AI6102: Machine Learning Methodologies & Applications

Team Project Instructions

General Information

- The term paper can be group-based or individual
- Each group consists of at most 3 members (≤ 3)
- The topic of the term paper is flexible, 4 options to choose:
  - Participate an ongoing Kaggle competition (has bonus marks based on some conditions)
  - Participate a completed Kaggle competition
  - Conduct a research project on machine learning (has bonus marks based on some conditions)
  - Survey an advanced topic of machine learning
- Submission: a report in PDF (NTULearn Assignment Link)
- **DEADLINE: 11:59pm, 30th April 2022**
- **Email Han Yucheng (YUCHENG002@e.ntu.edu.sg) your group members and name before 28th Feb 2022 to create groups on NTULearn**

Brief Description

- Ongoing Kaggle Competition
  - Any ongoing Kaggle Competition which will end by April 2023 with a duration within 6 months
- Completed Kaggle Competition
  - Pick one from the candidate list on the next slide
- Research Project
  - Any research project related to machine learning: either fundamental research (ML methods) or application
- Research Survey
  - Conduct a comprehensive survey on an advanced topic of machine learning

Kaggle Competition Candidates (Completed)

- Zillow Prize: Zillow's Home Value Prediction (Zestimate)
  url: https://www.kaggle.com/c/zillow-prize-1
- Sberbank Russian Housing Market
  url: https://www.kaggle.com/c/sberbank-russian-housing-market/
- Google Smartphone Decimeter Challenge 2022
  url: https://www.kaggle.com/competitions/smartphone-decimeter-2022
- Store Item Demand Forecasting Challenge
  url: https://www.kaggle.com/c/demand-forecasting-kernels-only/
- Nomad2018 Predicting Transparent Conductors
  url: https://www.kaggle.com/c/nomad2018-predict-transparent-conductors/
- HuBMAP - Hacking the Kidney
  url: https://www.kaggle.com/competitions/hubmap-kidney-segmentation
- PetFinder.my - Pawpularity Contest
  url: https://www.kaggle.com/competitions/petfinder-pawpularity-score
- Elo Merchant Category Recommendation
  url: https://www.kaggle.com/c/elo-merchant-category-recommendation/
- Northeastern SMILE Lab - Recognizing Faces in the Wild
  https://www.kaggle.com/c/recognizing-faces-in-the-wild/

Report – Kaggle Competition including both completed & ongoing

Recommended content (NeurIPS/CVPR/ICML latex template)

- Problem statement of the task in your own words
- Challenges of the problem
- Your proposed solution
  - Motivation (feature engineering, ML methods, etc)
  - Detailed steps with explanations
- Experiment
  - Conduct experiments to verify your motivations
  - Conduct experiments to show the performance of your proposed solution (both cross-validation results and the screenshot of leaderboard results)
- Conclusion
  - What you have learned from the competition

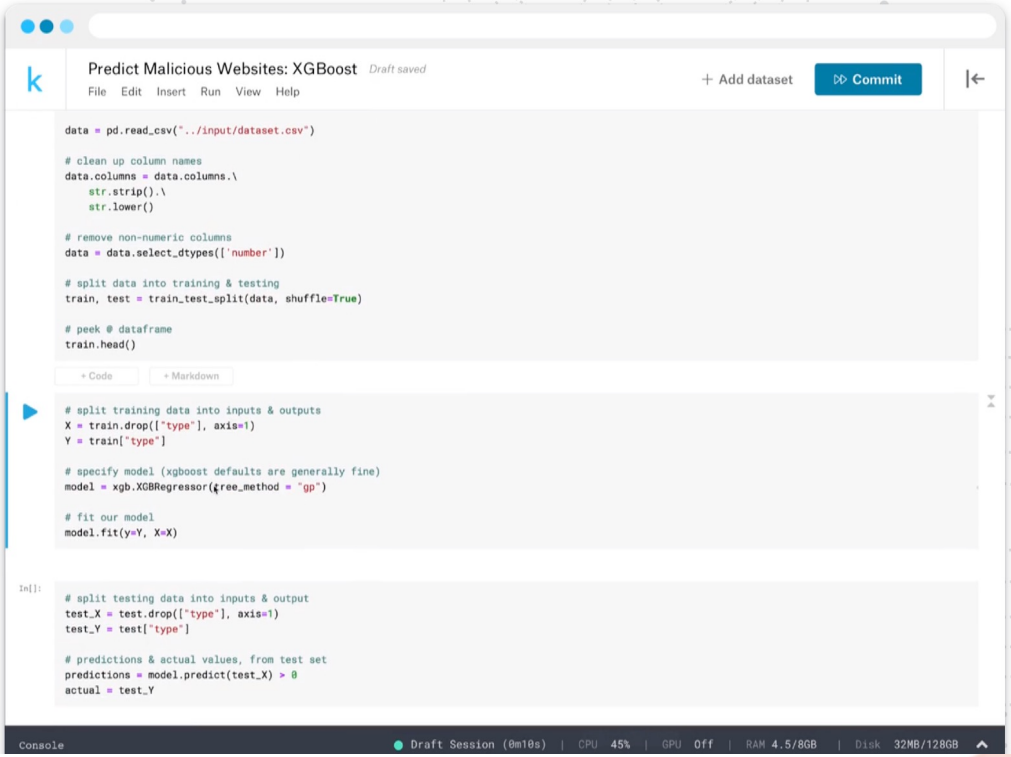
kaggle Competitions Datasets Code Discussions Courses ••• Search Sign In Register

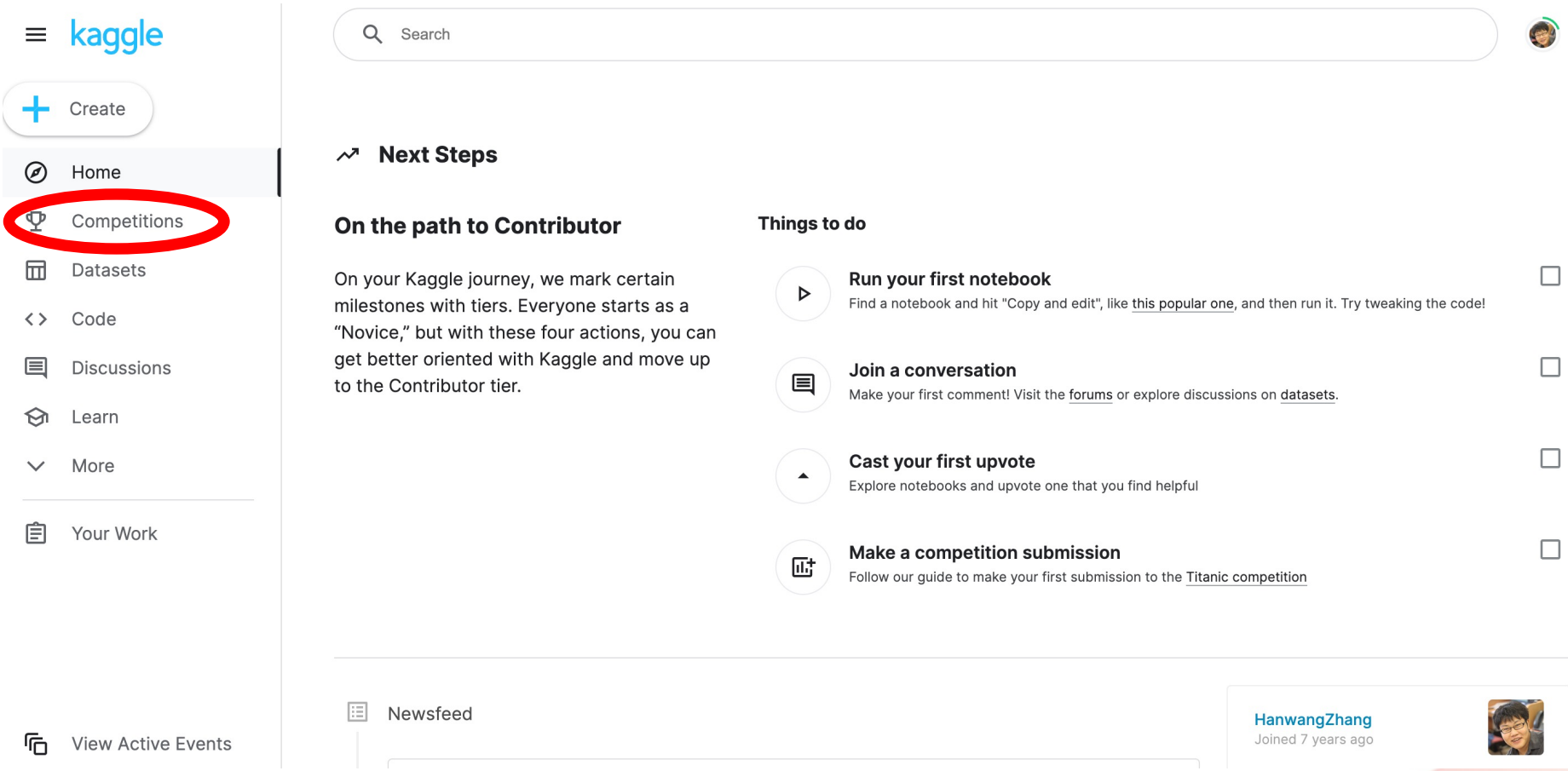
Start with more than a blinking cursor

Kaggle offers a no-setup, customizable, Jupyter Notebooks environment. Access GPUs at no cost to you and a huge repository of community published data & code.

REGISTER WITH GOOGLE

Register with Email




kaggle
Search
Create
Home
Competitions
Datasets
Code
Discussions
Learn
More
Your Work
View Active Events
Next Steps
On the path to Contributor
On your Kaggle journey, we mark certain milestones with tiers. Everyone starts as a "Novice," but with these four actions, you can get better oriented with Kaggle and move up to the Contributor tier.
Things to do
Run your first notebook
Find a notebook and hit "Copy and edit", like this popular one, and then run it. Try tweaking the code!
Join a conversation
Make your first comment! Visit the forums or explore discussions on datasets.
Cast your first upvote
Explore notebooks and upvote one that you find helpful
Make a competition submission
Follow our guide to make your first submission to the Titanic competition
Newsfeed
HanwangZhang
Joined 7 years ago

kaggle

Search

Create

Home

Competitions

Datasets

Code

Discussions

Learn

More

Your Work

View Active Events

Competitions

Grow your data science skills by competing in our exciting competitions. Find help in the documentation or learn about Community Competitions.

Host a Competition | Your Work

Search competitions | Filters

All Competitions — Everything, past & present

Featured — Premier challenges with prizes

Getting Started — Approachable ML fundamentals

Research — Scientific and scholarly challenges

Community — Created by fellow Kagglers

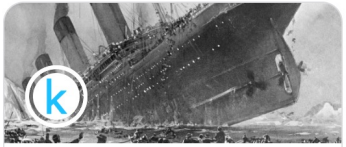
Playground — Fun practice problems

Get Started

See all

New to Kaggle?

These competitions are perfect for newcomers.



Titanic - Machine Learning

House Prices - Advanced

Spaceship Titanic

kaggle

Search

Create

Home

Competitions

Datasets

Code

Discussions

Learn

More

Your Work

View Active Events

Overview Data Code Discussion Leaderboard Rules

Late Submission ...

Overview

Description

Evaluation

Timeline

Prizes

You may have heard that blending predictions from model predictions can give better results than using the output of a single model. There are many different strategies that can be employed for this, and they are great to learn if you're looking for an effectively free boost in model scores. The November Tabular Playground is the chance to practice this skill!

Overview Data Code Discussion Leaderboard ules Late Submission ···

Leaderboard

Raw Data Refresh

Search leaderboard

Public Private

The private leaderboard is calculated with approximately 75% of the test data.
This competition has completed. This leaderboard reflects the final standings.

#	△	Team	Members	Score	Entries	Last	Solution
1	▲ 5			0.51845	82	1mo	
2	▼ 1	Cristian Sanabria		0.51852	138	1mo	

Report – Research Paper

Recommended content

- Problem statement of the research topic
- Related work – limitations of existing methods
- Your proposed method
  - Motivation
  - Detailed steps with explanations
- Experiment
  - Conduct experiments to verify your motivations
  - Conduct comparison experiments with SOTA methods
- Conclusion
  - Limitation of your proposed method, future work

Report – Research Survey

Recommended content

- Problem statement of the research topic
- Categorize existing methods based on your own insights
- Describe some representatives in each category
  - Motivation, high-level idea, detailed methodology
- Experiment
  - Conduct comparison experiments between SOTA methods
  - Analyze advantages and disadvantages of SOTA methods
- Conclusion
  - Potential new directions of the topic

Assessment

- Writing
  - Whether the organization is logically clear
  - Whether the descriptions on the proposed solution / proposed method / existing SOTA methods are clear
  - Whether there are many typos
- Experiments
  - Whether the experiments are extensive to verify the proposed solution / proposed method or to analyze the disadvantages or advantages between SOTA methods
- Bonus (+20%)
  - For ongoing Kaggle competition, if the ranking in leaderboard is top 10% (cutoff date: by 30th April 2023)
  - For research paper: if a novel idea is proposed compared with existing methods

Frequent Q&A

- Regarding completed Kaggle competitions, can we choose another Kaggle competition beyond the candidate list?
  - No, you can only choose one from the candidate list if you decide to choose a competed Kaggle competition for your term paper
- Are there requirements on the format of the report?
  - No hard requirements. Suggestions: Latex template, e.g., search CVPR 2023 Author Kit
  - Don’t forget include ALL team members NAME and Matric.
- Can we use other ML techniques beyond what are taught in this module, such as deep learning models?
  - Yes, you can
- Do we need to submit source codes?
  - No
- Can the report submission deadline be extended?
  - No, it is a hard deadline

Thank you!