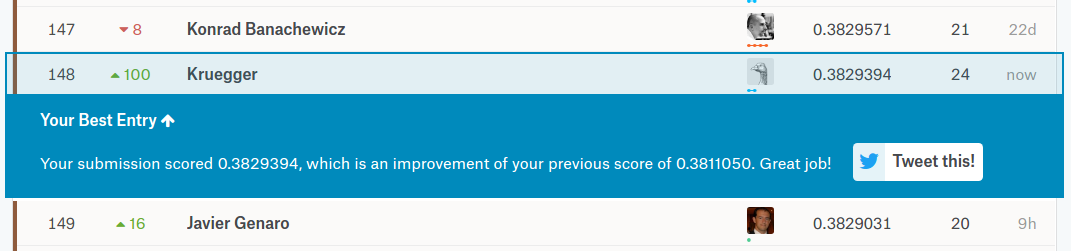
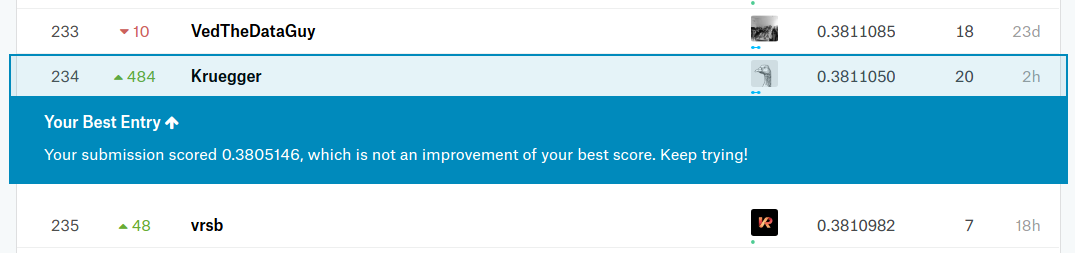
How to lift up to Bronze area in a week

posted in [Instacart Market Basket Analysis](https://www.kaggle.com/c/instacart-market-basket-analysis" \o "Instacart Market Basket Analysis) 2 days ago

https://www.kaggle.com/static/images/medals/discussion/goldL@1x.png

47



Hello!

I am novice kaggler and writing this post to help another novices not to give up competition even if you think that you are out of ideas and knowledge.

At first I want to say 'Thank you' to all people who shared their thoughts and expertize - your information and approaches very helpful and useful. Small list of these guys (but not compelte ) here:

@AlphaMeow @SVJ24 @raddar @Fabienvs @paulantoine @Li Li ... and many many others ...

It's my second serious competition here and second post on kaggle ) So I am not pretend to be experienced people, but want to share some thoughts and information compiled from discussions and kernel analysis.

Main topic - you can do it! Really, as mentioned early in some discussion, there is no Magic feature or leaking, just EDA/Feature engineering and some luck ) If you are sticking on 0.38 LB by some of public kernels - don't give up, just try understand what you can do and do it! )

Some basic ideas about this competition compiled from discussions:

* basic model (presented in high voted kernels) is binary classification with logloss over (order, product) pair and 1/0 as target from 'reordered' column on train dataset. You can also try bayesian/rnn/..., but this model is simple and well done.
* Correct CV!!! Mercedes show us that we can't underestimate importance of correct CV scheme. In this competition it is easy - just create folds based on disjointed user\_id.
* Treshold to convert probability to 1/0 is the key! Don't use default 0.5 - start with 0.2 and tune it on CV. The more advanced idea is to use different treshold for different orders. I give a link to some topics later.
* Feature engineering. In this competition it is the most important part. You have to read the book from post @Rodolfo Lomascolo ("Repeat Buyer Prediction for E-Commerce") <https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/36411>

Thats all to jump over 0.38. Really, just to try it )

And some advanced to improve your result:

* try to predict None as separate product in the order or try to predict basket size of the user - to implement F1 expectation scheme.
* add bayesian aproaches to your model
* ... here the place for your imagination ... )

And some links to the discussion i mention interested:

Data understanding

<https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/33128><https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/33448>

CV

<https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/36493>

General

<https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/35048>

What does 'Reordered' mean

<https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/33211>

NONE handling:

<https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/36134>

<https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/35716>

F1Score:

<https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/36544><https://www.kaggle.com/aikinogard/python-f1-score-function>

Features:

<https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/35468>

Model selection

<https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/33131>

Kernels:

<https://www.kaggle.com/fabienvs/instacart-xgboost-starter-lb-0-3791/code><https://www.kaggle.com/paulantoine/light-gbm-benchmark-0-3692><https://www.kaggle.com/nickycan/lb-0-3805009-python-edition><http://kelsh.tech/blog/2017/06/21/analytical-approach-to-kaggle-instacart-competetion/><http://blog.stylemyimage.com/post/kaggle-instacart-market-basket-analysis-competition-solution-with-c-and-vowpal-wabbit>

Advanced )

<https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/36312>

Thats all. Thank you for reading and happy kaggling!!!

@kruegger.

P.S. you can upvote if you want )