



# BrainWaves - Machine Learning

Oct 16, 2016, 08:00 AM IST - Oct 28, 2016, 03:05 PM IST

[INSTRUCTIONS](#)[PROBLEMS](#)[SUBMISSIONS](#)[LEADERBOARD](#)[ANALYTICS](#)[JUDGE](#)[← Problems / Predict the growth](#)

## Predict the growth

Max. Marks: 3

This problem is no longer available for practice. Apology for any inconvenience!

World is changing fast. We are experiencing a technological revolution as big as industrial revolution. One of the major drivers of this revolution is Machine learning. Banking and financial services are at forefront of digital revolution and are using machine learning in their businesses.

The problem you will be working on is related to trading in financial markets and it is quite challenging. You will be clustering financial assets and building forecasting model which are the initial steps needed to build a machine learning based quantitative trading strategy. The common problem in trading is to forecast a variable based on other variables. Since financial data is time related one has to use recent past information to capture all the dynamics. We have divided the problem into two parts. First problem is of clustering which is needed to understand the data and the second problem is of prediction/classification.

Following is the second part of the problem. **Given the time series data of 100 financial assets which are correlated, you have to build a model to forecast return of the financial asset from which Y is derived. You can also use the clustering info from the first problem as needed.**

### Data

The dataset is common for both the problems. The training file(train.csv) contains the time-series data. Below is the relevant explanation of the different columns.

Column	Explanation
Time	Integer representing the time factor
X1..X100	Stock prices belonging to various industries
Y	Y is forecast value of the financial variable which we are interested in. Y has +1 if the underlying asset goes up in the future and it has -1 if it goes down.

**Note:** The problem is to forecast return of financial asset from which Y is derived. The accuracy of '\*' models will be relatively low (closer to 50%). ?

[Download Data](#)

## Submission

You have to submit a file containing your predictions in the following format.

```
Time,Y
3001,1
3002,-1
3003,1
3004,-1
3005,1
```

## Evaluation

Submissions will be assigned scores based on classification accuracy.

$$\text{score} = 6 * \text{abs}(\text{accuracy} - 0.5)$$

Since this a binary classification problem, according to the above function, whether your accuracy is 0.3 or 0.7 does not matter. If you get 0.3, you could just flip your predictions and get an accuracy of 0.7.

**Note:** Your submission will be assigned score only on 50% of the data during the live contest phase. Complete evaluation will be done after the contest has finished.

**COMMENTS (68)** SORT BY: **Relevance**▼

Join Discussion...

Cancel

Post

**Nimesh Jha** 2 years ago

How do you know where you stand. I mean where do we see ranking leaderboard ?

▲ 6 votes ● Reply ● Message ● Permalink

**Thanish Batcha** 2 years ago

Can someone explain me what does the accuracy and the max.score in the home page of the contest means ? I have made the submission and I was able to see the score. But I am not able to find a leaderboard and see the score of other competitors. What is the structure of the challenge ?

▲ 0 votes ● Reply ● Message ● Permalink

**Thanish Batcha**  Edited 2 years ago

Okay after careful analysis I was able to figure out that accuracy column means the accepted submission out of the overall submissions made by the users. So at this time asset clustering has 13/14 and predict the growth has 10/16 which mean 13 out of 14 submissions were successful without any errors and similar way 10/16 were successful. The failed submissions were due to some runtime error. But still I am not sure why should this column be named "Accuracy".

Also the max .score column means the max you can score in that task. For example the evaluation metric for predict the growth is  $6 * \text{abs}(\text{accuracy} - 0.5)$  so if you make a 100% correct classification it is going to be  $6 * \text{abs}(1 - 0.5)$  which eventually is 3.

But still I am unable to find the leaderboard. If someone could help me with that it would be great. Thanks

▲ 1 vote ● Reply ● Message ● Permalink



**binga** 2 years ago

Well, in the second task, you could score 3 even if your classification accuracy is 0 !

▲ 2 votes ● Reply ● Message ● Permalink



**Chetan Gupta** 2 years ago

In binary classification 0% accuracy is just the class labels flipped.

▲ 0 votes ● Reply ● Message ● Permalink



**binga** 2 years ago

True. Typically, in competitions the evaluation metrics are monotonous (increasing / decreasing). Take accuracy, logloss or auc. In this one, that wasn't the case. Anyways, this works too! Thank you.

▲ 0 votes ● Reply ● Message ● Permalink



**Sharathkumar Anbu** Edited 2 years ago

Hi,

One correction. They have already mentioned that accuracy of the model is close to 50% only. Hence your accuracy can't go above 0.5. If the accuracy is 0.5, your score will be 0. Lower the score, higher the rank.

If you consider 3 as the highest score, even the model with 0% accuracy will get score as 3. since  $6 * \text{abs}(0 - 0.5) = 6 * 0.5 = 3$ .

▲ 1 vote ● Reply ● Message ● Permalink



**Thanish Batcha** Edited 2 years ago

Thanks Phani and Anbu for pointing out that even 0% acc will result in max score of 3. I missed out that part of the math. And Anbu can you please help me figure out where has it been said that accuracy of the model is close to 50% ? Do they mean the model we build will only predict accurately 50% of the time ?

Also the asset clustering says "The actual number of clusters are between 5 to 15" so I am confused how many clusters we have to choose(a definite number). Anybody has any thoughts ?

▲ 0 votes ● Reply ● Message ● Permalink



**Sharathkumar Anbu** 2 years ago

Please look under the sub heading "Data" Note: The problem is to forecast return of financial asset from which Y is derived. The accuracy of the models will be relatively low (closer to 50%).

▲ 0 votes ● Reply ● Message ● Permalink



**Iris.3** 2 years ago

Sharath is right. One of my seniors told me that most of the times it's great if you achieve 50-55% in financial classifications.

▲ 0 votes ● Reply ● Message ● Permalink



**Thanish Batcha** Edited 2 years ago

If that's the case then if we are able to achieve 55% classification accuracy then the score we might get is  $6 * \text{abs}(.55 - .5) = 0.3$  right ?

▲ 0 votes ● Reply ● Message ● Permalink



**Shahnawaz Akhtar** Edited 2 years ago

0% accuracy means you are getting every prediction wrong and thus if you flip all predictions you will get 100% accuracy. So it's not right to say that lower the score, higher the rank. What you could say though is farther away you accuracy is from 0.5, better will be score. That is why it is  $\text{abs}(\text{accuracy} - 0.5)$ . Btw, you can also use error in place of accuracy and result will be same. So you can either optimize for accuracy or for error.

Multiplication by 6 part is just because they are weighting overall score by taking 3 part from classification problem and 1 part from clustering problem.

▲ 1 vote ● Reply ● Message ● Permalink

**Nimesh Jha** 2 years ago

Competition Admin: could you please provide the leaderboard. please clarify if lower score is better or higher?

▲ 1 vote ● Reply ● Message ● Permalink

**Thanish Batcha** 2 years ago

Yes if the admin could clarify it can be better because if we get an accuracy of 40% then the score is  $6 * \text{abs}(0.4 - 0.5) = 0.6$ , if we get 60% accuracy then the score is  $6 * \text{abs}(0.6 - 0.5) = 0.6$ . No difference between the two I don't understand why this evaluation is followed

▲ 0 votes ● Reply ● Message ● Permalink

**Pankaj Mishra** 2 years ago

That's the weirdest evaluation metrics I have ever seen :3

▲ 1 vote ● Reply ● Message ● Permalink

**Vinay** 2 years ago

Is there any competition admin to clarify our doubts or do we have to solve our problems on our own?

▲ 0 votes ● Reply ● Message ● Permalink

**Thanish Batcha** 2 years ago

I have even dropped an e-mail to [contact@hackerearth.com](mailto:contact@hackerearth.com). No response yet.

▲ 0 votes ● Reply ● Message ● Permalink

**Praveen Kumar** 2 years ago

Hey Vinay, What are your doubts?

▲ 0 votes ● Reply ● Message ● Permalink

**Thanish Batcha** Edited 2 years ago

Hi Admin, Here are the doubts I am having

1. Is there a leaderboard to show the ranking after the submitted test file has been scored with 50% of the test data? Like everyone can see their position.
2. The error metric for predicting the growth is misleading "score =  $6 * \text{abs}(\text{accuracy} - 0.5)$ ". If we get a classification accuracy of 40% then the score is  $6 * \text{abs}(0.4 - 0.5) = 0.6$ , if we get 60% accuracy then the score is  $6 * \text{abs}(0.6 - 0.5) = 0.6$ . So which one is right/high?
3. For asset clustering it is said "cluster are between 5 to 15" which means the competitors can choose any no. of clusters between 5 to 15?
4. 5 submissions per day means 5 for predicting growth and 5 for asset clustering or the 5 is inclusive of both?

▲ 1 vote ● Reply ● Message ● Permalink

**Praveen Kumar** 2 years ago

Hey Thanish,

For 1 and 2, let me contact soc-gen and get this sorted out.

3-> the actual clusters are between 5 to 15. So, you will have to choose a no. between 5-15 and experiment to arrive at the right number.

4 -> 5 submissions per day is for an individual problem.

▲ 0 votes ● Reply ● Message ● Permalink

**Thanish Batcha** Edited 2 years ago

Hi Praveen, is there any update on the display of leaderboard ???

▲ 0 votes ● Reply ● Message ● Permalink

**Santhosh Kumar** 2 years ago

Hi Praveen,

Any updates on the Leader board & Evaluation Metric? At the moment we are all clueless about whether lower/higher score convey higher rank.

▲ 0 votes ● Reply ● Message ● Permalink

?

**Neel Jambhekar** 2 years ago



For the 2nd question, if you have 40% accuracy since you have binary classification you could just flip all your predictions ,which would get you 60% accuracy.Hence the metric has abs I think.

▲ 0 votes ● Reply ● Message ● Permalink



**Praveen Kumar** 2 years ago

That is absolutely correct.

▲ 0 votes ● Reply ● Message ● Permalink



**Sharathkumar Anbu** 2 years ago

Hi Praveen, Does lower score convey higher rank ? It would be great if the system could display accuracy as well as score.

▲ 0 votes ● Reply ● Message ● Permalink



**Vinay** 2 years ago

Hey, thanks Praveen.

Having the same doubt as Thanish Batcha i.e. "The error metric for predicting the growth is misleading "score =  $6 * \text{abs}(\text{accuracy} - 0.5)$ " .

If we get a classification accuracy of 40% then the score is  $6 * \text{abs}(0.4 - 0.5) = 0.6$ , if we get 60% accuracy then the score is  $6 * \text{abs}(0.6 - 0.5) = 0.6$ . So which one is right/high ?"

One more thing, if the required accuracy is 0.5, then the score must be close to '0', but the board is displaying "Max Score - 3". Is it going to be rectified or am I interpreting it wrong?

▲ 0 votes ● Reply ● Message ● Permalink



**Praveen Kumar** 2 years ago

Please see the explanation under the 'Evaluation' section above for clarification regarding 0.4/0.6. The problem just says that the actual accuracy is close to 0.5. The max score is set to be 3 when accuracy is 1, but you will probably never hit 0 or 1 as the accuracy.

▲ 0 votes ● Reply ● Message ● Permalink



**Vinay** 2 years ago

Ok , got it. Thanks

▲ 0 votes ● Reply ● Message ● Permalink



**Neel Jambhekar** 2 years ago

Admin, please provide a benchmark score for both the problems.

▲ 1 vote ● Reply ● Message ● Permalink



**Santhosh Kumar** Edited 2 years ago

@Admin/Praveen - Thanks for bringing up the LeaderBoard(LB). Is it possible to display the accuracy & then the score in LB?

For classification, let's say there is a model A which gives an accuracy of 0.5 then the score will be zero (  $6 * \text{abs}(0.5 - 0.5)$  ) and will lead to a lower rank in LB.

Let's say there is an another model B which has accuracy of 0.4 then the score will be 0.6 (  $6 * \text{abs}(0.4 - 0.5)$  ) which will result in a higher rank in LB.

Let's say there is an another model C which has accuracy of 0.6 then the score will be 0.6 (  $6 * \text{abs}(0.6 - 0.5)$  ) which will result in a rank equal to the model B rank.

So the metric used to calculate the score rewards both the models B(below benchmark) & C (above benchmark) equally.

Hence, One can built a poor model (not the ideal choice) by lowering the accuracy(easier to implement) to get a higher rank in LB or one can built a good model(ideal Choice) which beats the 0.5 benchmark(difficult) and get a higher rank in LB.

Currently I have a model which gives a score of zero(accuracy 0.5) however I have submitted the result of another model to get a rank of 9 (Klueless Team) in LB. Trying to beat the 0.5 benchmark but the efforts are not rewarded properly by the current evaluation metric.For eg: Accuracy of 0.54 will give only a score of 0.24(  $6 * (0.54 - 0.5)$  ) however it beats the benchmark(0.5).

If you could display the accuracy along with score then it will help us to evaluate our models better.

▲ 1 vote ● Reply ● Message ● Permalink



**Santhosh Kumar** 2 years ago

@Admin - Much appreciate if you could comment on the above post.

?

▲ 0 votes ● Reply ● Message ● Permalink



**Sharathkumar Anbu** 2 years ago

Whenever I submit a prediction file, it's being stuck at the stage "Uploaded. Saving..." What could be the reason

▲ 0 votes ● Reply ● Message ● Permalink



**Sharathkumar Anbu** 2 years ago

Could some one from hacker earth look in to the issue asap . I tried uploading from different browsers as well. I will lose out 5 submissions for today if the issue is not looked into. Please look into it asap

▲ 0 votes ● Reply ● Message ● Permalink



**Praveen Kumar** 2 years ago

Hey Sharath, Are you still facing issues uploading your files?

▲ 0 votes ● Reply ● Message ● Permalink



**Sharathkumar Anbu** 2 years ago

Yes Praveen. Please check asap. I have been waiting long for the issue to get resolved.

▲ 0 votes ● Reply ● Message ● Permalink



**Sudip Maji** ⚡ Admin 2 years ago

Hi Sharath, the issue has been fixed. Please go ahead and submit your solution.

▲ 0 votes ● Reply ● Permalink



**Sharathkumar Anbu** 2 years ago

Thank you Sudip. I have submitted. Does lower score convey higher rank ? It would be great if the system could display accuracy as well as score.

▲ 0 votes ● Reply ● Message ● Permalink



**Gaurav** 2 years ago

What are the sample scores you are getting in both the problems? Anyone willing to share? Provision of leaderboard would have made things clear!!

▲ 0 votes ● Reply ● Message ● Permalink



**Thanish Batcha** 2 years ago

For predict the growth I am getting less than 0.05. What is it you are getting ?

▲ 0 votes ● Reply ● Message ● Permalink



**Gaurav** 2 years ago

For predict the growth I am getting 0.1xx and for asset I am getting 0.0xx. I am loosing interest in the competition due to lack of information!!

▲ 0 votes ● Reply ● Message ● Permalink



**Thanish Batcha** 2 years ago

There is not definite number of cluster defined for cluster problem. It is blindly mentioned between 5 and 15. I don't know how many clusters to create

▲ 0 votes ● Reply ● Message ● Permalink



**RAJAT KUMAR** 2 years ago

Hey guys, Its clearly mentioned that the cluster are between 5 to 15 ( both inclusive) ,so the number of clusters are 12, right?

▲ 0 votes ● Reply ● Message ● Permalink



**Santhosh Kumar** 2 years ago

The number of clusters could be 5,6,7,8,9,10,11,12,13,14 and 15..

▲ 0 votes ● Reply ● Message ● Permalink



**Shahnawaz Akhtar** 2 years ago

I am getting 0.721 for classification and 0.35 for clustering... What about others

▲ 0 votes ● Reply ● Message ● Permalink

?

**Harshita Dudhe** 2 years ago

0.721 for classification? Is this the test accuracy?

▲ 0 votes ● Reply ● Message ● Permalink

**Shahnawaz Akhtar** 2 years ago

it's the score that i am getting.. what about you

▲ 0 votes ● Reply ● Message ● Permalink

**RAJAT KUMAR** 2 years ago

How r u treated time factor? and which algorithm u used for classification?

▲ 0 votes ● Reply ● Message ● Permalink

**Shahnawaz Akhtar** 2 years ago

you need to include information about previous time periods

▲ 0 votes ● Reply ● Message ● Permalink

**Santhosh Kumar** 2 years ago

How to include information about previous time periods?

▲ 0 votes ● Reply ● Message ● Permalink

**Shahnawaz Akhtar** 2 years ago

for time t, add values from time t-1, t-2, t-3 etc... Also try time t+1, t+2, t+3 etc

▲ 0 votes ● Reply ● Message ● Permalink

**mallu baba** 2 years ago

Please dont mislead people. Your method is not working

▲ 0 votes ● Reply ● Message ● Permalink

**Shahnawaz Akhtar** 2 years ago

Just because you can't implement doesn't mean that it's not working..

▲ 0 votes ● Reply ● Message ● Permalink

**mallu baba** 2 years ago

I am using Recurrent Neural Networks with time series information and it is not helping much. So i am assuming there must be a better approach

▲ 0 votes ● Reply ● Message ● Permalink

**Santhosh Kumar** Edited 2 years ago

@shahnawaz Akhtar - Thanks for the reply. It helps!!!

▲ 0 votes ● Reply ● Message ● Permalink

**Santhosh Kumar** 2 years ago

I got 0.518(without using any algorithms, yet to build the models) for classification &amp; 0.38 for clustering..

▲ 0 votes ● Reply ● Message ● Permalink

**Vinay** 2 years ago

Or a leaderboard should work for time being.

▲ 0 votes ● Reply ● Message ● Permalink

**Ashwin Dhaka** 2 years ago

Hi Guys,

The leader board is now on and you can check your respective rankings there along with your current score.

▲ 0 votes ● Reply ● Message ● Permalink

**Thanish Batcha** 2 years ago

?





Thanks for bringing it up admin

▲ 0 votes ● Reply ● Message ● Permalink



**Nimesh Jha** 2 years ago

I understand that the Leaderboard score =  $\text{max\_score}(\text{Asset Clustering}) + \text{max\_score}(\text{predict the growth})$ .  
correct ??

▲ 0 votes ● Reply ● Message ● Permalink



**Praveen Kumar** 2 years ago

yes, that is correct.

▲ 0 votes ● Reply ● Message ● Permalink



**Santhosh Kumar** 2 years ago

How many teams will be selected from the qualifying round?Is it just Top 3 teams/Individuals?

▲ 0 votes ● Reply ● Message ● Permalink



**Praveen Kumar** 2 years ago

The number of teams that will be selected for the offline round will be decided by soc-gen.

▲ 0 votes ● Reply ● Message ● Permalink



**Siddharth Gupta** 2 years ago

Should we submit source file on every submission?

▲ 0 votes ● Reply ● Message ● Permalink



**Praveen Kumar** 2 years ago

You need not submit source file every time but make sure that the source file and a text file discussing about the your approach is submitted before the contest ends.

▲ 0 votes ● Reply ● Message ● Permalink



**Santhosh Kumar** 2 years ago

Hi Praveen/Admin - Is there a way to view the uploaded source file?

▲ 0 votes ● Reply ● Message ● Permalink



**Sudip Maji** ⚡ Admin 2 years ago

Hi guys, the total number of submissions allowed for each problem from a team has been increased to 30.

▲ 0 votes ● Reply ● Permalink



**Lithin Reddy Marla** 2 years ago

Can per day submission limit be increased too?I started to solve the problem yesterday only !

▲ 0 votes ● Reply ● Message ● Permalink



**Santhosh Kumar** 2 years ago

when shall we expect the final results.

▲ 0 votes ● Reply ● Message ● Permalink

About Us

Innovation Management

Technical Recruitment

University Program

Developers Wiki

Blog

Press

Careers

Reach Us





