Question 1-1 is available in the file “Assignment4_Q-1-1.pdf”

Question 2:

1-a. What is the accuracy on the training set?

```
Accuracy of train: 1.0
```

This is because the decision Tree always overfits to the training data.

1-b. What is the accuracy on the test set?

```
Accuracy of test: 0.4
```

2. What is the effect of restricting the maximum depth of the tree? Try different depths and find the best value.

```
Depth of the original tree 5

Current Tree depth: 1
Accuracy of train: 0.55
  Construction type
0         Apartment
1             House
2             House
3         Apartment
4         Apartment
['Apartment' 'Apartment' 'House' 'House' 'House']
Accuracy of test: 0.4

Current Tree depth: 2
Accuracy of train: 0.75
  Construction type
0         Apartment
1             House
2             House
3         Apartment
4         Apartment
['Condo' 'Condo' 'House' 'Apartment' 'Apartment']
Accuracy of test: 0.6

Current Tree depth: 3
Accuracy of train: 0.9
  Construction type
0         Apartment
1             House
2             House
3         Apartment
4         Apartment
['Condo' 'Condo' 'Condo' 'Apartment' 'Apartment']
Accuracy of test: 0.4
```

```
Current Tree depth: 4
Accuracy of train: 0.95
  Construction type
0         Apartment
1             House
2             House
3         Apartment
4         Apartment
['Condo' 'Condo' 'Condo' 'Apartment' 'Apartment']
Accuracy of test: 0.4

Current Tree depth: 5
Accuracy of train: 1.0
  Construction type
0         Apartment
1             House
2             House
3         Apartment
4         Apartment
['Condo' 'Condo' 'Condo' 'Apartment' 'Apartment']
Accuracy of test: 0.4
```

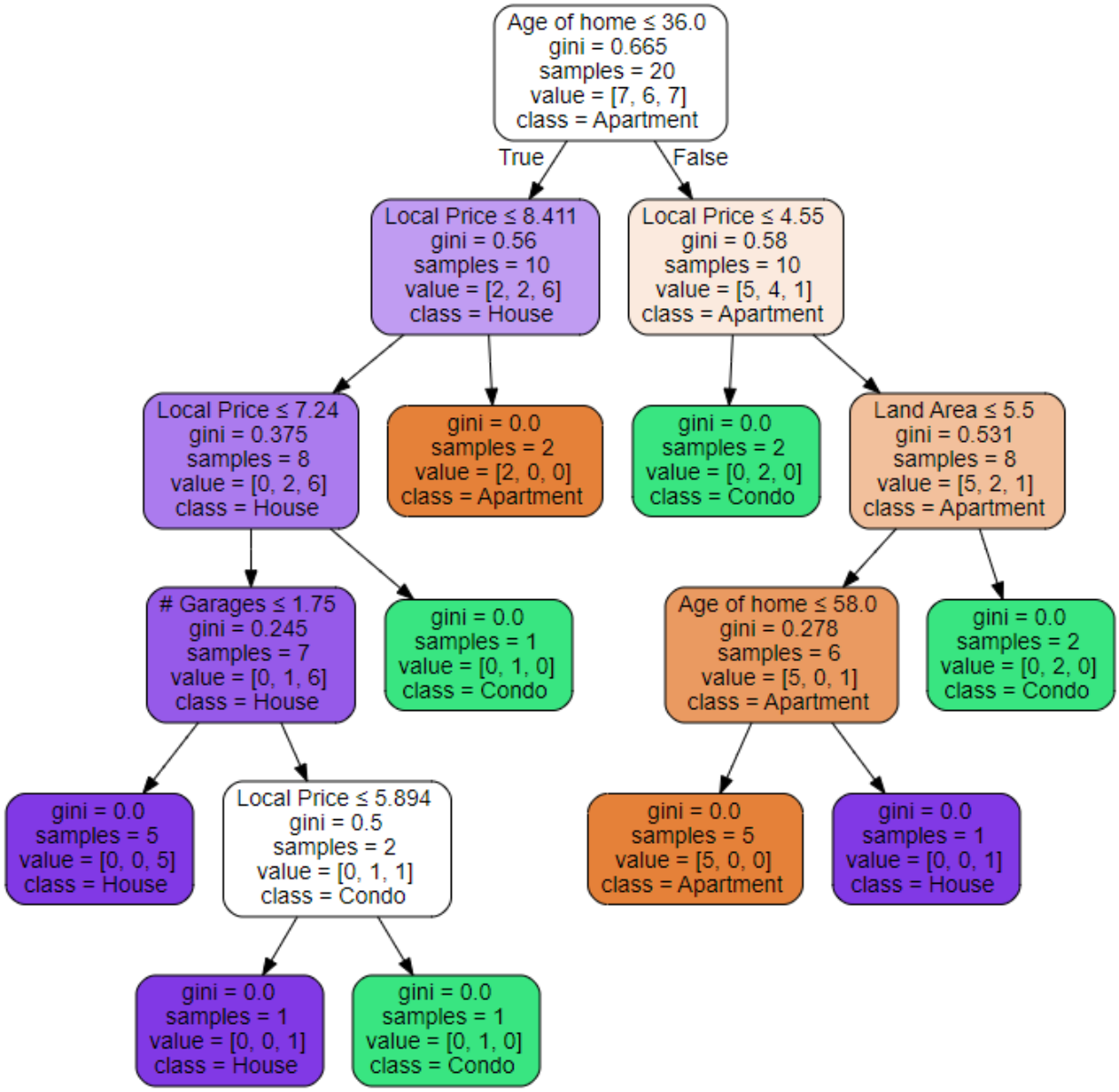
The effect of restring the maximum depth of the tree increases the test accuracy and drops the train accuracy.

3. Why does restricting the depth have such a strong effect on the classifier performance?

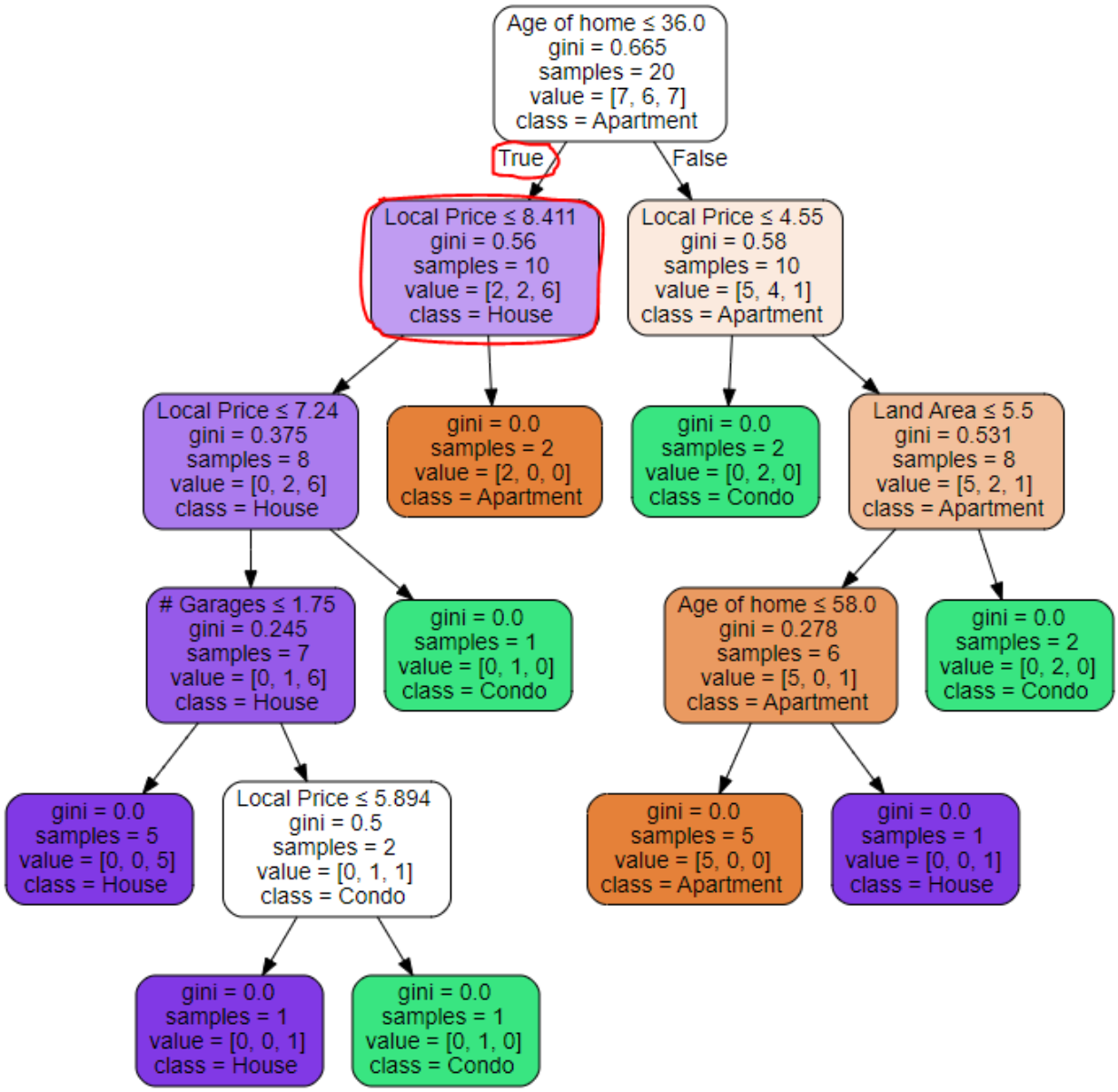
Restricting the depth of the tree or early stopping has a strong effect on the classifier performance because it prevents the model to not overfit to the training data, reduces variance, and introduces some bias towards tree.

4. Visualize the resulting tree. Perform the inference on this tree manually (i.e. show/trace the path taken towards classification) and provide a classification for the following example:

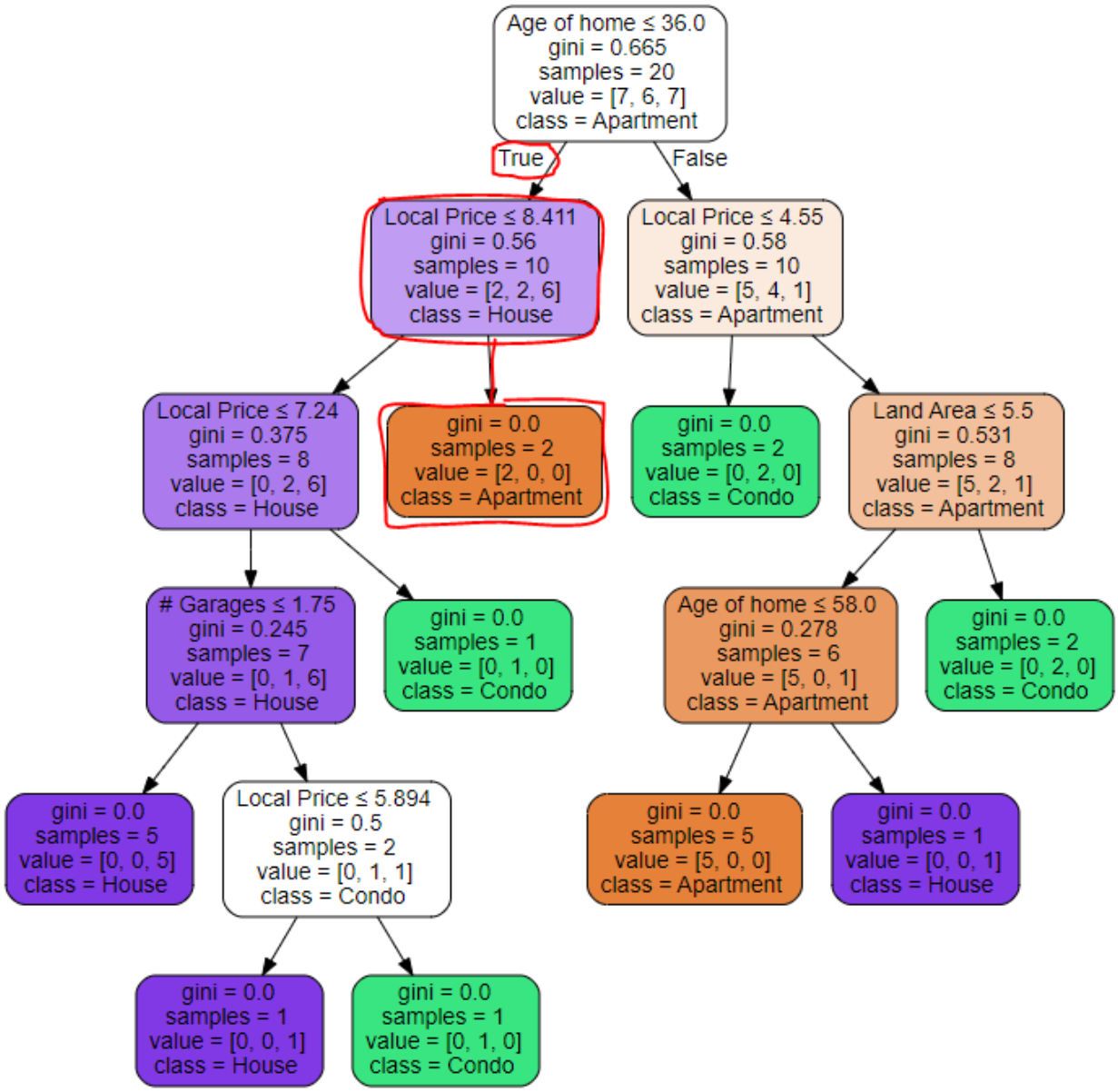
Local Price	9.0384
Bathrooms	1
Land Area	7.8
Living area	1.5
# Garages	1.5
# Rooms	7
# Bedrooms	3
Age of home	23


Age of home ≤ 36.0
gini = 0.665
samples = 20
value = [7, 6, 7]
class = Apartment
True
False
Local Price ≤ 8.411
gini = 0.56
samples = 10
value = [2, 2, 6]
class = House
Local Price ≤ 4.55
gini = 0.58
samples = 10
value = [5, 4, 1]
class = Apartment
Local Price ≤ 7.24
gini = 0.375
samples = 8
value = [0, 2, 6]
class = House
gini = 0.0
samples = 2
value = [2, 0, 0]
class = Apartment
gini = 0.0
samples = 2
value = [0, 2, 0]
class = Condo
Land Area ≤ 5.5
gini = 0.531
samples = 8
value = [5, 2, 1]
class = Apartment
# Garages ≤ 1.75
gini = 0.245
samples = 7
value = [0, 1, 6]
class = House
gini = 0.0
samples = 1
value = [0, 1, 0]
class = Condo
Age of home ≤ 58.0
gini = 0.278
samples = 6
value = [5, 0, 1]
class = Apartment
gini = 0.0
samples = 2
value = [0, 2, 0]
class = Condo
gini = 0.0
samples = 5
value = [0, 0, 5]
class = House
Local Price ≤ 5.894
gini = 0.5
samples = 2
value = [0, 1, 1]
class = Condo
gini = 0.0
samples = 5
value = [5, 0, 0]
class = Apartment
gini = 0.0
samples = 1
value = [0, 0, 1]
class = House
gini = 0.0
samples = 1
value = [0, 0, 1]
class = House
gini = 0.0
samples = 1
value = [0, 1, 0]
class = Condo

Age of home: 23, so move to left child.



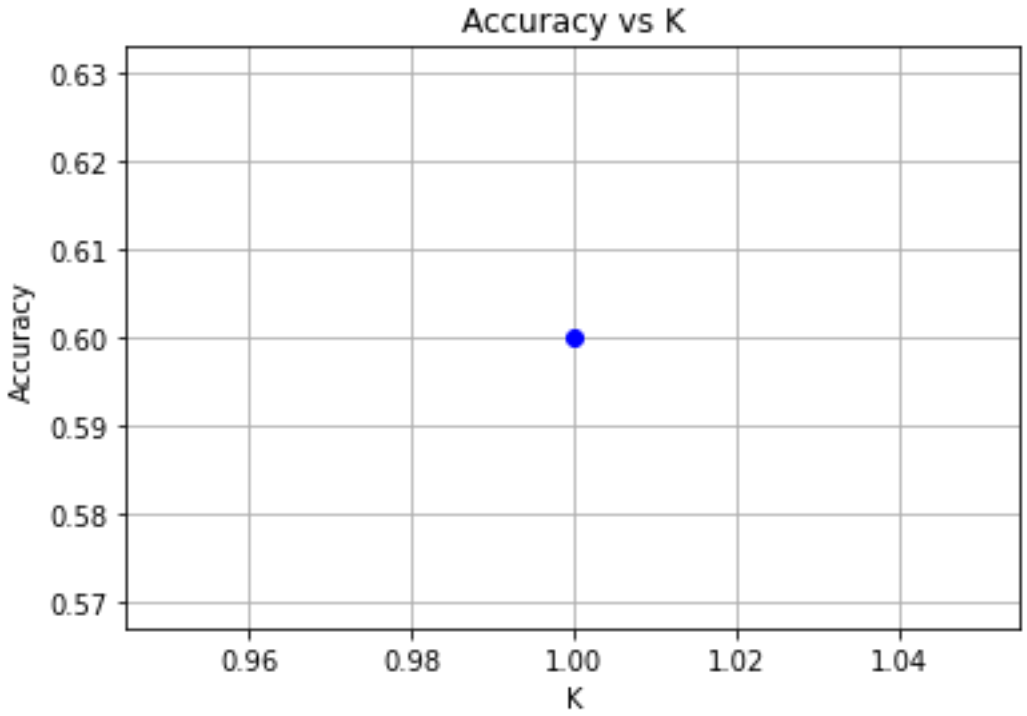
Local Price: 9.0384, so move to right child.



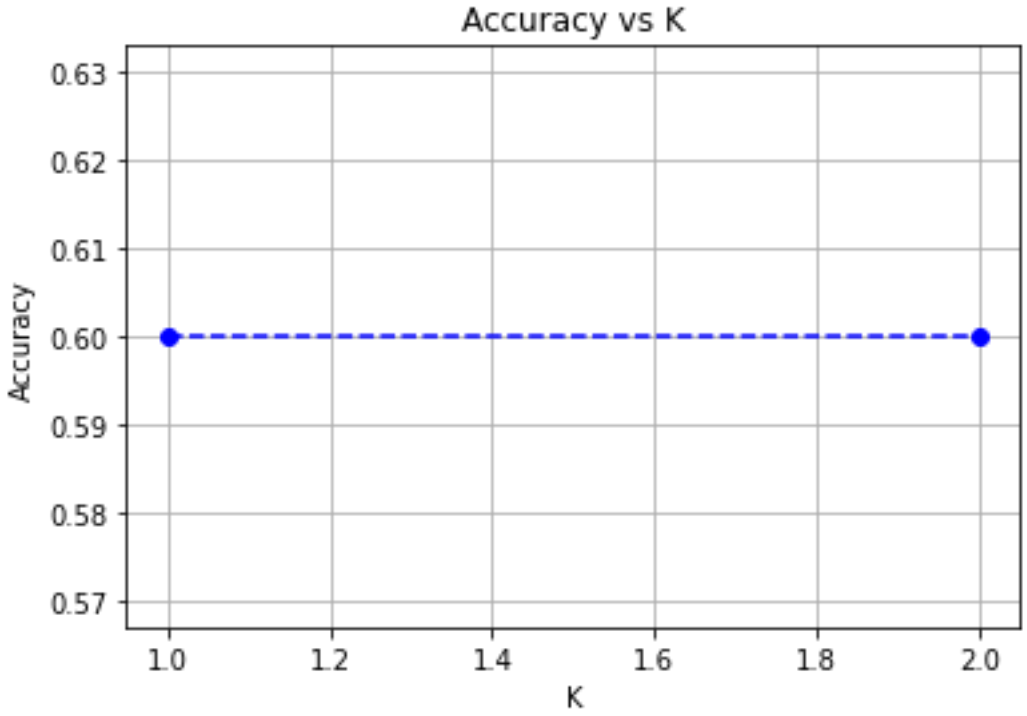
Therefore, the class of the given example is Apartment by above decision tree.

Question 3

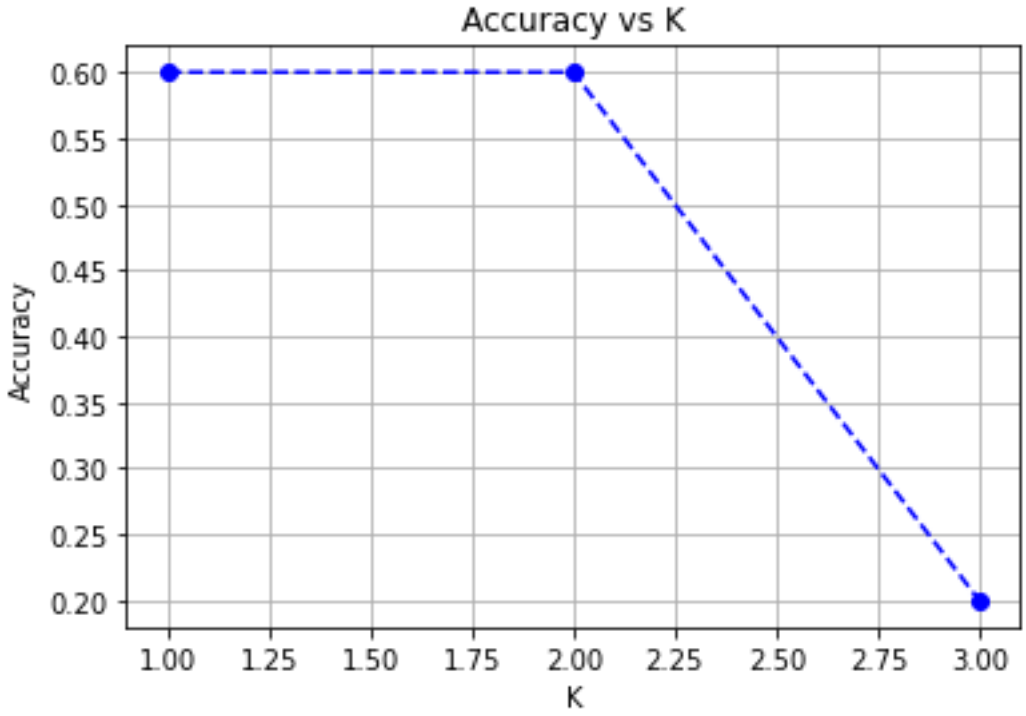
```
['Apartment' 'Apartment' 'Condo' 'Apartment' 'Apartment']
Accuracy of test of K=1: 0.6
```



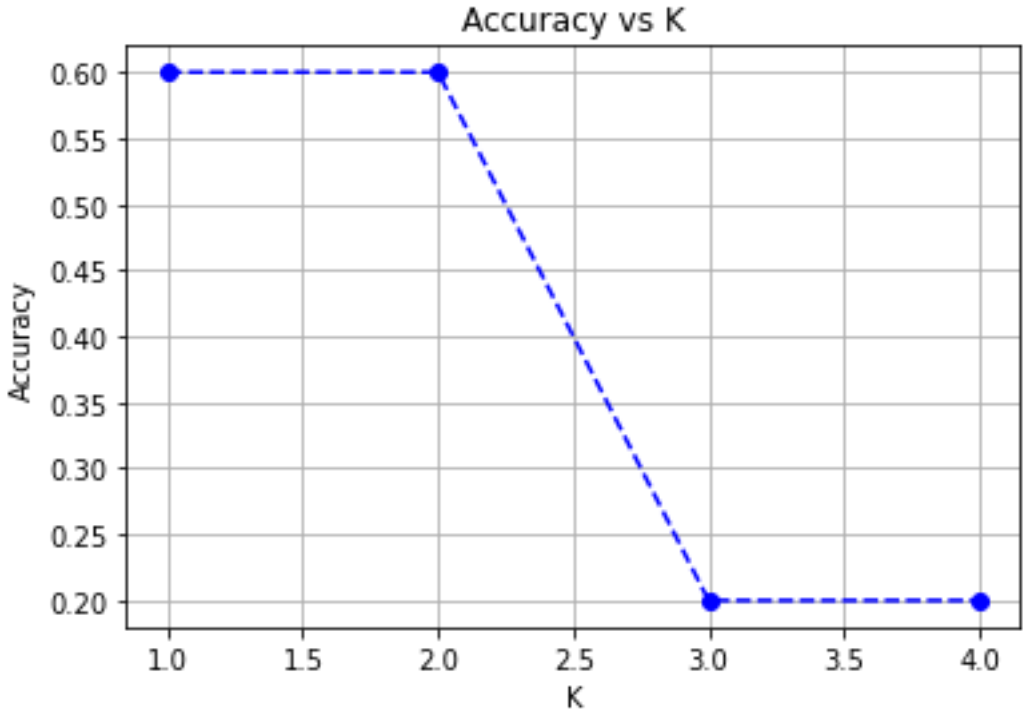

```
['Apartment' 'Apartment' 'Condo' 'Apartment' 'Apartment']
Accuracy of test of K=2: 0.6
```



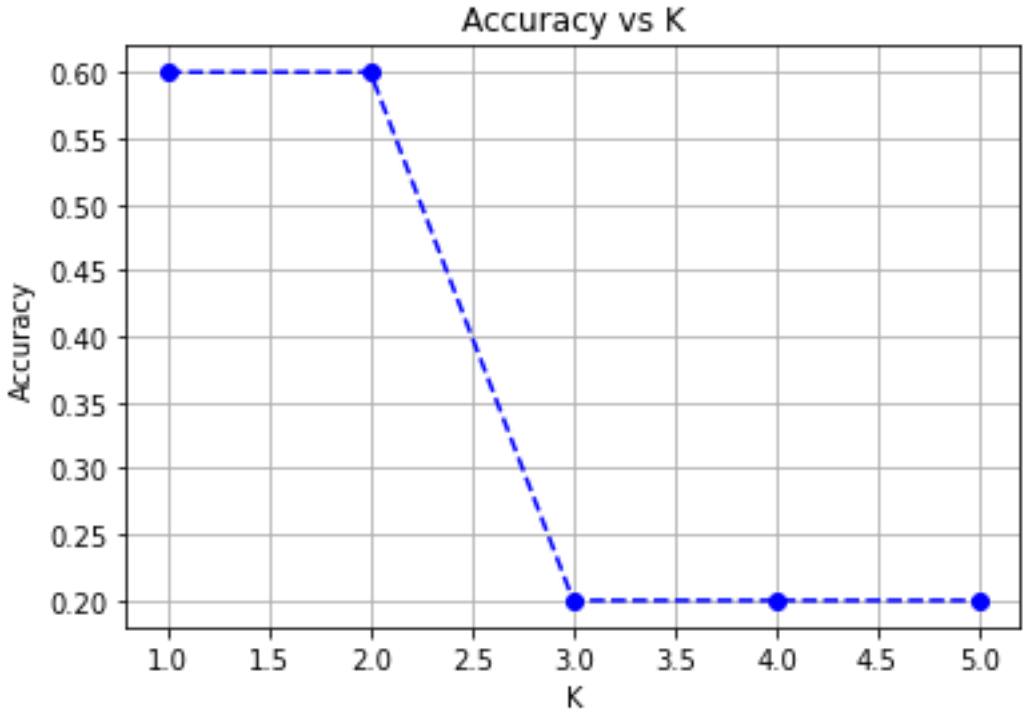
```
['Apartment' 'Apartment' 'Apartment' 'House' 'House']
Accuracy of test of K=3: 0.2
```




```
['Apartment' 'Apartment' 'Apartment' 'House' 'House']
Accuracy of test of K=4: 0.2
```



```
['Apartment' 'Apartment' 'Apartment' 'House' 'House']
Accuracy of test of K=5: 0.2
```



As the plot demonstrates, when the parameter nearest neighbor(K) increase, we can observe that the accuracy of the result decreases. This happens because the K-NN decision boundary will include much more data points that are irrelevant to it's neighbors.