ARTIFICIAL INTELLIGENCE MIDTERM REPORT

Ayşe AKIŞIK - 171805007

Ayşen ALPASLAN - 171805060

Betül Berna SOYLU - 171805019

Saliha APAK - 171805022

1. **How did the use of raw and preprocessed data affect learning outcome?**

Answer:

When we use raw data, especially our mse values give results in a wide range. For this reason, it becomes difficult for us to interpret the mse and R2 values we obtain for each algorithm. But when we use preprocessed data, the range gets smaller. Thus, it becomes easier for us to interpret the results. The learning results of the algorithms with preprocessed data are also more similar.

2. **How much learning performance did you achieve with preprocessed applied datasets? (All data, highest half selected with cross-correlation, highest half selected with min-mse)**

Answer:

Screenshots of the results have been added in the order given below. The mse and r2 values we obtained were used to evaluate the performance. When we evaluate these results, the performance of the learning process we perform with all data is higher. Because our dataset is small, performance has decreased in the datasets that we selected half of them and perform the learning process. If we want to rank from higher learning to low learning, a sequence is formed as follows.


```
MSE - LR: 0.471859
R2 - LR: 0.547699
MSE - RIDGE: 0.471858
R2 - RIDGE: 0.547700
MSE - LASSO: 0.470892
R2 - LASSO: 0.548752
MSE - KNN: 0.388637
R2 - KNN: 0.626473
MSE - ELASTIC: 0.471073
R2 - ELASTIC: 0.548594
MSE - DECISIONTREE: 0.478261
R2 - DECISIONTREE: 0.503959
MSE - RANDOMFOREST: 0.347455
R2 - RANDOMFOREST: 0.667009
MSE - SVR: 0.354806
R2 - SVR: 0.679534
```

```
MSE - LR: 0.517802
R2 - LR: 0.488533
MSE - RIDGE: 0.517801
R2 - RIDGE: 0.488535
MSE - LASSO: 0.517651
R2 - LASSO: 0.488946
MSE - KNN: 0.423387
R2 - KNN: 0.563908
MSE - ELASTIC: 0.517572
R2 - ELASTIC: 0.488917
MSE - DECISIONTREE: 0.493383
R2 - DECISIONTREE: 0.521577
MSE - RANDOMFOREST: 0.352139
R2 - RANDOMFOREST: 0.652665
MSE - SVR: 0.395516
R2 - SVR: 0.631820
```

```
MSE - LR: 0.539873
R2 - LR: 0.464421
MSE - RIDGE: 0.539872
R2 - RIDGE: 0.464423
MSE - LASSO: 0.539866
R2 - LASSO: 0.464737
MSE - KNN: 0.430927
R2 - KNN: 0.554302
MSE - ELASTIC: 0.539771
R2 - ELASTIC: 0.464755
MSE - DECISIONTREE: 0.634098
R2 - DECISIONTREE: 0.305277
MSE - RANDOMFOREST: 0.403357
R2 - RANDOMFOREST: 0.590271
MSE - SVR: 0.477722
R2 - SVR: 0.534100
```

3. Which algorithm learned better as a result of all operations?

Answer:

The algorithm that performs the best learning is determined by looking at the lowest value among mse values or the highest value among R2 values. We used the mse value while determining it.

Among the average "mse" values in the algorithms found in each data type, the lowest mse value, that is, the best learning (accuracy) value, and the algorithm with this value were calculated. As seen in the screenshots below, it is seen that the best learning for each data type is realized in the "Random Forest" algorithm.

```
MSE - LR: 87.147352
R2 - LR: 0.547699
MSE - RIDGE: 87.164862
R2 - RIDGE: 0.548665
MSE - LASSO: 87.853690
R2 - LASSO: 0.545788
MSE - KNN: 71.653128
R2 - KNN: 0.606157
MSE - ELASTIC: 91.819798
R2 - ELASTIC: 0.521485
MSE - DECISIONTREE: 90.167353
R2 - DECISIONTREE: 0.483345
MSE - RANDOMFOREST: 63.225642
R2 - RANDOMFOREST: 0.647156
MSE - SVR: 95.139759
R2 - SVR: 0.497739

----------------------------------

Best learning algorithm (Accuracy):
RANDOMFOREST -> 63.225641965324414
```
code_1_raw

```
MSE - LR: 0.471859
R2 - LR: 0.547699
MSE - RIDGE: 0.471858
R2 - RIDGE: 0.547700
MSE - LASSO: 0.470892
R2 - LASSO: 0.548752
MSE - KNN: 0.388637
R2 - KNN: 0.626473
MSE - ELASTIC: 0.471073
R2 - ELASTIC: 0.548594
MSE - DECISIONTREE: 0.473981
R2 - DECISIONTREE: 0.479887
MSE - RANDOMFOREST: 0.345640
R2 - RANDOMFOREST: 0.656033
MSE - SVR: 0.354806
R2 - SVR: 0.679534

----------------------------------

Best learning algorithm (Accuracy):
RANDOMFOREST -> 0.3456403767719357
```
code_2_preprocessed

```
MSE - LR: 0.539873
R2 - LR: 0.464421
MSE - RIDGE: 0.539872
R2 - RIDGE: 0.464423
MSE - LASSO: 0.539866
R2 - LASSO: 0.464737
MSE - KNN: 0.430927
R2 - KNN: 0.554302
MSE - ELASTIC: 0.539771
R2 - ELASTIC: 0.464755
MSE - DECISIONTREE: 0.641020
R2 - DECISIONTREE: 0.305058
MSE - RANDOMFOREST: 0.388532
R2 - RANDOMFOREST: 0.592282
MSE - SVR: 0.477722
R2 - SVR: 0.534100

----------------------------------

Best learning algorithm (Accuracy):
RANDOMFOREST -> 0.3885323767934114
```
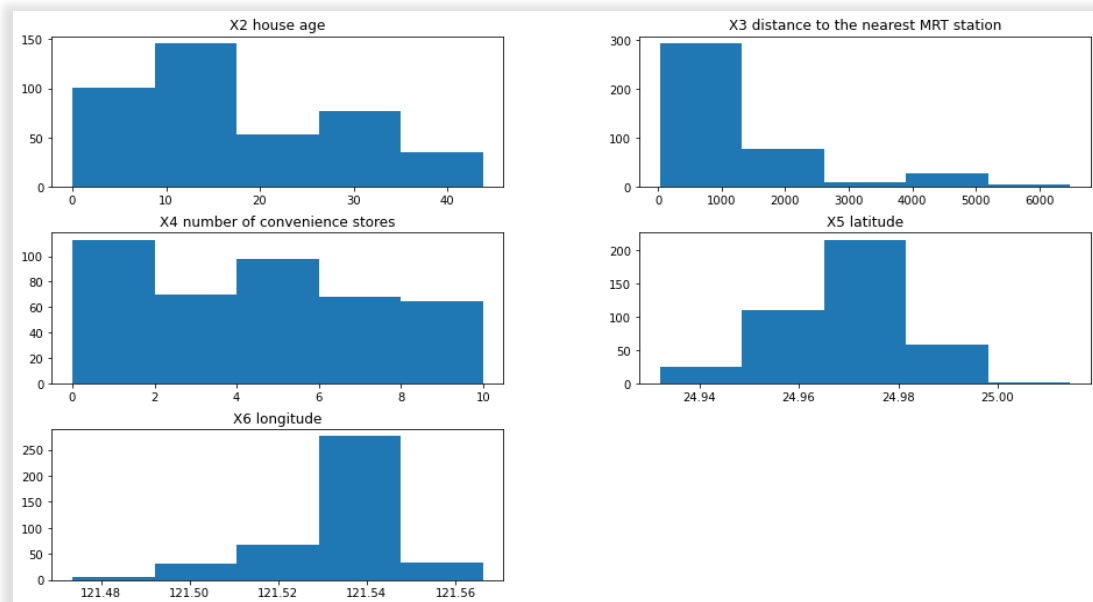*code_3_preprocessed_top_cross_correlation*

```
MSE - LR: 0.517802
R2 - LR: 0.488533
MSE - RIDGE: 0.517801
R2 - RIDGE: 0.488535
MSE - LASSO: 0.517651
R2 - LASSO: 0.488946
MSE - KNN: 0.423387
R2 - KNN: 0.563908
MSE - ELASTIC: 0.517572
R2 - ELASTIC: 0.488917
MSE - DECISIONTREE: 0.461338
R2 - DECISIONTREE: 0.536457
MSE - RANDOMFOREST: 0.348288
R2 - RANDOMFOREST: 0.657047
MSE - SVR: 0.395516
R2 - SVR: 0.631820

----------------------------------

Best learning algorithm (Accuracy):
RANDOMFOREST -> 0.3482881351688879
```
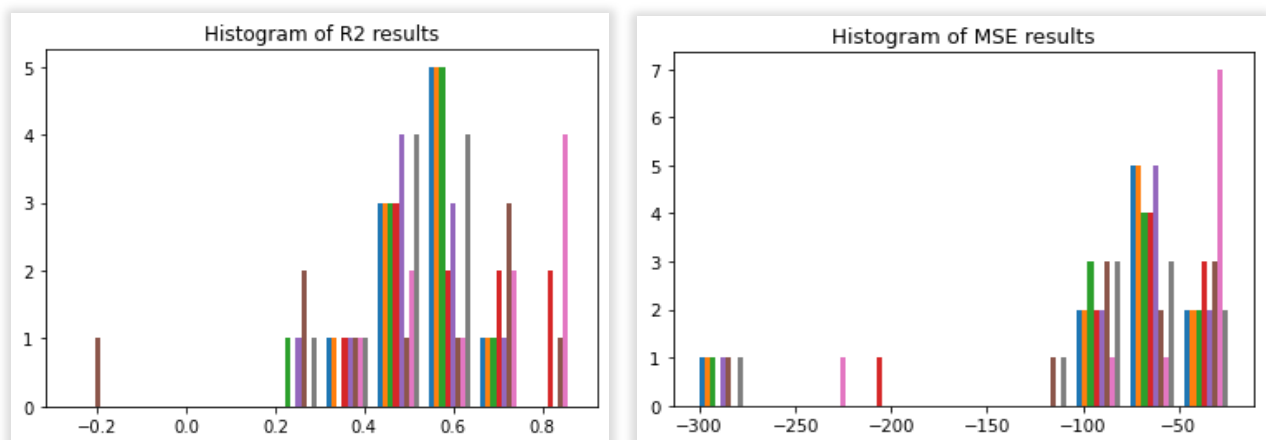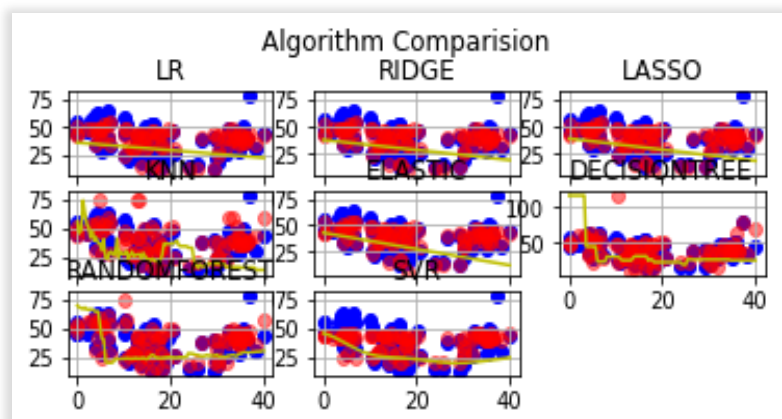*code_4_preprocessed_min_mse*

code_1_raw Plots:

o We have 6 different variables in our data set, 5 inputs and 1 output. Distributions of the input data are obtained by comparing them with the output data. Its graph for "code_1_raw" is as follows.
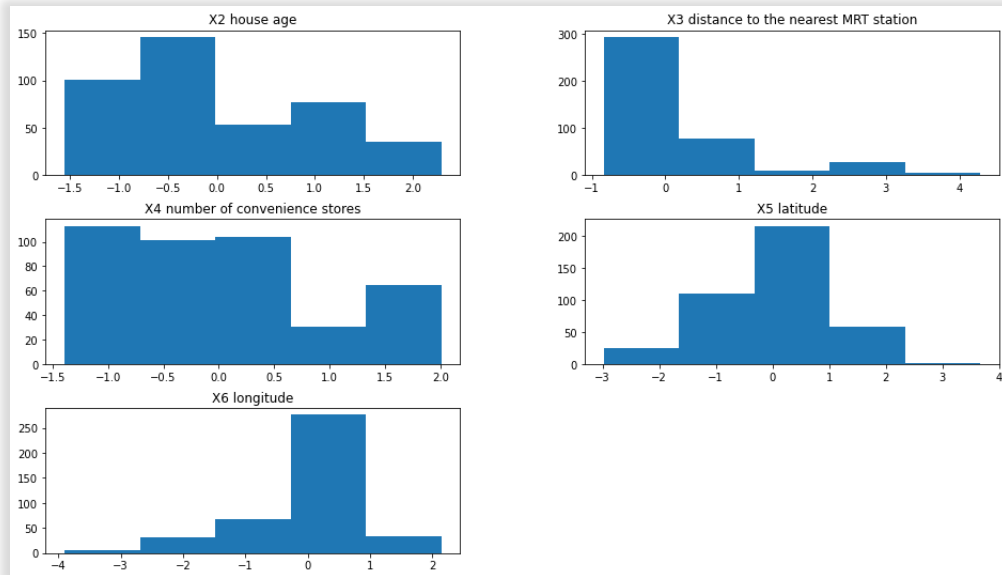

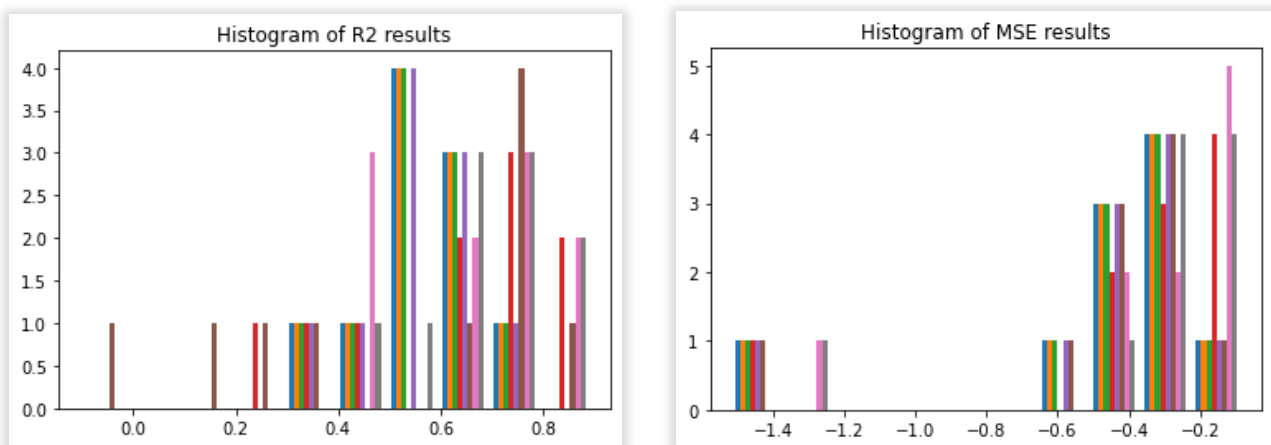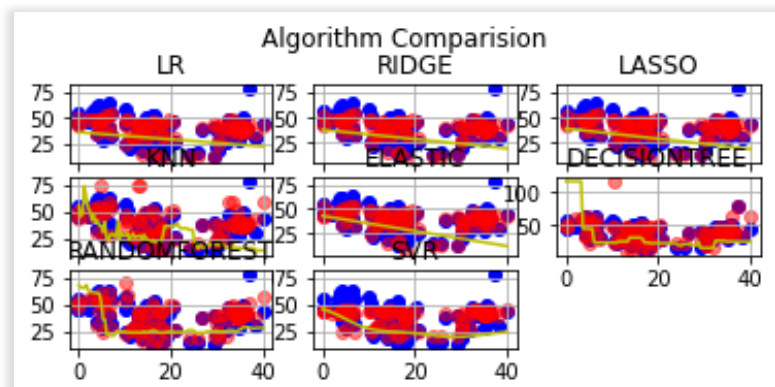o Histogram displays of mse and R2 values calculated after 10 fold process are as follows.

o The graphs of the learning that all data shows as a result of the algorithms are shown on the side.

code_2_preproccesed Plots:

o   We have 6 different variables in our data set, 5 inputs and 1 output. Distributions of the input data are obtained by comparing them with the output data. Its graph for "code_2_preprocessed" is as follows.

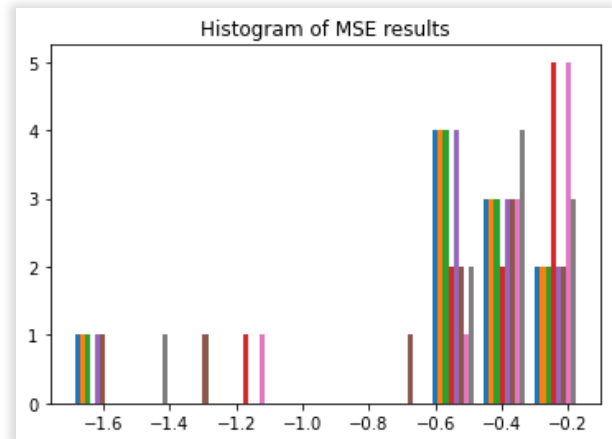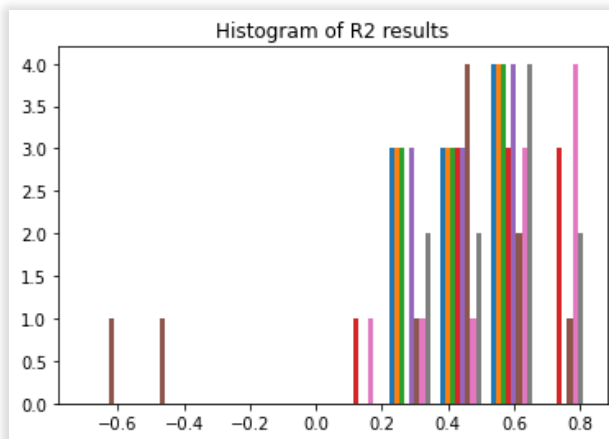o   Histogram displays of mse and R2 values calculated after 10 fold process are as follows.

o   The graphs of the learning that all data shows as a result of the algorithms are shown on the side.
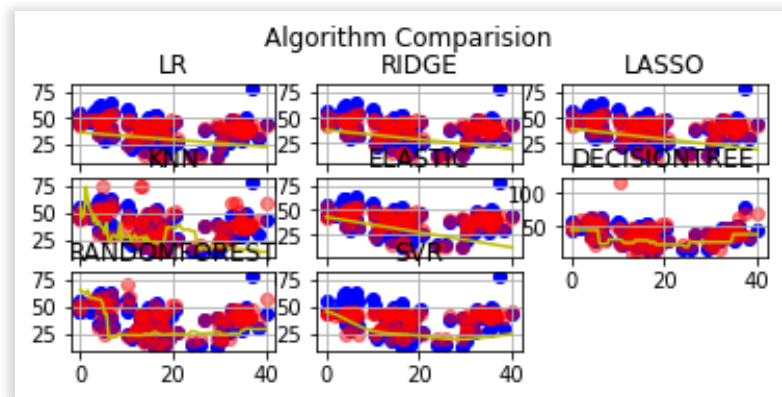
code_3_ preproccesed_top_cross_correlation Plots:

o   Histogram displays of mse and R2 values calculated after 10 fold process are as follows.
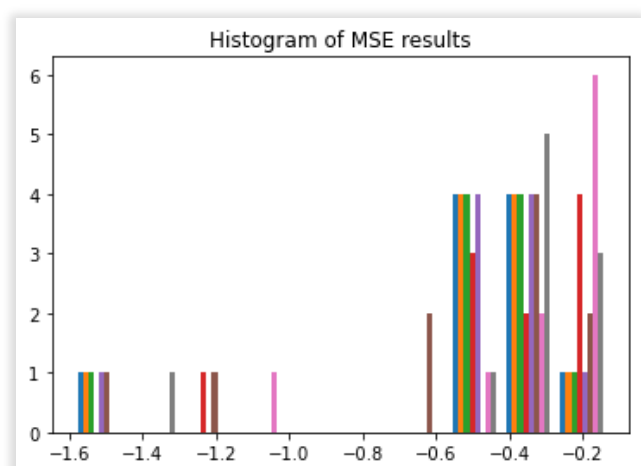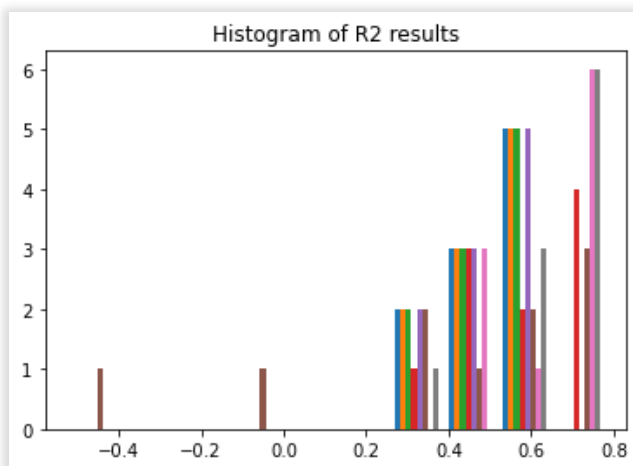

o   The graphs of the learning that all data shows as a result of the algorithms are shown on the side.

code_4_ preproccesed_min_mse Plots:

o   Histogram displays of mse and R2 values calculated after 10 fold process are as follows.

○ The graphs of the learning that all data shows as a result of the algorithms are shown on the side.

Algorithm Comparision