Hawaii Machine Learning Challenge

Venue Sponsor: Hawaii Technology Development Corporation

https://www.kaggle.com/t/85d5f53d2a0244a7bf1283be7381849d

Table of Contents

Table of Contents

Getting Started

- **Sign up for a Kaggle account**
  https://www.kaggle.com
- **Go to competition link**
  https://www.kaggle.com/t/85d5f53d2a0244a7bf1283be7381849d
- **(optional) sign up for the Hawaii ML Meetup**
  https://www.meetup.com/Hawaii-Machine-Learning/
- **(optional) get an invite to the Hawaii ML Slack**
  send an email to hawaiimachinelearning@gmail.com

Table of Contents

Competition Details – Task

Predict the quantity of items sold in online retail transactions.

- You're given some sales data
- From the data, predict the number items sold in a transaction
- You are evaluated on how well your predictions match reality
- Whoever's predictions are closest to reality wins!

Competition Details – Dataset

- Real data from online retail transactions
- Data is anonymized to discourage cheating
- Training set (with quantity) / test set (without quantity)

`id` - an anonymous id unique to a transaction

`date` - date of the purchase

`time` - time of the purchase

`invoice_id` - the id of an invoice

`stock_id` - id of a stock

`customer_id` - id of a customer

`country` - the country the item(s) was purchased in

`description` - text describing the item(s) purchased

`unit_price` - price for each item

`quantity` - total quantity purchased during the transaction

Competition Details – Dataset

id	date	time	invoice_id	stock_id	customer_	country	description	unit_price	quantity
330477	8/30/2011	12:26	357	198	7	united kingdom	french blue metal door sign 3	2.46	1
113814	11/1/2011	14:54	6757	2747	410	united kingdom	paper lantern 9 point deluxe star	6.65	4
51510	7/11/2011	16:21	52	755	7	united kingdom	fruit salad bag charm	2.46	1
257784	10/11/2011	14:52	3642	1942	1722	france	set of 10 led dolly lights	6.25	2
431803	11/14/2011	14:59	2105	39	1159	united kingdom	glitter christmas star	0.39	5
473125	12/3/2010	11:35	1768	506	1009	united kingdom	set 20 napkins fairy cakes design	0.85	1
223094	11/15/2011	17:00	779	1295	7	united kingdom	gingham heart decoration	1.63	3
314901	10/31/2011	14:09	404	923	13	united kingdom	calendar paper cut design	5.79	1
377668	11/7/2011	11:55	13526	251	1829	united kingdom	ivory knitted mug cosy	1.65	6
238904	7/5/2011	12:15	2319	2789	773	united kingdom	mirrored wall art gents	0.75	1
477169	2/7/2011	11:56	7553	3455	596	united kingdom	carousel ponies baby bib	1.25	1
70332	7/22/2011	9:20	997	2880	7	united kingdom	candy spot heart decoration	0.83	3
190264	9/23/2011	15:17	11229	260	3397	united kingdom	biscuit tin vintage christmas	2.89	6

Competition Details – Evaluation Metric

- Participants are ranked on how well they can predict test data
- Test data is split into public and private
- Evaluation metric is Root Mean Squared Log Error (RMSLE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_i+1)-\log(a_i+1))^2},$$

Competition Details – Rules

- Participants can submit up to 5 submissions per day. Participants will be allowed to submit 2 submissions for evaluation on the test set.
- Private sharing of information is not allowed. If in doubt, post to the discussions forum https://www.kaggle.com/c/hawaiiml0/discussion
- Multiple accounts are not allowed.

Table of Contents

- Getting Started
- Competition Details
  - Task
  - Dataset
  - Evaluation metric
  - Rules
- Kernels
- Prizes
- Timeline
- Q&A

Kernels

- Use Google's resources to run your code.
- Share your work with the community using Kaggle kernels.
- All public kernels will be subjectively evaluated a panel of judges using the following criteria:
  - *Originality* – does the kernel contain original material
  - *Helpfulness* – does the kernel provide valuable information for the task at hand, number of upvotes is one measure of helpfulness
  - *Correctness* – are the technical details of the kernel correct

Table of Contents

- Getting Started
- Competition Details
  - Task
  - Dataset
  - Evaluation metric
  - Rules
- Kernels
- Prizes
- Timeline
- Q&A

Prizes (contingent on funding)

Top 3 teams with the lowest test RMSLE will be awarded

- First Prize: ($100)
- Second Prize: ($50)
- Third Prize: ($25)

Winner of the Best Kernel will receive ($25)

Table of Contents

Timeline

Event	Date	Description
Start	Week 0, March 30 (Friday)	introduction, explain rules, form teams
Tutorial 1	Week 1, April 2 (Monday)	EDA and feature engineering
Tutorial 2	Week 2, April 9 (Monday)	training and prediction
Tutorial 3	Week 3, April 16 (Monday)	cross validation and model selection
Tutorial 4	Week 4, April 23 (Monday)	ensembling and blending
Finish	Week 5a, April 30 (Monday)	no new submissions allowed
Wrap-up	Week 5b, May 4 (Friday)	award prizes, discuss solutions

Table of Contents

- Getting Started
- Competition Details
  - Task
  - Dataset
  - Evaluation metric
  - Rules
- Kernels
- Prizes
- Timeline
- Q&A

You’re All Winners!

You're All Winners!

Except for Cheaters!