Hackathon Info session - Wed April 12th, 11AM-Noon

Hackathon website and participation form:
https://dsi.udel.edu/events/dsi-symposium-2023/hackathon/

Prof. Sunita Chandrasekaran, Department of Computer and Information Science
Prof. Benjamin Bagozzi, Department of Political Science & International Relations

Key Milestones

- Industry/academic groups propose problems/teams: **May 5th deadline**

- Open invitation to UD (under)graduate students: **June 5th deadline**

- Hackathon (In-person on UD's STAR Campus): **July 17th-19th**

- Hackathon Awards/Presentations: **September 22nd** (DSI Symposium)

DS+AI Hackathon Focus

Goal: Collaboratively solve problems using data science, machine learning, and/or artificial intelligence

Example problem statement: How do we improve the predictive performance of machine learning models for cancer challenges such as drug response problem? …Especially given small datasets that can limit the performance of models and cause them to overfit?
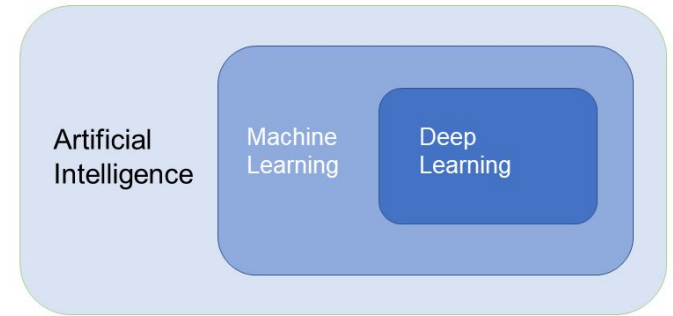
Potential hackathon outcomes: Perform the necessary preprocessing steps to clean up the data and do any feature engineering necessary before training the model including creating train, test, and validation splits for your dataset. Choose a machine learning model that can suit the nature of your data and the problem you are dealing with. Some parameters include classification vs. regression problem, etc. Perform hyperparameter optimization to improve model performance.
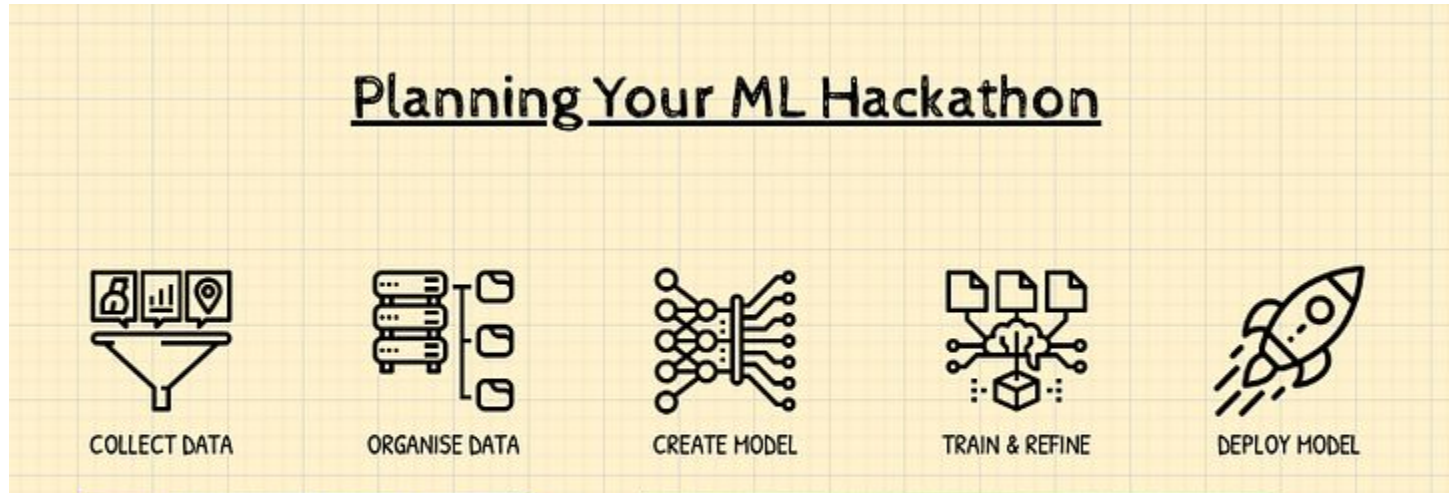
A Typical Team's Composition

- Aiming for teams of four
- A domain scientist (Physicist, chemist, soil scientist, epidemiologist…)
  - To explain/understand the scientific outcomes
- One or more computer scientist(s) or data scientist(s)
  - To analyze data, perform statistical modeling, provide predictive outcomes
- Grad/undergraduate students, postdocs, research staff, DSI fellows
- Each team will have a lead

Why and When AI/ML?

- When the problem requires determinations
- When there are hundreds or thousands of variables and features involved in the decision making
- When there is a need to predict results on new data
- When patterns in large volumes of data need to be determined
- When there is a need to advance recommendation engines

Artificial Intelligence   Machine Learning   Deep Learning

A Potential Hackathon Workflow

Planning Your ML Hackathon

COLLECT DATA · ORGANISE DATA · CREATE MODEL · TRAIN & REFINE · DEPLOY MODEL

Critical Questions to Frame Prior to the Hackathon

- What do you want to accomplish from the hackathon?

- What is your research question? Why do you think ML/AI is suitable for your problem? Could you have solved the problem without ML/AI?

- What are your needs in terms of software, ML/AI expertise, data, programming languages, and compute resources?

Potential Outcomes

- Identify a good starting point
- Learn relevant tools pertaining to the type of datasets under study
- Learn about data preparation/wrangling/cleaning/processing/feature engineering
- Learn data characteristics - Complete? Balanced? Imbalanced? Distributed? Heterogeneous?
- Learn about ML solutions
- Learn about validation methods
- Training the next-generation workforce on ML
- Mentors are building skills
- Knowledge gained outside of one's comfort zone
- Building effective communication strategies between domain and computer scientists

Hackathon Mentors

- Mentors will be assigned per each team
- Floating mentors will also assist
- Mentors will have:
  - A background on Python (and other potential languages)
  - Knowledge of ML pipelines and working with datasets
  - Knowledge of using UD DARWIN and other types of cluster
- Key mentors will be recruited from:
  - DSI Fellows
  - CIS PhD students
  - Interested senior undergrad students

Best Practices

- Use version control like GitHub where possible for code sharing
- There will be impromptu scrum sessions
  - Gather thoughts/results and present to the rest of the audience
- Save code, outcomes and summaries of findings
- Develop results into formal presentations and other outputs
- Practice good governance and collaboration
  - FAIR policies
  - Team science
  - Welcoming new participants
  - Communication across academic and industry participants

Sensitive Code/Project?

- We are working to develop a pre-participation agreement to manage IP-related issues for industry teams and problem proposers

- This relates to code developed during the Hackathon, any data results/outputs, and formal presentations of the above at the DSI Symposium

- For now, please let us know if you have any specific concerns, needs, or questions on this particular item, and we will work to incorporate these.

Hackathon Awards

- Awards will be given by the DSI/AICoE to three teams
- Award focus may include best ML/code-pipeline, best speedup, and/or best visualization
- Award winners will be identified immediately following the Hackathon, and these winners will then have until the DSI's Data Science Symposium on September 22nd to further develop their analysis, code, and results
- Presentations by award winning teams will be featured in a special session at the Data Science Symposium (https://dsi.udel.edu/events/dsi-symposium-2023/)

Pre-requisites (Self-learning Materials)

https://developers.google.com/machine-learning/crash-course

https://hackathon.guide/

https://www.hackerearth.com/hackathon/

Hackathon Organizers and Contact Info

- Sunita Chandrasekaran - schandra@udel.edu
- Benjamin Bagozzi - bagozzib@udel.edu
- Pinar Kullu - pkullu@udel.edu
- Sam Smith - sesmith@udel.edu
- Lynette Carney - lynettec@udel.edu

Hackathon:
https://dsi.udel.edu/events/dsi-symposium-2023/hackathon/

Data Science Institute: https://dsi.udel.edu/

AI Center of Excellence: https://sites.udel.edu/ai/