

# otto group

# Completed • \$10,000 • 3,514 teams

# **Otto Group Product Classification Challenge**

Tue 17 Mar 2015 - Mon 18 May 2015 (8 months ago)

Dashboard	▼	Competition Forum	
All Forums » Otto	Group Product Class	ification Challenge	Search
Prev <b>《</b> Topic		41599 via TSNE, meta bagging	Next Topic <b>&gt;&gt;&gt;</b>
		Start Watching View all posts	
		<b>1</b> 2	>

84

As promised, this will get you 0.41599 public via T-SNE and meta bagging. The key here is meta bagging. And proper representation of the data with various transforms.

I'll post an update to a potentially higher scoring single model with meta bagging neural networks where the first stage uses the mean of rbf svm, rf, and gbm.

1 Attachment —

ottoHomeBagG4.R (2.78 KB)

#1 | Posted 8 months ago



Mike Kim

▲ Th

That's awesome. Thanks Mike!

0

#2 | Posted 8 months ago



**Phil Culliton** 



Here's the meta bagged ANN with Lasagne and sklearn. I'm not sure what the score is. You can do slightly better using mean over geomean and doing more epochs.

1 Attachment —

lasagneSeed21.py (4.16 KB)

#3 | Posted 8 months ag



Mike Kim

Nice. - appreciate u throwing the code up



#4 | Posted 8 months ago



T. Scharf



Does the transform you do in first step



(Look like a logit on square root ) of the initial matrix -

Does have a name or anything?

Congrats on awesome finish

#5 | Posted 8 months ago



T. Scharf



Great work, Mike. Could you give more details on what you mean "meta bagging"? Thanks!



 $\Delta$ 

#6 | Posted 8 months ago



cloudfarm



Thanks! This is something new to me



From the python example, it seems he is training on part of the dataset with a SVM, random forsest, and GBM, then using the output of these models to train a bag of 50 neural networks.

Clarification: the neural networks are trained on the original features with the geometric mean of the other classifiers added as an extra feature.

#7 | Posted 8 months ago | Edited 8 months ago



Devin



Really cool, thanks!



#8 | Posted 8 months ago



Nicholas Guttenberg



Appreciate the sharing Mike. Congrats!



#9 | Posted 8 months ago



Patrick Chan



Great work!



#10 | Posted 8 months ago



Jiming Ye



That is awesome. I learned something new today.



Thanks!

#11 | Posted 8 months ago





Great work Mike! Really cool ideas here. Thanks!



#12 | Posted 8 months ago



A Morales



It's really a beautiful way to do ensembling. Seems I have a lot of new things to learn.



Could you explain more about this, why you choose the first 10000 training data?

#13 | Posted 8 months ago



Jiming Ye



#### Cool! Thanks Mike!





fyears



#### Jiming Ye wrote:

It's really a beautiful way to do ensembling. Seems I have a lot of new things to learn.

Could you explain more about this, why you choose the first 10000 training data?

Because SVM does not scale too well with too many rows. Also it's possible to over train the first stage estimators. The intuition for tuning this meta bag is somewhat strange. I did try more and less data for the first stage. In the R Code I just use out of bag which means the exact number is variable depending on run.

I did not do a lot of tuning here. So I have no idea what the possibilities are for scores using the framework. The tuning does not correspond to "common sense" for me.

#15 | Posted 8 months ago



### Mike Kim



Thanks for sharing the code.



This competition has been a learning experience and now I am learning more:)

#16 | Posted 8 months ago



VMK



## Mike Kim wrote:

#### Jiming Ye wrote:

It's really a beautiful way to do ensembling. Seems I have a lot of new things to learn.

Could you explain more about this, why you choose the first 10000 training data?

Because SVM does not scale too well with too many rows. Also it's possible to over train the first stage estimators. The intuition for tuning this meta bag is somewhat strange. I did try more and less data for the

first stage. In the R Code I just use out of bag which means the exact number is variable depending on run.

I did not do a lot of tuning here. So I have no idea what the possibilities are for scores using the framework. The tuning does not correspond to "common sense" for me.

#### Thanks!

#17 | Posted 8 months ag



Jiming Ye



Very interesting indeed, thanks for sharing. I wish there was a way to upvote twice;)



#18 | Posted 8 months ago



Matt



Thanks Mike!



#10 | Dostod 9 months o



**Abhishek** 



Amazing insight - thanks Mike

0

#20 | D--+--| 0 -----+|-----

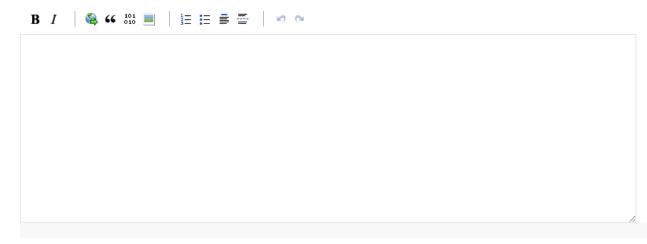


Run2

1 2

2





+ Add attachment(s) ...

Post Reply

☑ Email me when someone replies

Start Watching

View all posts

« Back to forum



© 2016 Kaggle Inc

About Our Team Careers Terms Privacy Contact/Support





