# Fall 2023 IST 557: Data Mining: Techniques and Applications

Instructor: Lu Lin

**Individual Project II: Image Classification (20 points + 1 bonus point)**
**Due Date: 11:59 PM, Friday, Oct 27, 2023**

**Goal:** This project is to expose students to real-world Kaggle Data Mining competition for practicing basic skills on classification task. DO NOT CHEAT. You can only learn data mining and machine learning by getting your hands dirty. Enjoy it.

**Logistics:** To successfully complete the Kaggle competition, follow the following rules:

(1) ONLY create ONE account for the competition.

(2) Welcome to use ANY deep learning models (e.g., CNN), with any hyperparameters (e.g., number of layers, number of filters, size of filters etc). You may follow the Pytorch tutorial to design your classifier.

(3) You can submit your prediction **10 times each day in maximum**.

**Kaggle Competition:** The kaggle competition for this project can be accessed via this link: https://www.kaggle.com/t/f5c929cb4bc5458c8ac69c7e2406286c. This competition is created only for students enrolled in IST 557 to participate. **Please do not spread this link.**

**Submission Checklist:** You need to submit the following to **Individual Project II on CANVAS**:

(1) **Code file** (runable Jupyter notebook or python file) used to train your model and make predictions for the final submission. **(if missing, -5 pts)**

(2) **PDF report** (within 3 pages with proper fomatting) summarizing the following:

- Kaggle account name, your name and PSU email **(2 pts)**
- Screenshot of your final rank in the leaderboard **(2 pts)**
- Explanation about your model, including the model type, the value and meaning of the hyperparameters **(2 pts)**
- Explanation about your optimization procedure, including number of iteration, learning rate, optimization method you use **(2 pts)**
- What did you tried and what were your findings? **(2 pts)**

**Grading Rubric:** Total 20 + 1 bonus points consists of two parts:

- Performance (Accuracy) on the **Public** leaderboard **(10 points + 1 bonus point)**
  - Accuracy $\in [\mathbf{0.50}, 1.00)$: 10 points
  - Accuracy $\in [\mathbf{0.40}, 0.50)$: 8 points
  - Accuracy $\in [\mathbf{0.30}, 0.40)$: 6 points
  - Accuracy $\in [\mathbf{0.20}, 0.30)$: 4 points
  - Accuracy $\in [\mathbf{0.10}, 0.20)$: 2 points
  - Accuracy $\in [\mathbf{0.00}, 0.10)$: 0 points
  - **Students that achieve Top 3 highest Accuracy on the Private Leaderboard are awarded 1 bonus point.**
- Quality of report (or Jupyter notebook) **(10 points)**: it can be a brief document, as long as you demonstrate the information about your model and optimization setting.