Walmart sells prediction

Dhiraj Bankar  
DSC680 - Summer 2021  
<https://github.com/dhirajbankar/DSC680>

# Any surprises from your domain from these data?

I choose retail as a domain. I had very close relation with retail domain as I worked for 5 years in past with Sears. This dataset is very clean so I don’t even need to do any cleaning operations. Only 8 attributes are there. If I talk about the factors retail analysis is mostly the holiday seasons and the offers by the retailer. We don’t have offer lever information but we have dates which I want analyze.

# The dataset is what you thought it was?

I would say there is not perfect dataset and there is not 100% guaranty of prediction model which I am working on. I feel there should be some additional attributes would have been there that could have made model much more reliable and give correct results.

1. Cleanliness of store
2. Behavior of staff

# Have you had to adjust your approach or research questions?

For now model will perform good an I have enough data to create model and I don’t think there will be any problems in current approach.

Except As I stated above. The analysis question within the current dataset is all concerning characteristic the attributes that are a lot of possible to predict the Walmart sell a given store. The end result are going to be very useful and includes a heap of sensible applications within the US.

# Is your method working?

Yes as per train and test data the model is performing very well.

I am planning to take a unique approach in terms of selecting the strategy. In alternative words- rather than simply selecting one classification algorithmic rule to predict a disease, I will be able to be employing a combination of various algorithms in predicting the end result. As an example, since this is often a classification drawback, I’m planning to apply completely different classification algorithms like logistic regression, Polynomial Features, random forests etc.

The main reason for taking this approach is just to get a global picture of how different algorithms analyze the same dataset and go further to understand on how to handle the results when the outcome from different algorithms is vastly different. Do we need to pick and choose or do we perform a logical elimination of one algorithm versus the other algorithm?

# What challenges are you having?

I don’t think the give model will be overfit as I have enough data to tune the model.However, I still decided to go with the dataset as I want to learn different ways by which I can work with a smaller dataset and still get a better accuracy. The other issue I anticipate is to deal with different accuracies I get from models and how do I choose one over the other when these individual models give me vastly different accuracies.