

Team Assignment Options

Weizhi Zhu

Assignment I: Team Project Options



	Kaggle Competition	<u>Entrepreneurship Competition</u>	CogX-AIC Conference	Other
Overview	2-Sigma Kaggle contest in financial modeling	Application details to be can be found here	Also known as “CognitionX”, Europe’s largest AI conferece	Other project(s) initiated by student(s) to be approved by instructors
Scope	Public	Public	Private	Public or private
Team Size	Any	Any	5 members	Any
Internship	No	No	Yes ¹	No
Interviews	N/A	N/A	Required	N/A
Nature	Trading	Various	NLP	Any
Deadline	May 31	April 15	May 31	May 31

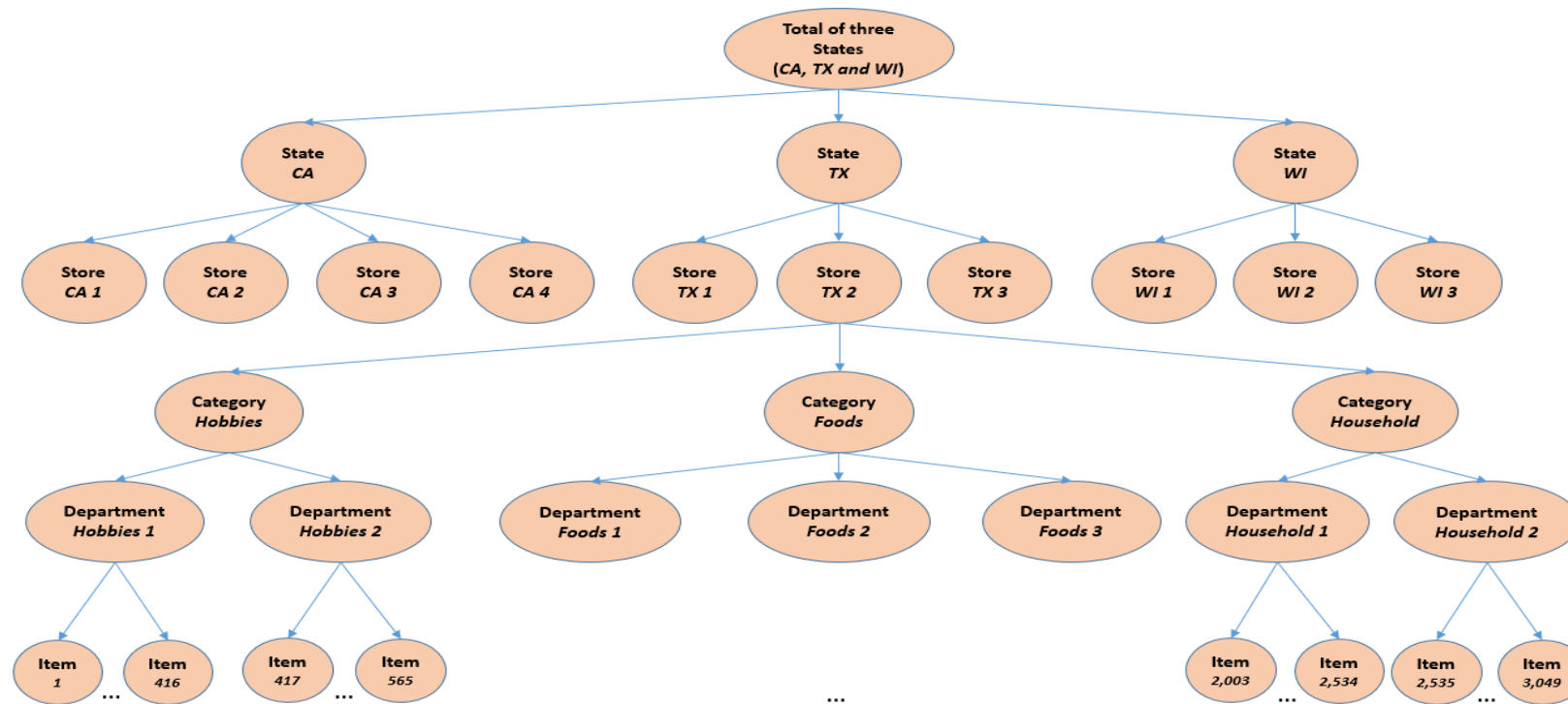
Note: 1. High performers will be invited to interviews for full-time positions at Alpha Intelligence Capital

M5 Estimate the unit sales of Walmart retail goods

- M(MOFC Research) - provides business support/training on prediction.
- 5 – Season 5.
- Multi-dimension(high-dimension) time series prediction with auxiliary features.
- Active(Mar - June), with prize up to 50,000USD.
- Share certain similarities to predicting the movement of stock market in their DNA.

Dataset – Time Series

- Sales of over 30,000 items during ~5.4 years.
- Hierarchical structure of items is available.



Dataset – Auxiliary features

- Sales price of item at given time and store.
- Calendar and events.

2/6/2011	11102 Sunday	2	2	2011 d_9	SuperBowl	Sporting	1	1	1
2/7/2011	11102 Monday	3	2	2011 d_10			1	1	0
2/8/2011	11102 Tuesday	4	2	2011 d_11			1	0	1
	Wednesd								
2/9/2011	11102 ay	5	2	2011 d_12			1	1	1
2/10/2011	11102 Thursday	6	2	2011 d_13			1	0	0
2/11/2011	11102 Friday	7	2	2011 d_14			0	1	1
2/12/2011	11103 Saturday	1	2	2011 d_15			0	1	1
2/13/2011	11103 Sunday	2	2	2011 d_16			0	1	0
2/14/2011	11103 Monday	3	2	2011 d_17	ValentinesDay	Cultural	0	0	1
2/15/2011	11103 Tuesday	4	2	2011 d_18			0	1	1
	Wednesd								
2/16/2011	11103 ay	5	2	2011 d_19			0	0	0
2/17/2011	11103 Thursday	6	2	2011 d_20			0	0	0
2/18/2011	11103 Friday	7	2	2011 d_21			0	0	0
2/19/2011	11104 Saturday	1	2	2011 d_22			0	0	0
2/20/2011	11104 Sunday	2	2	2011 d_23			0	0	0
2/21/2011	11104 Monday	3	2	2011 d_24	PresidentsDay	National	0	0	0

Dataset - Goal

Participants of M5 are asked to provide **28 days ahead point forecasts (PFs)** for all the series of the competition,

<https://www.kaggle.com/c/m5-forecasting-accuracy>

as well as the corresponding **median and 50%, 67%, 95%, and 99% prediction intervals (PIs)**.

<https://www.kaggle.com/c/m5-forecasting-uncertainty>

Each with its separate prizes of \$50,000.

Other options – Implement your trading strategy

Quantopian Community



Learn

Explore [tutorials](#) and ask questions in our [community](#) of over 300,000 members.



Develop

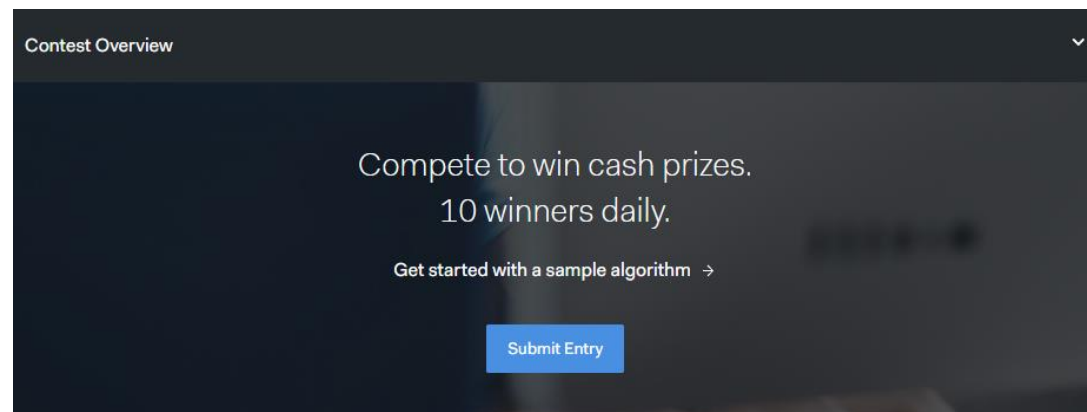
Test ideas in an interactive [research](#) environment with dozens of pre-integrated [datasets](#).



Compete

Enter our daily [contest](#) and community [challenges](#) for an opportunity to win cash prizes.

<p>TUTORIAL 1</p> <p>Getting Started</p> <p>Basic Quantopian Lessons</p>	<p>TUTORIAL 2</p> <p>Pipeline</p> <p>Dynamic Security Selection</p>
<p>TUTORIAL 3</p> <p>Alphas</p> <p>Analyze Factors With Alphas</p>	<p>TUTORIAL 4</p> <p>Writing a Contest Algorithm</p> <p>How to Meet Contest Requirements</p>




Other options - NLP and Unsupervised learning

Dataset

COVID-19 Open Research Dataset Challenge (CORD-19)

An AI challenge with AI2, CZI, MSR, Georgetown, NIH & The White House

 Allen Institute For AI and 7 collaborators • updated 3 days ago (Version 3)

[Data](#) [Tasks \(10\)](#) [Kernels \(17\)](#) [Discussion \(7\)](#) [Activity](#) [Metadata](#)

What is known about transmission, incubation, and environmental stability?

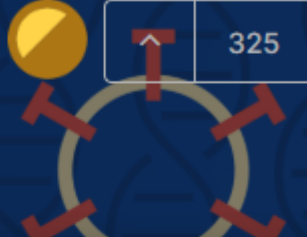
COVID-19 Open Research Dataset Challenge (CORD-19)
Paul Mooney · 0 Submissions

What do we know about COVID-19 risk factors?

COVID-19 Open Research Dataset Challenge (CORD-19)
Paul Mooney · 0 Submissions

Sample task with sample submission

Help us understand how geography affects virality.
devrishi · 1 Submission

 325

Topic Modeling (LDA) on CORD-19 Paper Abstracts

Python notebook using data from [COVID-19 Open Research Dataset Challenge \(CORD-19\)](#) · 27 views · 2h ago

Print Topics

In [6]:

```
top_topics = model.top_topics(corpus)
for i, (topic, sc) in enumerate(top_topics):
    print("\nTopic {}: ".format(i) + ", ".join([w for score,w in topic]))
```

Topic 0: health, disease, outbreak, data, case, public, transmission, influenza, model, epidemic, infectious, study, control, risk, population, method, infection, result, system, pandemic

Topic 1: cell, mouse, infection, response, immune, ifn, expression, induced, infected, viral, disease, il, type, specific, level, macrophage, cytokine, role, interferon, human

Topic 2: cov, sars, human, mers, coronavirus, antibody, bat, respiratory, syndrome, host, specie, protein, epitope, infection, severe, viral, coronaviruses, novel, receptor, vaccine

Topic 3: gene, protein, host, sequence, analysis, structure, genome, viral, interaction, study, different, genetic, expression, using, two, data, function, identified, based, mutation