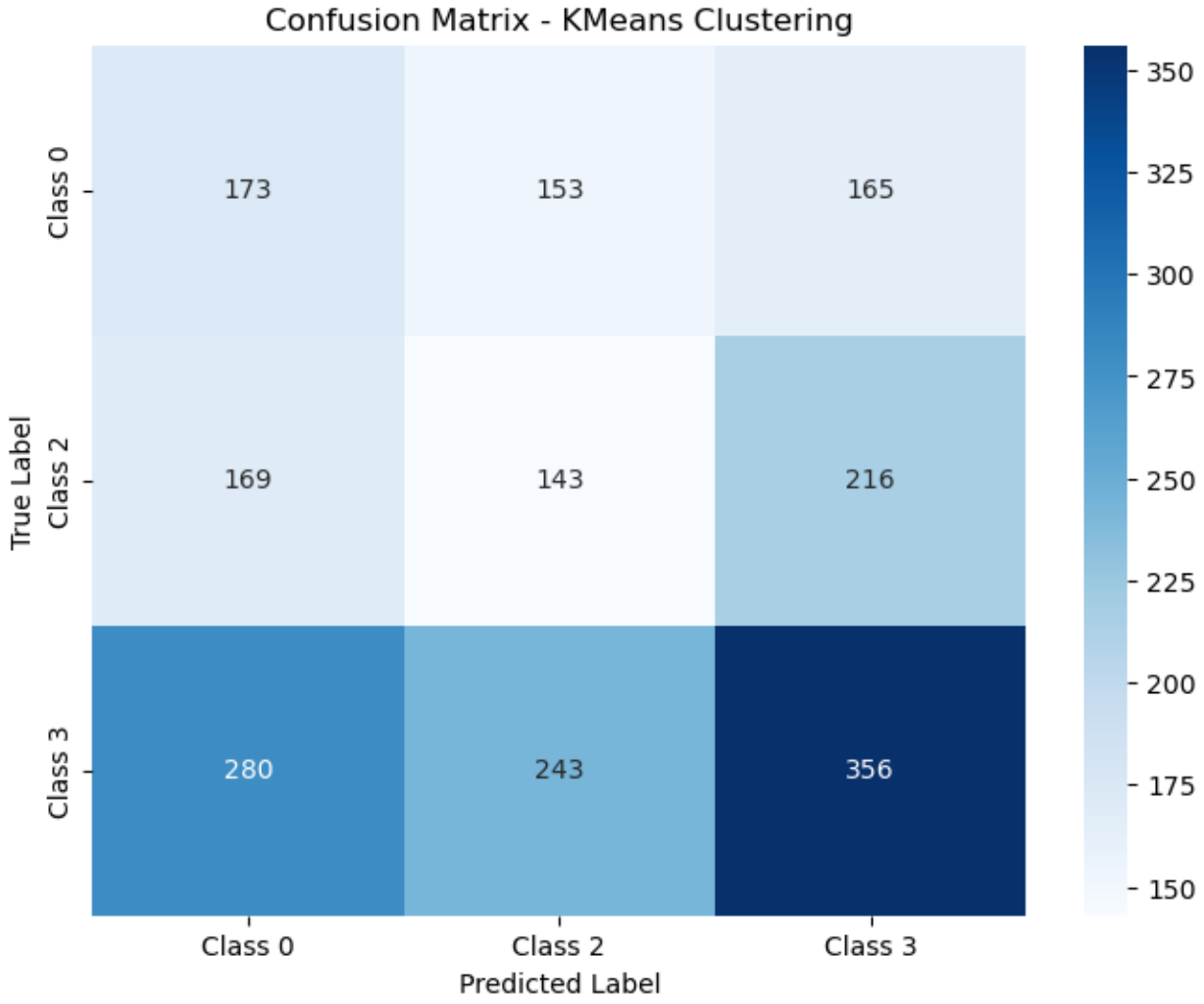
KMEANS MODEL

General Information on Dataset

- **Hyperparameters Used: k=3**

Results Details

- **Performance on Testing Data: 35%**


Confusion Matrix - KMeans Clustering
True Label
Class 0
Class 2
Class 3
173
153
165
169
143
216
280
243
356
Class 0
Class 2
Class 3
Predicted Label
350
325
300
275
250
225
200
175
150