Winning Model Documentation Guidelines

A. MODEL SUMMARY

A1. Background on you/your team

- Competition Name: Predict Future Sales
- Team Name: Alexander Deineha
- Public Leaderboard Score: 0.91203
- Public Leaderboard Place: 999
- Name: Alexander Deineha
- Location: Nice, France
- Email: alexanderdeineha@gmail.com

A2. Background on you/your team

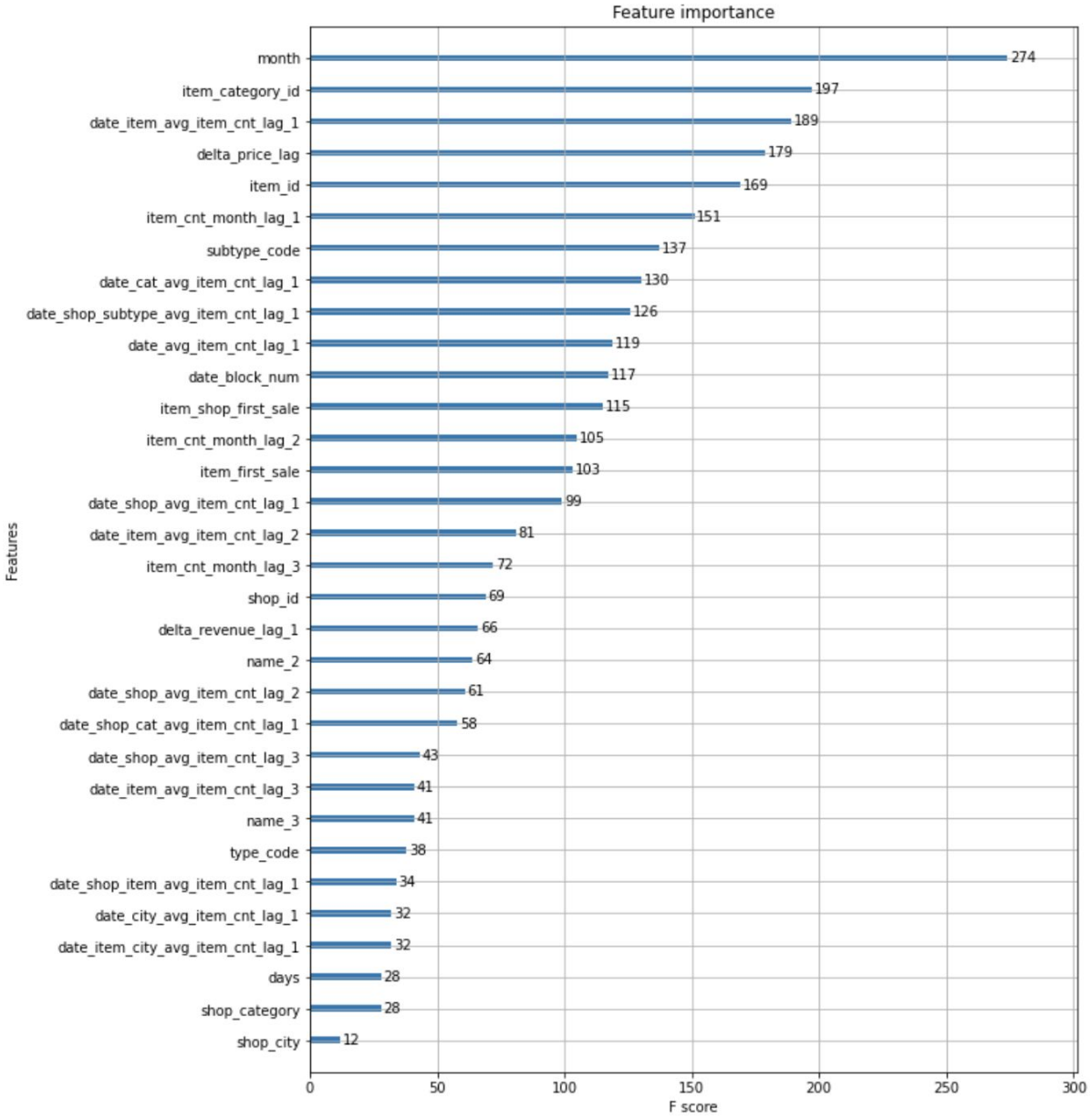
- Currently I am a student of Master program in computer science, I have 2 years of experience of working as ML engineer.
- I have participated in several Kaggle competitions.
- This competition has interesting task and data.
- I spend about a week for this competition.

A3. Summary

- I used Neural Network, XGBoost and ElasticNet
- The most important features are 'date_block_num', 'shop_id', 'item_id'
- It takes about 2 hours to train models

A4. Features Selection / Engineering

- Here is plot of features importance for XGBoost



- I added features by groups and monitored the performance
- I have done only Label Encoding and standard scaling of numeric features for NN.
- I have found that lags and mean encoding are important for model.
- I haven't used any external data.

A5. Training Method(s)

- I used classical training methods for XGBoost and NN
- I stacked models with ElasticNet.
- Different models were weighted by ElasticNet

A6. Interesting findings

- I think most important part is model ensembling.
- Most of other people used only one model.
- For me was interesting the way shops and items for test set were selected.

A7. Simple Features and Methods

- It is possible to get 90-95% of final performance using this features: date_block_num, shop_id, item_id, shop_city, item_category_id, month, delta_price_lag, date_shop_item_avg_item_cnt_lag_1, item_cnt_month_lag_1, item_cnt_month_lag_2
- Most important model is XGBoost
- Simplified model score is 0.97960
- One model with 10 features score is 0.97960

A8. Model Execution Time

- It takes 2 hours to train all the models
- It takes about 20 min to generate predictions using my model.
- It takes 20 min to train the simplified model (referenced in section A6).
- It takes 5min to generate predictions from the simplified model.

A9. References

Citations to references, websites, blog posts, and external sources of information where appropriate.

- For this competition I used Neural Network and XGBoost
- They were validated on date block number 33 and trained on previous data
- Further models were ensembled using ElasticNet
- ElasticNet was trained on block number 33