Bathrooms Classifier

Aldo N. Cao Romero
June 2021

Use case

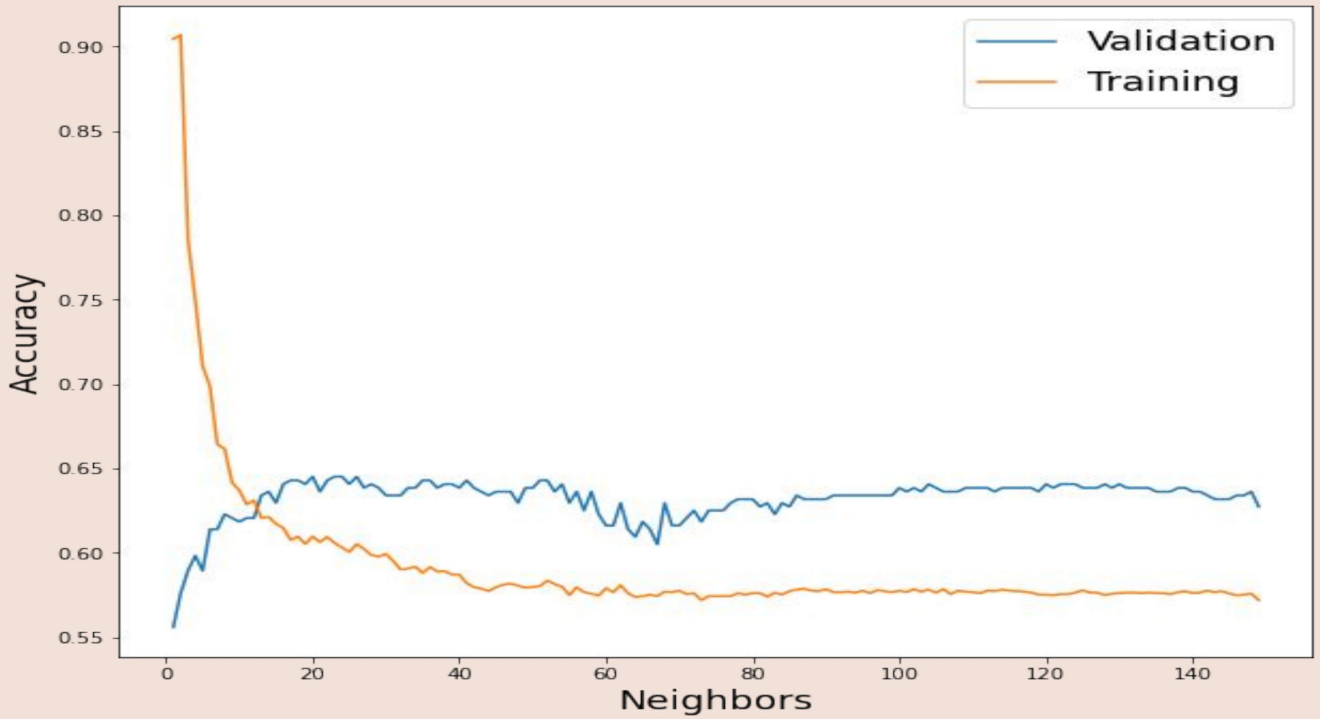
The creation of a model that classifies the houses by the number of bathrooms and, in this way, predict which houses can qualify for adding a new bathroom

We take the cleansed data of the raw house data, and took the variables we will use for the model

	sqrt_ft	bedrooms	bathrooms
0	7471.0	5	6
1	5333.0	4	4
2	6800.0	4	7
3	6622.0	4	6
4	7132.0	5	6
...	...	...	...
4479	2106.0	3	2
4480	3601.0	5	3
4481	2318.0	4	3
4482	3724.0	4	4
4483	4317.0	4	4

The model was trained using the KNN classifier along with the normalization:

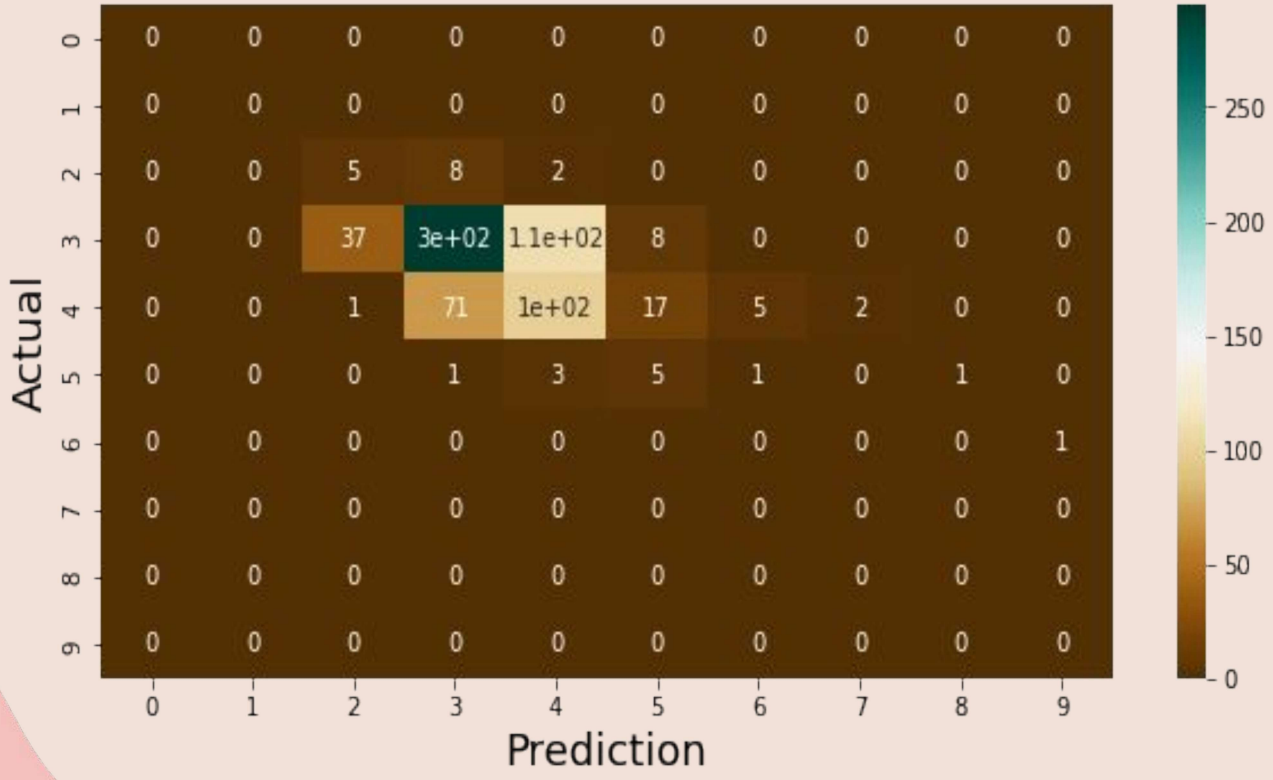
$$X = \frac{X - \min(X)}{\max(X) - \min(X)}.$$



The amount of neighbors that gives us the best accuracy for the model was 14, which corresponds to an accuracy of 62%

	Neighbors	accuracy_Train	accuracy_Val
10	11	0.620536	0.628605
11	12	0.620536	0.630984
12	13	0.633929	0.620577
13	14	0.636161	0.621172
14	15	0.629464	0.617009
15	16	0.640625	0.614630
16	17	0.642857	0.607493
17	18	0.642857	0.609575
18	19	0.640625	0.605114
19	20	0.645089	0.609575

Then, we test our model with the test data and obtain an accuracy of 60% and we get the following behavior:

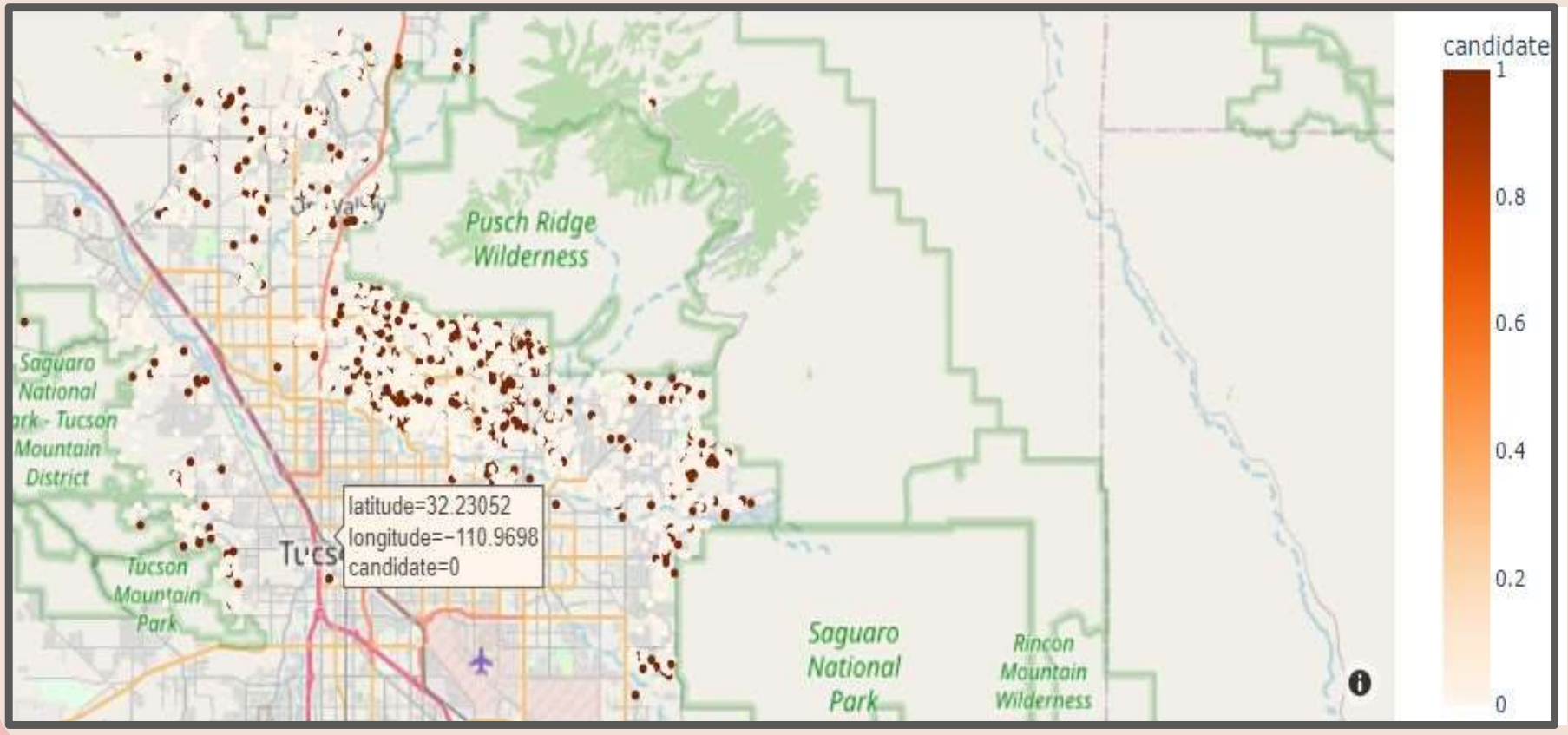


	Actual	prediction
0	3	3.0
1	3	3.0
2	3	4.0
3	2	3.0
4	3	3.0
5	3	3.0
6	4	3.0
7	3	3.0
8	2	3.0
9	3	3.0
10	3	3.0
11	3	4.0
12	3	3.0
13	3	3.0
14	2	2.0

It was taken the values that the model predicts larger than the actual value and we assume the house might have the possibility of adding a new bathroom. It was obtained a total of 734 candidates of adding a new bathroom

	sqrt_ft	bedrooms	bathrooms	predicted	candidate
0	7471.0	5	6	6.0	0
1	5333.0	4	4	5.0	1
2	6800.0	4	7	6.0	0
3	6622.0	4	6	6.0	0
4	7132.0	5	6	6.0	0

We locate in the map the houses that are possibly able to have a new bathroom



Conclusions

- We create a model that classifies the houses with a certain amount of bathrooms and find the optimal quantity of K-Neighbors

- By getting the confusion matrix, we were able to observe that the predictions differs only by one in many cases.

- We propose that, in the houses where the predicted value was larger than the actual value, those houses can be candidates of adding a new bathroom.

- We locate in the the map the houses that we can offer the possibility of adding a bathroom

Appendix

```
cand=[]
c=0
for i in range(len(Y)):
  if Y_Hat[i]>Y[i]:
    cand.append(1)
    c=c+1
  else:
    cand.append(0)
print('The number of houses that can be candidates is', c)
```

```
fig = px.scatter_mapbox(df,lat = 'latitude', lon = 'longitude', color = 'candidate',
                        center=dict(lon=-110.9, lat=32.3),
                       zoom = 9, mapbox_style = 'open-street-map',color_continuous_scale="oranges")
fig
```