More Administrivia: Projects

- 3 % of course grade (20% for full project: 3 + 5 + 12)
- **Team information due Fri Feb 20.**
  - 3 members per team.
  - Submit information on google form (announcement soon).
- **Project proposal due Wed Mar 1.**
  - A proposal template document will be released in the coming days.
  - A project mentor will be assigned to you based on your proposal.

- Guidance on project topics: See next 2 slides.

"Standard" Projects

- The recommended option barring exceptional cases.
- Tied closely to any one from a pre-approved list of Kaggle projects (announcement soon).
- **Part 1: Implementation**
  - **Option 1:** Extensive evaluation of design decisions in pre-existing codebases.
    - Evaluate the design decisions in existing Kaggle submissions, e.g. current leading submissions, or other codebases on the web for this problem. Always cite and acknowledge.
    - Recommendations:
      - For tabular datasets, significant feature engineering and try several models
      - For image datasets, try different neural network architectures etc.
  - **Option 2:** New ML approach. A new ML approach, not directly building on top of current codebases
    - Typically a learning strategy (e.g., semi-supervised learning) or a neural network architecture
- **Part 2: Evaluation**
  - **Part 2a:** Systematic evaluation of hyperparameters (e.g., regularization, learning rate, etc.)
  - **Part 2b:** Evaluate on test data distributions different from training data
    - E.g. Add synthetic noise to test set, train-test split based on demographic features or time
    - Plot performance measures vs. degree of shift (e.g. for demographic features, include X% fraction of minority in the training set, where X is degree of shift)
    - Particularly interesting to identify "small" shifts that break the model.

- No collaboration outside your project team.
- Public submission to Kaggle leaderboard at end of project period together with code. **You will not be graded only on leaderboard position though.** More creative and ambitious projects will be held to lower final performance standards than more incremental projects.

Standard Project Options 1-5

- <u>Project topic</u>: Animal video analysis
  - Kaggle link / link to resource: https://www.kaggle.com/competitions/iwildcam2022-fgvc9
  - Brief Description: To count the number of animals in a sequence of images
- <u>Project topic</u>: Cassava Leaf Image Classification
  - Kaggle link / link to resource: https://www.kaggle.com/competitions/cassava-leaf-disease-classification
  - Brief Description: To classify each cassava leaf image into four disease categories or as a healthy leaf (5th category)
- <u>Project topic</u>: Recommending new TV shows
  - Kaggle link / link to resource: https://www.kaggle.com/competitions/recommender-system-2022-challenge-polimi
  - Brief Description: To recommend a TV show the user has not yet interacted with in any way
- <u>Project topic</u>: Forecasting stock prices
  - Kaggle link / link to resource: https://www.kaggle.com/code/ysthehurricane/advanced-stock-pred-using-svr-rfr-knn-lstm-gru/data
  - Brief Description: To predict Reliance Industries Limited Stocks using trading data from 2020 to 2021. Task: train models to predict stock prices 1 day, 2 days, 1 week, and 2 weeks into the future.
- <u>Project topic</u>: Forecasting power demand
  - Kaggle link / link to resource: https://github.com/zhouhaoyi/ETDataset
  - Brief Description: The electric power distribution problem is the distribution of electricity to different areas depends on its sequential usage. But predicting the following demand of a specific area is difficult, as it varies with weekdays, holidays, seasons, weather, temperatures, etc. This project is to predict the electrical transformers' oil temperature based on super long-term real-world data with high precision.

Standard Project Options 6-7

- Project topic: Natural Language Processing - Sentiment Analysis
  - Kaggle link / link to resource: https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews
  - Brief Description: Analyzing sentiments in Amazon food reviews. Also provides short review summaries in addition to the full reviews.

- Project topic: Reinforcement Learning for crop management
  - Kaggle Link / link to resource: https://gitlab.inria.fr/rgautron/gym_dssat_pdi
  - Brief description: No Kaggle for this. May be more challenging than other topics. Learning policies for making crop management decisions. Baseline models included in: https://arxiv.org/pdf/2207.03270.pdf

"Non-Standard" Projects

- Strongly recommended that you use the "standard" option from the last slide.

- If you have good reason to go beyond this, e.g., you would like to propose a project tied to your PhD research, you could do so, but these submissions will go through greater scrutiny for approval.