

## A1. Background on you/your team

- Competition Name: Predict Future Sales
- Team Name: SCML
- Private Leaderboard Score: 0.915642
- Private Leaderboard Place: 2460
- 

[For each team member]

- Name: Hassan Eldeeb
- Location: Tartu, Estonia
- Email: [ha.eldeeb@gmail.com](mailto:ha.eldeeb@gmail.com)

## A2. Background on you/your team

- What is your academic/professional background?
  - PhD student
- Did you have any prior experience that helped you succeed in this competition?
  - yes
- What made you decide to enter this competition?
  - Participation in a coursera course
- How much time did you spend on the competition?
  - Two days

## A3. Summary

Two level ensemble model is used. In the first meta model of level, I have ensembled XGBoost, Random Forest, and KNN. In the second meta model, I have used XGBoost again on different features. Finally, the output of the two meta models have been averaged.

## A4. Features Selection / Engineering

- What were the most important features?
  - Mean encoded features and window rolling based features (see the notebook)

- How did you select features?
  - Since the number of raw features are very small, I kept all the engineered features
- Did you make any important feature transformations?
  - Yes, see the notebook for more details
- Did you find any interesting interactions between features?
  - Yes, grouping the shop\_id and product\_id then aggregating the target column improved the final score.
- Did you use external data? (if permitted)
  - no

## A5. Training Method(s)

- What training methods did you use?
  - Two-level ensembling
- Did you ensemble the models?
  - yes
- If you did ensemble, how did you weight the different models?
  - I have used to ensemble models. The base models of the first ensemble have been combined using a linear model (stacking), while the base models of the second ensemble have been combined using weighted average.

## A6. Interesting findings

- What was the most important trick you used?
  - Adding multi level ensemble
- What do you think set you apart from others in the competition?
  - Ensemble of different ensembles.
- Did you find any interesting relationships in the data that don't fit in the sections above?
  - No

## A7. Simple Features and Methods

- Is there a subset of features that would get 90-95% of your final performance? Which features? \*
  - I have used few features, so I do not think this hypothesis is valid
- What model that was most important? \*
  - XGBoost

## A8. Model Execution Time

- How long does it take to train your model?
  - 2 hours (see the machine specs on the readme file)
- How long does it take to generate predictions using your model?
  - seconds