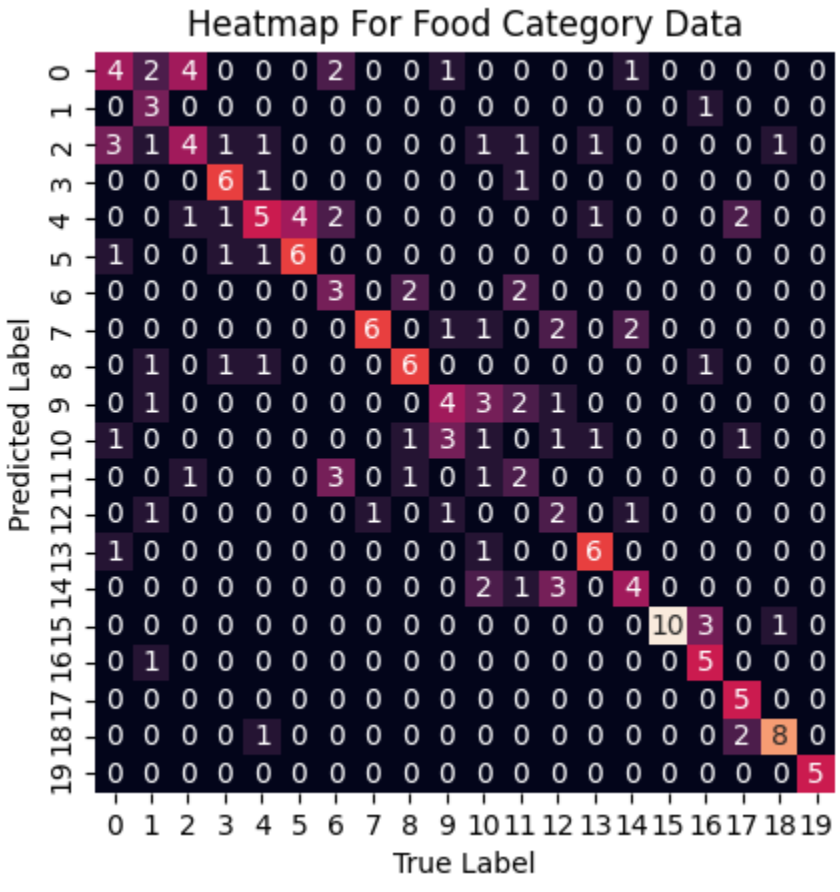
Homework Three
Caroline Smith
23 April 2023

Model Used: Random Forest Classifier

RFC Independent Classification Report

	precision	recall	f1-score	support
1	0.31	0.50	0.38	10
2	0.67	0.40	0.50	10
3	0.38	0.50	0.43	10
4	0.56	0.50	0.53	10
5	0.50	0.50	0.50	10
6	0.60	0.90	0.72	10
7	0.33	0.30	0.32	10
8	0.60	0.86	0.71	7
9	0.88	0.70	0.78	10
10	0.50	0.40	0.44	10
11	0.18	0.20	0.19	10
12	0.20	0.11	0.14	9
13	0.57	0.44	0.50	9
14	0.62	0.56	0.59	9
15	0.36	0.50	0.42	8
16	0.77	1.00	0.87	10
17	0.67	0.40	0.50	10
18	0.83	0.50	0.62	10
19	0.64	0.70	0.67	10
20	1.00	1.00	1.00	5
accuracy			0.53	187
macro avg	0.56	0.55	0.54	187
weighted avg	0.55	0.53	0.53	187



Misclassification: It was very difficult to get a high accuracy score on this dataset. In the example above, I used the Random Forest Classifier, which reached a peak of 56%. I also tested KNN and Decision Tree, but both could not break 50% accuracy. It appears that the data is misclassified, in my opinion, because of a few reasons: first of all, the sample data for each type of food are very similar to each other, so it was difficult to differentiate. This would lead to misclassification as there aren't many defining features in the dataset. Another reason why the data could be misclassified is that there aren't enough of each food type in the data; the data set only had 933 entries, which, considering it has 20 different categories, really is not very large. Thus, the model doesn't have much to work off of.