

Kamryn Parker

Professor Kennington

CS 334

Homework 5

Unfortunately, my homework assignment was unable to gain an accuracy better than 5% like I wanted to. I was only able to increase my model by about 4-4.5%. However, I am glad I was able to test and try different scaling and model improvement methods. I think one of the main reasons this model may have failed was lack of more data. I think the way I collected data may have been more time consuming than other methods. I used a website with month by month data so I had to copy and paste the data into a CSV format before I loaded it into python.

One thing that really helped my model was including data about rain in Seattle the day before the day it rained in Boise. This helped greatly in improving my model. A change I may make to help more would be to add maybe another city and have them 2 days prior to the current tested day in Boise.

One hot encoding was something I tried but I was not very successful with it since I didn't have any categorical data in my data set. I ended up making some with seasons data but I easily changed it to a numerical value instead which ended up helping my model. I think if I had more categorical data in my data set like wind direction or maybe what the day looks like (cloudy, sunny, etc.) would have helped more with making coded data that way.

The last thing I wanted to try but was not able to test out was ways to fix class imbalance. I think if I was able to make a distinction of which classes were imbalanced and how I could fix them. I think balancing the class would have been beneficial because it would balance the overall data set not just scaling each column individually.

