Featured Prediction Competition

Corporación Favorita Grocery Sales Forecasting

Can you accurately predict sales for a large grocery chain?

$30,000
Prize Money

Corporación Favorita · 1,697 teams · 16 hours ago

Louis T.
13th place

13th place solution

posted in Corporación Favorita Grocery Sales Forecasting 12 hours ago

15

Feature Engineering

After reading a couple of solutions posted, I think this is my biggest trump card compared to the others. I started by thinking about my own grocery shopping habit.

- I tends to shop on a weekend. (timing)
- I will stock up if things I want is on promotion. (on promotion)
- If I already have enough of something at home, it's much less likely for me to buy more even if it is on discount. (historical sales)
- If I have enough beef at home, it's much less likely for me to buy more any other type of meat as well (sales by class)

The key insight is that most people maintain a certain level of stock for a particular group of goods at home, not individual goods. I think most people missed this. Or in classic economic theory, there is a substitution effect among goods. I believe including these class-based features boosted my model by a large margin. I added these features at the very early stage so I don't have a good estimate on how much the impact this.

My formal education is in economics which definitely helped, but at the same time, I think these features are pretty generic with limited domain expertise

involved.

Model

I used sliding window approach like everyone else.

My base model is a simple DNN. started with a batch norm layer, then followed by 6 dense layers with a .3 dropout layer between every two. Activation with 'elu' with the only exception of last layer of 'relu' to force the output to be positive.

I started with train the model using the last 2 years data and after a lot of tuning the best public score, I managed to achieve with this 0.510.

I stuck at 0.510 for a long time then because I have been training with RMSLE without weighting, and my intuition also tells me, the behaviour for perishable and non-perishable might be different, I decided to train the perishable and non-perishable separately. Doing this, I managed to improve the public score by another .001.

My final model is an ensemble of the above two variants and AhmetErdem's public kernel. this improved the public score to 0.507

Validation

Initially, I used the last 16 days across all series in train dataset. And the performance has been quite consistent compared to the public score.

However, after I started training the perishable and non-perishable, I noticed further improving my local cv actually decreased my public score so I added another 16 days to my validation set, which turns out to be quite helpful in training the perishable goods.

Options

Comments (5) Sort by Hotness

Click here to enter a comment...

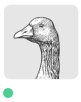
ZedYeung • (326th in this Competition) • 12 hours ago • Options • Reply 1

I wonder what is the insight that traing separately according to perishable is helpful but traing them together with perishable as categorical is useless.

Louis T. • (13th in this Competition) • 11 hours ago • Options • Reply 1

My intuition is that perishable goods like food might have a very different pattern than non-perishable goods. My economic training tells me that people are probably more sensitive to price (hence promotion) for perishable good as they observed the price more frequently. A lot of people can probably tell how much was the meat they brought last weekend but wouldn't be able to remember how much was the soap they brought last time. Also, for non-perishable goods, I used a longer look back horizon (64 days instead of 28).

That being said, I went back to check the private score of those two, my single model approach (perishable as a feature) actually performed a little better (both are 0.517) on private, but 0.002 less on the public. So I might actually over-fitted a little. However, I still believe there might be some merit in training them separately, but validation might be tricky.

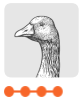


ZedYeung • (326th in this Competition) • 10 hours ago • Options • Reply 0

My solution is somewhat similar with yours-- NN stuck at 0.510 using the last 2 years data. And use last 16 days and 16 days since 7.26 as same as Ceshine Lee's kernel. But I have not try perishable feature since this feature get low importance score in lgb model. I wonder have you preprocessed your data? I had removed those data with too many 0. Maybe that is not a good decesion.

FabSchreiber • (188th in this Competition) • 9 hours ago • Options • Reply 0

Congrats, Luis and thanks for sharing your solution. Very interesting. I was trying to look into DNNs for this competition, but did get them working. I think it would be great if you could enlighten others with sharing your code, if you have the time. :-)



Evdilos_Ikaria • (653rd in this Competition) • 8 hours ago • Options • Reply 0

Very good job LuisT. This feature engineering based on your habits is super!