Kaggle Team Name: Feeling

Kaggle Email: hfl22@uw.edu

What did you try and why? What was your final choice of model and why?

At the first glance of this data set, I immediately identified that it is a classification problem. Therefore, I tried the decision tree model first. I followed the steps from the assignment 4, and was hesitated to balance the amount of data between > 50K and <= 50K since I found out that the number of data for <= 50K is far more than the other. This balanced training data set resulted to get a 77 accuracy on Kaggle. I then chose to not balance the training data set and followed the same steps for the rest, then this decision tree model resulted in around 85. I then chose to use Adaboost model and I got 86.885 in accuracy. However, after I googled for other classification models that I chose gradient boosting for my final choice. It gave an 87.346 in accuracy.

What were your results: What do you think your final accuracy will be and what gives you confidence in that prediction?

I used GridSearchCV to find the fittest hyper parameters for the GradientBoosting Classifier. Therefore I am confident with the final accuracy.

Are there any ethical considerations we should think about before acting on that plan?

I think Kroger should not focus on merely on the aspect of using the model for benefitting the company. Indeed the model could identify the locations where richer families and individuals reside, but this might cause the unfairness to the other communities with poorer income. As a grocery company, its focus should more focus on bring more communities with equally accessible resources and bringing conveniences to all communities if possible.