What changed

In my project proposal, I was going to detect sodium content based on ingredient names. I had to change the dataset to this one <https://www.kaggle.com/datasets/shuyangli94/food-com-recipes-and-user-interactions> .

In addition, I realized that this was a regression task and not a classification task.

When I first ran the first model, it had an R^2 score of .0006, which is rather bad. I realized that having one feature be the only item that informs the decision of the model may be a bad idea. I added features that potentially could be valid, such as Total Fat as well to help with this, and ended up with a better R^2 score.

Metrics

Ridge Regressor, unoptimized, R^2: 0.1085832212602621

Ridge Regressor, optimized, R^2: 0.17840529028848973

Ridge Regressor, unoptimized, Mean Squared Error: 5626.903748728214

Ridge Regressor, optimized, Mean Squared Error: 5186.164835877213

Decision Tree, unoptimized, Mean Squared Error: 8945.683617086921

Decision Tree, optimized, Mean Squared Error: 6392.548619901867

Linear Regression, unoptimized, Mean Squared Error: 9.616381550995421e+18

Linear Regression, optimized, Mean Squared Error: 5942.265297720374

Ridge seemed to perform the best with the lowest mean squared error, even the unoptimized seemed to do better compared to other optimized models with the lowest mean square error. The Linear Regression model unoptimized performed the worst, but optimized performed decently well. Overall, none of the models performed very well.