KNN algorithmn

```
% Load the data from social network file
Social_network = readtable("C:\Users\pc user\Documents\MATLAB\Classification\K-Nearest Neighbor\Social_Network_Ads.csv")
```

Social_network = 400×5 table

	UserID	Gender	Age	EstimatedSalary	Purchased
1	15624510	'Male'	19	19000	'Not Purchased'
2	15810944	'Male'	35	20000	'Not Purchased'
3	15668575	'Female'	26	43000	'Not Purchased'
4	15603246	'Female'	27	57000	'Not Purchased'
5	15804002	'Male'	19	76000	'Not Purchased'
6	15728773	'Male'	27	58000	'Not Purchased'
7	15598044	'Female'	27	84000	'Not Purchased'
8	15694829	'Female'	32	150000	'Purchased'
9	15600575	'Male'	25	33000	'Not Purchased'
10	15727311	'Female'	35	65000	'Not Purchased'
11	15570769	'Female'	26	80000	'Not Purchased'
12	15606274	'Female'	26	52000	'Not Purchased'
13	15746139	'Male'	20	86000	'Not Purchased'
14	15704987	'Male'	32	18000	'Not Purchased'
15	15628972	'Male'	18	82000	'Not Purchased'
16	15697686	'Male'	29	80000	'Not Purchased'
17	15733883	'Male'	47	25000	'Purchased'
18	15617482	'Male'	45	26000	'Purchased'
19	15704583	'Male'	46	28000	'Purchased'
20	15621083	'Female'	48	29000	'Purchased'
21	15649487	'Male'	45	22000	'Purchased'
22	15736760	'Female'	47	49000	'Purchased'
23	15714658	'Male'	48	41000	'Purchased'
24	15599081	'Female'	45	22000	'Purchased'
25	15705113	'Male'	46	23000	'Purchased'
26	15631159	'Male'	47	20000	'Purchased'
27	15792818	'Male'	49	28000	'Purchased'
28	15633531	'Female'	47	30000	'Purchased'
29	15744529	'Male'	29	43000	'Not Purchased'

	UserID	Gender	Age	EstimatedSalary	Purchased
30	15669656	'Male'	31	18000	'Not Purchased'
31	15581198	'Male'	31	74000	'Not Purchased'
32	15729054	'Female'	27	137000	'Purchased'
33	15573452	'Female'	21	16000	'Not Purchased'
34	15776733	'Female'	28	44000	'Not Purchased'
35	15724858	'Male'	27	90000	'Not Purchased'
36	15713144	'Male'	35	27000	'Not Purchased'
37	15690188	'Female'	33	28000	'Not Purchased'
38	15689425	'Male'	30	49000	'Not Purchased'
39	15671766	'Female'	26	72000	'Not Purchased'
40	15782806	'Female'	27	31000	'Not Purchased'
41	15764419	'Female'	27	17000	'Not Purchased'
42	15591915	'Female'	33	51000	'Not Purchased'
43	15772798	'Male'	35	108000	'Not Purchased'
44	15792008	'Male'	30	15000	'Not Purchased'
45	15715541	'Female'	28	84000	'Not Purchased'
46	15639277	'Male'	23	20000	'Not Purchased'
47	15798850	'Male'	25	79000	'Not Purchased'
48	15776348	'Female'	27	54000	'Not Purchased'
49	15727696	'Male'	30	135000	'Purchased'
50	15793813	'Female'	31	89000	'Not Purchased'
51	15694395	'Female'	24	32000	'Not Purchased'
52	15764195	'Female'	18	44000	'Not Purchased'
53	15744919	'Female'	29	83000	'Not Purchased'
54	15671655	'Female'	35	23000	'Not Purchased'
55	15654901	'Female'	27	58000	'Not Purchased'
56	15649136	'Female'	24	55000	'Not Purchased'
57	15775562	'Female'	23	48000	'Not Purchased'
58	15807481	'Male'	28	79000	'Not Purchased'
59	15642885	'Male'	22	18000	'Not Purchased'
60	15789109	'Female'	32	117000	'Not Purchased'
61	15814004	'Male'	27	20000	'Not Purchased'
62	15673619	'Male'	25	87000	'Not Purchased'

	UserID	Gender	Age	EstimatedSalary	Purchased
63	15595135	'Female'	23	66000	'Not Purchased'
64	15583681	'Male'	32	120000	'Purchased'
65	15605000	'Female'	59	83000	'Not Purchased'
66	15718071	'Male'	24	58000	'Not Purchased'
67	15679760	'Male'	24	19000	'Not Purchased'
68	15654574	'Female'	23	82000	'Not Purchased'
69	15577178	'Female'	22	63000	'Not Purchased'
70	15595324	'Female'	31	68000	'Not Purchased'
71	15756932	'Male'	25	80000	'Not Purchased'
72	15726358	'Female'	24	27000	'Not Purchased'
73	15595228	'Female'	20	23000	'Not Purchased'
74	15782530	'Female'	33	113000	'Not Purchased'
75	15592877	'Male'	32	18000	'Not Purchased'
76	15651983	'Male'	34	112000	'Purchased'
77	15746737	'Male'	18	52000	'Not Purchased'
78	15774179	'Female'	22	27000	'Not Purchased'
79	15667265	'Female'	28	87000	'Not Purchased'
80	15655123	'Female'	26	17000	'Not Purchased'
81	15595917	'Male'	30	80000	'Not Purchased'
82	15668385	'Male'	39	42000	'Not Purchased'
83	15709476	'Male'	20	49000	'Not Purchased'
84	15711218	'Male'	35	88000	'Not Purchased'
85	15798659	'Female'	30	62000	'Not Purchased'
86	15663939	'Female'	31	118000	'Purchased'
87	15694946	'Male'	24	55000	'Not Purchased'
88	15631912	'Female'	28	85000	'Not Purchased'
89	15768816	'Male'	26	81000	'Not Purchased'
90	15682268	'Male'	35	50000	'Not Purchased'
91	15684801	'Male'	22	81000	'Not Purchased'
92	15636428	'Female'	30	116000	'Not Purchased'
93	15809823	'Male'	26	15000	'Not Purchased'
94	15699284	'Female'	29	28000	'Not Purchased'
95	15786993	'Female'	29	83000	'Not Purchased'

	UserID	Gender	Age	EstimatedSalary	Purchased
96	15709441	'Female'	35	44000	'Not Purchased'
97	15710257	'Female'	35	25000	'Not Purchased'
98	15582492	'Male'	28	123000	'Purchased'
99	15575694	'Male'	35	73000	'Not Purchased'
100	15756820	'Female'	28	37000	'Not Purchased'

⋮

```
% check if there are missing values
ismissing(Social_network)
```

```
ans = 400×5 logical array
   0   0   0   0   0
   0   0   0   0   0
   0   0   0   0   0
   0   0   0   0   0
   0   0   0   0   0
   0   0   0   0   0
   0   0   0   0   0
   0   0   0   0   0
   0   0   0   0   0
   0   0   0   0   0
     ⋮
```

```
% check for outliers
isoutlier(Social_network.Age)
```

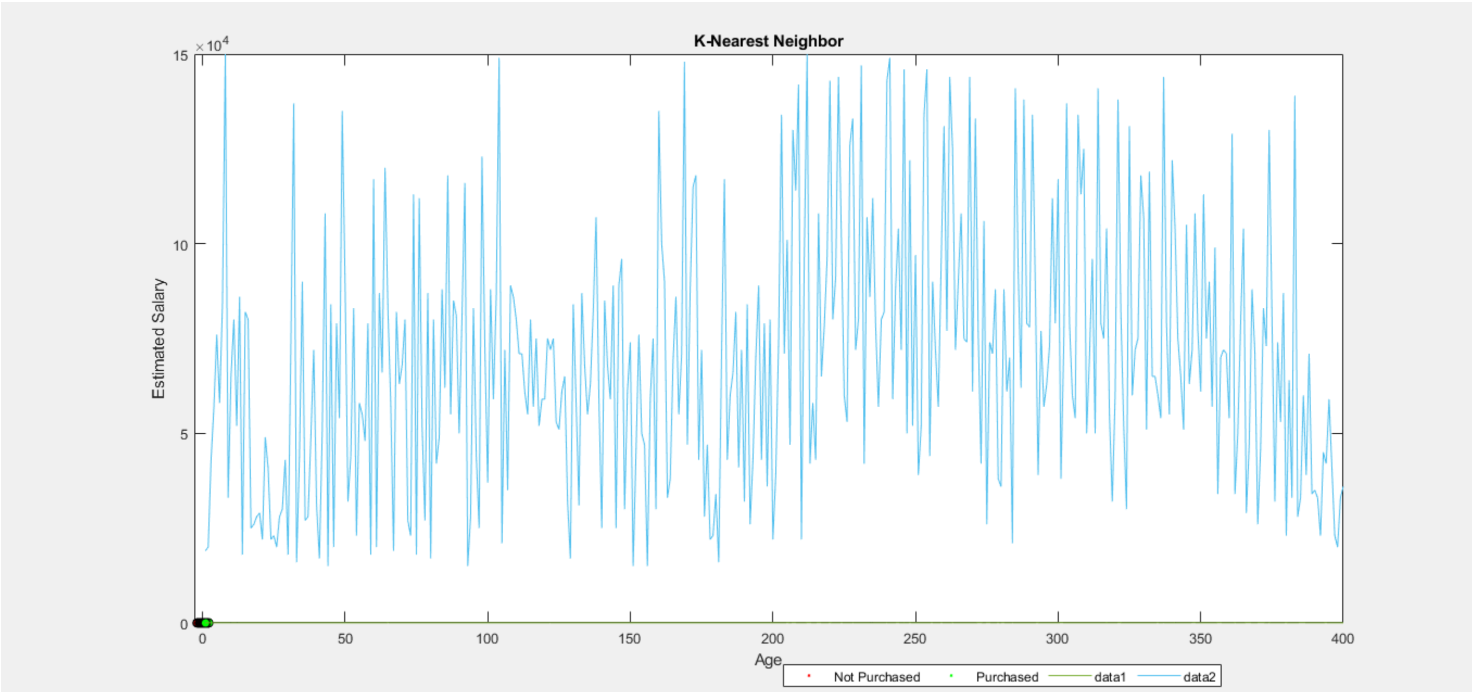
```
ans = 400×1 logical array
   0
   0
   0
   0
   0
   0
   0
   0
   0
   0
     ⋮
```

```
isoutlier(Social_network.EstimatedSalary)
```

```
ans = 400×1 logical array
   0
   0
   0
   0
   0
   0
   0
   0
   0
   0
```

⋮

```
plot(Social_network.Age)
plot(Social_network.EstimatedSalary)
```




```
% Remove the userID column
Social_network.UserID = []
```

Social_network = 400×4 table

	Gender	Age	EstimatedSalary	Purchased
1	'Male'	19	19000	'Not Purchased'
2	'Male'	35	20000	'Not Purchased'
3	'Female'	26	43000	'Not Purchased'
4	'Female'	27	57000	'Not Purchased'
5	'Male'	19	76000	'Not Purchased'
6	'Male'	27	58000	'Not Purchased'
7	'Female'	27	84000	'Not Purchased'
8	'Female'	32	150000	'Purchased'
9	'Male'	25	33000	'Not Purchased'
10	'Female'	35	65000	'Not Purchased'
11	'Female'	26	80000	'Not Purchased'
12	'Female'	26	52000	'Not Purchased'
13	'Male'	20	86000	'Not Purchased'

	Gender	Age	EstimatedSalary	Purchased
14	'Male'	32	18000	'Not Purchased'
15	'Male'	18	82000	'Not Purchased'
16	'Male'	29	80000	'Not Purchased'
17	'Male'	47	25000	'Purchased'
18	'Male'	45	26000	'Purchased'
19	'Male'	46	28000	'Purchased'
20	'Female'	48	29000	'Purchased'
21	'Male'	45	22000	'Purchased'
22	'Female'	47	49000	'Purchased'
23	'Male'	48	41000	'Purchased'
24	'Female'	45	22000	'Purchased'
25	'Male'	46	23000	'Purchased'
26	'Male'	47	20000	'Purchased'
27	'Male'	49	28000	'Purchased'
28	'Female'	47	30000	'Purchased'
29	'Male'	29	43000	'Not Purchased'
30	'Male'	31	18000	'Not Purchased'
31	'Male'	31	74000	'Not Purchased'
32	'Female'	27	137000	'Purchased'
33	'Female'	21	16000	'Not Purchased'
34	'Female'	28	44000	'Not Purchased'
35	'Male'	27	90000	'Not Purchased'
36	'Male'	35	27000	'Not Purchased'
37	'Female'	33	28000	'Not Purchased'
38	'Male'	30	49000	'Not Purchased'
39	'Female'	26	72000	'Not Purchased'
40	'Female'	27	31000	'Not Purchased'
41	'Female'	27	17000	'Not Purchased'
42	'Female'	33	51000	'Not Purchased'
43	'Male'	35	108000	'Not Purchased'
44	'Male'	30	15000	'Not Purchased'
45	'Female'	28	84000	'Not Purchased'
46	'Male'	23	20000	'Not Purchased'

	Gender	Age	EstimatedSalary	Purchased
47	'Male'	25	79000	'Not Purchased'
48	'Female'	27	54000	'Not Purchased'
49	'Male'	30	135000	'Purchased'
50	'Female'	31	89000	'Not Purchased'
51	'Female'	24	32000	'Not Purchased'
52	'Female'	18	44000	'Not Purchased'
53	'Female'	29	83000	'Not Purchased'
54	'Female'	35	23000	'Not Purchased'
55	'Female'	27	58000	'Not Purchased'
56	'Female'	24	55000	'Not Purchased'
57	'Female'	23	48000	'Not Purchased'
58	'Male'	28	79000	'Not Purchased'
59	'Male'	22	18000	'Not Purchased'
60	'Female'	32	117000	'Not Purchased'
61	'Male'	27	20000	'Not Purchased'
62	'Male'	25	87000	'Not Purchased'
63	'Female'	23	66000	'Not Purchased'
64	'Male'	32	120000	'Purchased'
65	'Female'	59	83000	'Not Purchased'
66	'Male'	24	58000	'Not Purchased'
67	'Male'	24	19000	'Not Purchased'
68	'Female'	23	82000	'Not Purchased'
69	'Female'	22	63000	'Not Purchased'
70	'Female'	31	68000	'Not Purchased'
71	'Male'	25	80000	'Not Purchased'
72	'Female'	24	27000	'Not Purchased'
73	'Female'	20	23000	'Not Purchased'
74	'Female'	33	113000	'Not Purchased'
75	'Male'	32	18000	'Not Purchased'
76	'Male'	34	112000	'Purchased'
77	'Male'	18	52000	'Not Purchased'
78	'Female'	22	27000	'Not Purchased'
79	'Female'	28	87000	'Not Purchased'

	Gender	Age	EstimatedSalary	Purchased
80	'Female'	26	17000	'Not Purchased'
81	'Male'	30	80000	'Not Purchased'
82	'Male'	39	42000	'Not Purchased'
83	'Male'	20	49000	'Not Purchased'
84	'Male'	35	88000	'Not Purchased'
85	'Female'	30	62000	'Not Purchased'
86	'Female'	31	118000	'Purchased'
87	'Male'	24	55000	'Not Purchased'
88	'Female'	28	85000	'Not Purchased'
89	'Male'	26	81000	'Not Purchased'
90	'Male'	35	50000	'Not Purchased'
91	'Male'	22	81000	'Not Purchased'
92	'Female'	30	116000	'Not Purchased'
93	'Male'	26	15000	'Not Purchased'
94	'Female'	29	28000	'Not Purchased'
95	'Female'	29	83000	'Not Purchased'
96	'Female'	35	44000	'Not Purchased'
97	'Female'	35	25000	'Not Purchased'
98	'Male'	28	123000	'Purchased'
99	'Male'	35	73000	'Not Purchased'
100	'Female'	28	37000	'Not Purchased'

⋮

Feature Scaling

```
% We need to fit Age and Estimated Salary columns to same scale
stand_age = (Social_network.Age -mean(Social_network.Age))/std(Social_network.Age);
Social_network.Age = stand_age;
stand_estimatesalary = (Social_network.EstimatedSalary -
mean(Social_network.EstimatedSalary))/std(Social_network.EstimatedSalary);
Social_network.EstimatedSalary = stand_estimatesalary;
Social_network
```

Social_network = 400×4 table

	Gender	Age	EstimatedSalary	Purchased
1	'Male'	-1.7796	-1.4882	'Not Purchased'
2	'Male'	-0.2533	-1.4589	'Not Purchased'

	Gender	Age	EstimatedSalary	Purchased
3	'Female'	-1.1118	-0.7843	'Not Purchased'
4	'Female'	-1.0164	-0.3737	'Not Purchased'
5	'Male'	-1.7796	0.1835	'Not Purchased'
6	'Male'	-1.0164	-0.3444	'Not Purchased'
7	'Female'	-1.0164	0.4181	'Not Purchased'
8	'Female'	-0.5395	2.3538	'Purchased'
9	'Male'	-1.2072	-1.0776	'Not Purchased'
10	'Female'	-0.2533	-0.1391	'Not Purchased'
11	'Female'	-1.1118	0.3008	'Not Purchased'
12	'Female'	-1.1118	-0.5204	'Not Purchased'
13	'Male'	-1.6842	0.4768	'Not Purchased'
14	'Male'	-0.5395	-1.5175	'Not Purchased'
15	'Male'	-1.8750	0.3595	'Not Purchased'
16	'Male'	-0.8256	0.3008	'Not Purchased'
17	'Male'	0.8915	-1.3122	'Purchased'
18	'Male'	0.7007	-1.2829	'Purchased'
19	'Male'	0.7961	-1.2242	'Purchased'
20	'Female'	0.9868	-1.1949	'Purchased'
21	'Male'	0.7007	-1.4002	'Purchased'
22	'Female'	0.8915	-0.6083	'Purchased'
23	'Male'	0.9868	-0.8430	'Purchased'
24	'Female'	0.7007	-1.4002	'Purchased'
25	'Male'	0.7961	-1.3709	'Purchased'
26	'Male'	0.8915	-1.4589	'Purchased'
27	'Male'	1.0822	-1.2242	'Purchased'
28	'Female'	0.8915	-1.1656	'Purchased'
29	'Male'	-0.8256	-0.7843	'Not Purchased'
30	'Male'	-0.6348	-1.5175	'Not Purchased'
31	'Male'	-0.6348	0.1249	'Not Purchased'
32	'Female'	-1.0164	1.9725	'Purchased'
33	'Female'	-1.5888	-1.5762	'Not Purchased'
34	'Female'	-0.9210	-0.7550	'Not Purchased'
35	'Male'	-1.0164	0.5941	'Not Purchased'

	Gender	Age	EstimatedSalary	Purchased
36	'Male'	-0.2533	-1.2536	'Not Purchased'
37	'Female'	-0.4441	-1.2242	'Not Purchased'
38	'Male'	-0.7302	-0.6083	'Not Purchased'
39	'Female'	-1.1118	0.0662	'Not Purchased'
40	'Female'	-1.0164	-1.1362	'Not Purchased'
41	'Female'	-1.0164	-1.5468	'Not Purchased'
42	'Female'	-0.4441	-0.5497	'Not Purchased'
43	'Male'	-0.2533	1.1220	'Not Purchased'
44	'Male'	-0.7302	-1.6055	'Not Purchased'
45	'Female'	-0.9210	0.4181	'Not Purchased'
46	'Male'	-1.3980	-1.4589	'Not Purchased'
47	'Male'	-1.2072	0.2715	'Not Purchased'
48	'Female'	-1.0164	-0.4617	'Not Purchased'
49	'Male'	-0.7302	1.9139	'Purchased'
50	'Female'	-0.6348	0.5648	'Not Purchased'
51	'Female'	-1.3026	-1.1069	'Not Purchased'
52	'Female'	-1.8750	-0.7550	'Not Purchased'
53	'Female'	-0.8256	0.3888	'Not Purchased'
54	'Female'	-0.2533	-1.3709	'Not Purchased'
55	'Female'	-1.0164	-0.3444	'Not Purchased'
56	'Female'	-1.3026	-0.4324	'Not Purchased'
57	'Female'	-1.3980	-0.6377	'Not Purchased'
58	'Male'	-0.9210	0.2715	'Not Purchased'
59	'Male'	-1.4934	-1.5175	'Not Purchased'
60	'Female'	-0.5395	1.3860	'Not Purchased'
61	'Male'	-1.0164	-1.4589	'Not Purchased'
62	'Male'	-1.2072	0.5061	'Not Purchased'
63	'Female'	-1.3980	-0.1098	'Not Purchased'
64	'Male'	-0.5395	1.4740	'Purchased'
65	'Female'	2.0362	0.3888	'Not Purchased'
66	'Male'	-1.3026	-0.3444	'Not Purchased'
67	'Male'	-1.3026	-1.4882	'Not Purchased'
68	'Female'	-1.3980	0.3595	'Not Purchased'

	Gender	Age	EstimatedSalary	Purchased
69	'Female'	-1.4934	-0.1977	'Not Purchased'
70	'Female'	-0.6348	-0.0511	'Not Purchased'
71	'Male'	-1.2072	0.3008	'Not Purchased'
72	'Female'	-1.3026	-1.2536	'Not Purchased'
73	'Female'	-1.6842	-1.3709	'Not Purchased'
74	'Female'	-0.4441	1.2687	'Not Purchased'
75	'Male'	-0.5395	-1.5175	'Not Purchased'
76	'Male'	-0.3487	1.2393	'Purchased'
77	'Male'	-1.8750	-0.5204	'Not Purchased'
78	'Female'	-1.4934	-1.2536	'Not Purchased'
79	'Female'	-0.9210	0.5061	'Not Purchased'
80	'Female'	-1.1118	-1.5468	'Not Purchased'
81	'Male'	-0.7302	0.3008	'Not Purchased'
82	'Male'	0.1283	-0.8136	'Not Purchased'
83	'Male'	-1.6842	-0.6083	'Not Purchased'
84	'Male'	-0.2533	0.5355	'Not Purchased'
85	'Female'	-0.7302	-0.2271	'Not Purchased'
86	'Female'	-0.6348	1.4153	'Purchased'
87	'Male'	-1.3026	-0.4324	'Not Purchased'
88	'Female'	-0.9210	0.4475	'Not Purchased'
89	'Male'	-1.1118	0.3302	'Not Purchased'
90	'Male'	-0.2533	-0.5790	'Not Purchased'
91	'Male'	-1.4934	0.3302	'Not Purchased'
92	'Female'	-0.7302	1.3566	'Not Purchased'
93	'Male'	-1.1118	-1.6055	'Not Purchased'
94	'Female'	-0.8256	-1.2242	'Not Purchased'
95	'Female'	-0.8256	0.3888	'Not Purchased'
96	'Female'	-0.2533	-0.7550	'Not Purchased'
97	'Female'	-0.2533	-1.3122	'Not Purchased'
98	'Male'	-0.9210	1.5619	'Purchased'
99	'Male'	-0.2533	0.0955	'Not Purchased'
100	'Female'	-0.9210	-0.9603	'Not Purchased'

⋮

Let's Build KNN Model

```
% KNN model
classification_model = fitcknn(Social_network,'Purchased ~ Age+EstimatedSalary');
classification_model
```

```
classification_model = 
  ClassificationKNN
           PredictorNames: {'Age'  'EstimatedSalary'}
             ResponseName: 'Purchased'
    CategoricalPredictors: []
               ClassNames: {'Not Purchased'  'Purchased'}
           ScoreTransform: 'none'
          NumObservations: 400
                 Distance: 'euclidean'
             NumNeighbors: 1


  Properties, Methods
```

```
% Let's split the data into training and test data
cv = cvpartition(classification_model.NumObservations, 'HoldOut', 0.2);
cv
```

```
cv = 
Hold-out cross validation partition
   NumObservations: 400
       NumTestSets: 1
         TrainSize: 320
          TestSize: 80
          IsCustom: 0
```

```
% Let's build KNN model based on training data
cross_validation_model = crossval(classification_model, 'cvpartition',cv);
cross_validation_model
```

```
cross_validation_model = 
  ClassificationPartitionedModel
    CrossValidatedModel: 'KNN'
         PredictorNames: {'Age'  'EstimatedSalary'}
           ResponseName: 'Purchased'
        NumObservations: 400
                  KFold: 1
              Partition: [1×1 cvpartition]
             ClassNames: {'Not Purchased'  'Purchased'}
         ScoreTransform: 'none'


  Properties, Methods
```

Predict the classifier on the testing set

```
% Predict on the testing set
Predictions = predict(cross_validation_model.Trained{1},
Social_network(test(cv),1:end-1));
Predictions
```

```
Predictions = 80×1 cell
'Not Purchased'
'Not Purchased'
'Not Purchased'
'Not Purchased'
'Purchased'
'Not Purchased'
'Not Purchased'
'Purchased'
'Not Purchased'
'Not Purchased'
      ⋮
```

Analyze the predictions

```
% Let's use a Confusion Matrix
Results = confusionmat(cross_validation_model.Y(test(cv)), Predictions);
Results
```

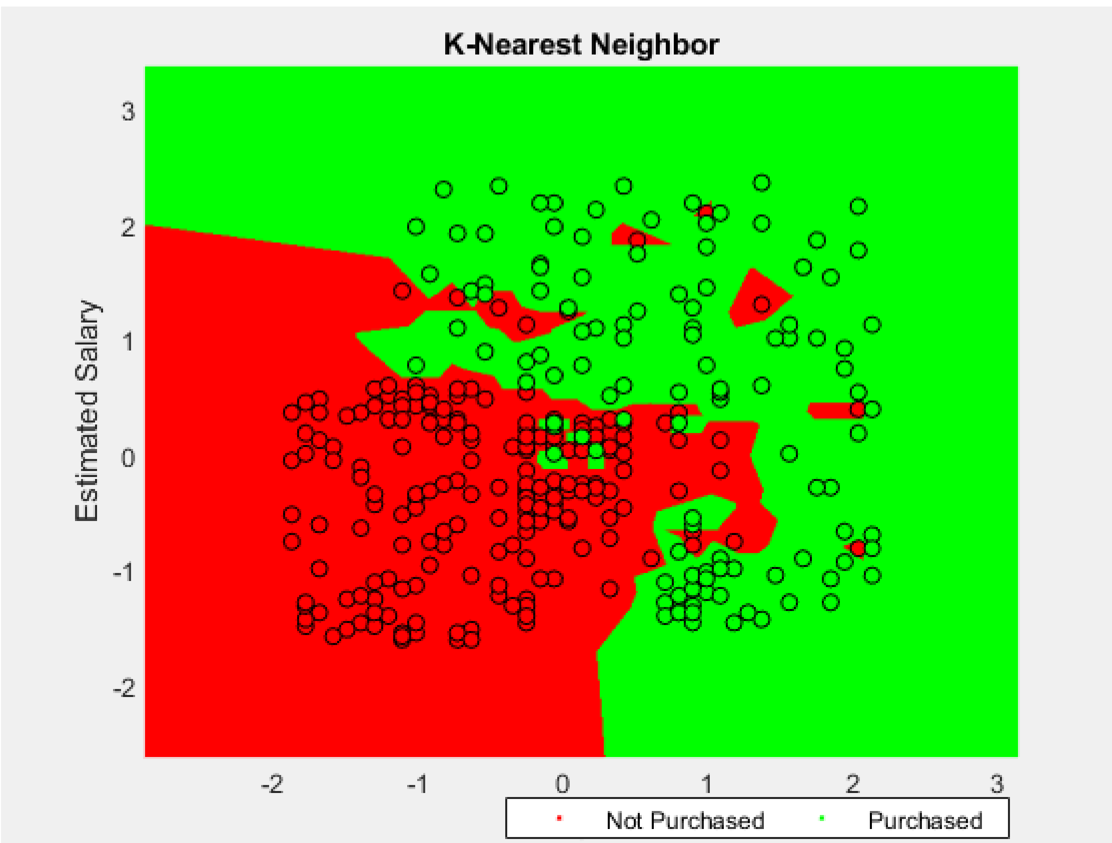
```
Results = 2×2
    42     9
     8    21
```

Visualizing the KNN model results

```
% setting labels and classifier names
labels = unique(Social_network.Purchased);
classifier_name = 'K-Nearest Neighbor';

% computing the minimum and maximum Age and Salary
Age_range = min(Social_network.Age(training(cv)))
-1:0.01:max(Social_network.Age(training(cv)))+1;
Estimated_salary_range = min(Social_network.EstimatedSalary(training(cv)))
-1:0.01:max(Social_network.EstimatedSalary(training(cv)))+1;
```

```
% Let's create 2D co-ordinates
[xx1, xx2] = meshgrid(Age_range,Estimated_salary_range);
XGrid = [xx1(:) xx2(:)];

% Create prediction meshgrid
predictions_meshgrid = predict(cross_validation_model.Trained{1},XGrid);

% Plot the predictions on the grid
figure;

gscatter(xx1(:), xx2(:), predictions_meshgrid, 'rgb');

hold on

training_data = Social_network(training(cv),:);
```

```
Y = ismember(training_data.Purchased,labels(1));

scatter(training_data.Age(Y), training_data.EstimatedSalary(Y), 'o',
'MarkerEdgeColor','black','MarkerFaceColor','red');
scatter(training_data.Age(~Y),training_data.EstimatedSalary(~Y), 'o',
'MarkerEdgeColor','black','MarkerFaceColor','green');

xlabel('Age');
ylabel('Estimated Salary');

title(classifier_name);
legend off;
axis tight;

legend(labels,'Location',[0.45,0.01,0.45,0.05],'Orientation','Horizontal');

% Let's bring the figure to the front
set(0, 'DefaultFigureVisible', 'on');
```




```
figure(gcf);
```