# Important Dates

**• Proposal due: Feb 28, 2023  
• Intermediate project report due: Mar 31, 2023  
• Tentative Final project presentation: Apr 12 / Apr 14 / Apr 19 / Apr 21 /  
Apr 26 / Apr 28, 2023  
• Final project report due: Apr 30, 2023**

# Deliverables

* **Code  
  The code (proposal, report, and program) should be maintained in GitHub, with commits reflecting the efforts from each team member.**
* **Project proposal  
  Project proposal (1-2 pages) should cover:**

**• Project title  
• Team members  
• Description of the problem.  
• A brief survey of what have been done and how the proposed work is  
different.  
• Preliminary plan (milestones) and Reference (a list of papers)**

* **Group Member**
* [hwang218@hawk.iit.edu](mailto:hwang218@hawk.iit.edu) Bo: 2nd semester, CS, fresh man
* [psubramanian@hawk.iit.edu](mailto:psubramanian@hawk.iit.edu) JD: Internship, final semester
* [xzhou70@hawk.iit.edu](mailto:xzhou70@hawk.iit.edu)May: Developer, working in Japan before, Salesforce Consultant
* Project Orientation:
* App
* 2 Projects:
* Main Proposal: JD’s
* Backup Proposal: May

|  |  |  |
| --- | --- | --- |
| Name - Project | Pros (+) | Cons (-) |
| Bo | Interesting, Novel | Dataset |
| JD | Practical, Dataset, Coding, Quick  Prediction System  Recommendation System? | Can be a little bit change, Easy?!  Can be a little bit complicated? |
| May | Real-life related, Dataset, Coding. Classify | Result Identity? |

* **Member’s Task:**
* **Project Analyzing – Requirements, Dataset, Input-Output**
* **Data Preprocessing – Noise Data, Type, Missing Data**
* **Model Creation – Model Function, Optional Parameters**
* **Model Training – Training Dataset**
* **Model Testing – Testing Dataset**
* **Report Writing – Survey, Project Description, Result Analyzing, Conclusion, etc.**
* Links
* GitHub: https://github.com/May-Xiaoting-Zhou/CS584\_Machine-Learning\_Project
* [Kaggle: Your Machine Learning and Data Science Community](https://www.kaggle.com/)
* [**https://www.overleaf.com/**](https://www.overleaf.com/)

# 1st Discussion: 10 Feb 2023 Friday

1. **Warm-up & self-Introduction**
2. **Determine the orientation of the project: Theory Or Application?**

* **May: Application**
* **JD: Theory (prior project)**
* **Bo: Application**

1. **Clarify everyone's tasks**
2. **Set up schedule for Proposal due (28)**

* **Tasks for next week:**

1. **Clarify the requirement for each orientation 15-Feb Wed, orientation should be determined**
2. **First version of Proposal should be made by 22-Feb Wed.**

* **Next discussion date (Done)**
* **Peer Review?**

**Ideas from Bo:**

1. **Complex -> App, Graph after training to speak Chinese, CNN**
2. **Rating for photo editor & designer. Input (original photo) 🡪 output (the best edited photo)**

**May:**

* **Computer Vision**
* [CoronaHack -Chest X-Ray-Dataset | Kaggle](https://www.kaggle.com/datasets/praveengovi/coronahack-chest-xraydataset)

An imaging method to identify the corona virus.

Classify the X Rays of Healthy vs Pneumonia (Corona) affected patients

Dataset: Chest\_Xray\_Corona\_Metadata.csv

* Input: X Ray Image
* Output: Classification of Healthy or Affected
* [Lung and Colon Cancer Histopathological Images | Kaggle](https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images)

An imaging method to identify Cancer.

This dataset contains 25,000 [histopathological images](https://en.wikipedia.org/wiki/Histopathology) with 5 classes.

There are five classes in the dataset, each with 5,000 images, being:

* Lung benign tissue
* Lung adenocarcinoma
* Lung squamous cell carcinoma
* Colon adenocarcinoma
* Colon benign tissue
* Input: five classes’ images
* Output: Classification of Healthy or Cancer
* **Prediction (TBD)**
* [Store Sales - Time Series Forecasting | Kaggle](https://www.kaggle.com/competitions/store-sales-time-series-forecasting/overview)

# 2nd Discussion: 15 Feb 2023 Wed

1. **Determine the orientation of the project**

**Note: please prepare the description of the problem. (Done)**

1. **Clarify everyone's tasks**
2. **Define the task for next discussion (details)**

**Orientation -> Topic (problem) -> details (papers, algorithm, etc.)**

# 3rd Discussion: ~~22 Feb 2023 Wed~~ 24 Feb 2023

1. **First version of Proposal: Review & Discuss**
2. **Some details to be determined: Member’s task, schedule etc.**
3. **Others: Coding part - Python?**

* **CS584\_Project-Proposal**
* It is strongly recommended to use **LaTeX** to format the project report, and use the ACM SIG or CVPR template. How to use it?
* **Project proposal  
  Project proposal (1-2 pages) should cover:**

**• Project title  
• Team members  
• Description of the problem  
• A brief survey of what have been done and how the proposed work is  
different  
• Preliminary plan (milestones) and Reference (a list of papers)**

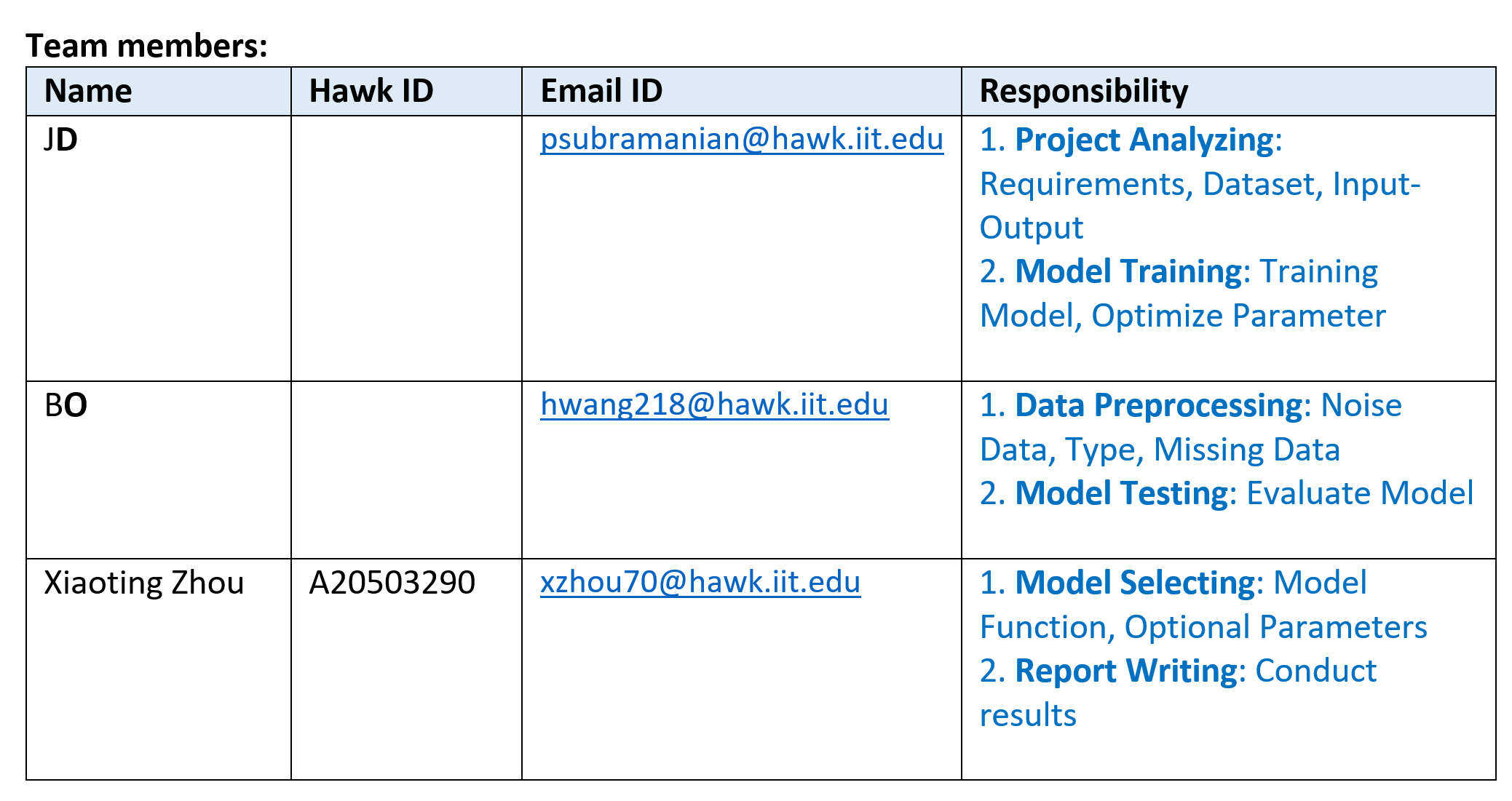
# 4th Discussion: 24 Mar 2023 Fri

1. **Intermediate Project Report discussion**

**The intermediate project report (3-5 pages) should cover:  
• a high quality introduction and problem description  
• description of the data used in the project  
• what have you done so far  
• what remains to be done**

**What we have done so far:**

**Prepare data set , Project Analysis, Decide ML Language (Python), Packages built-in functions Linear/Classification, IDE (Pytorch)**

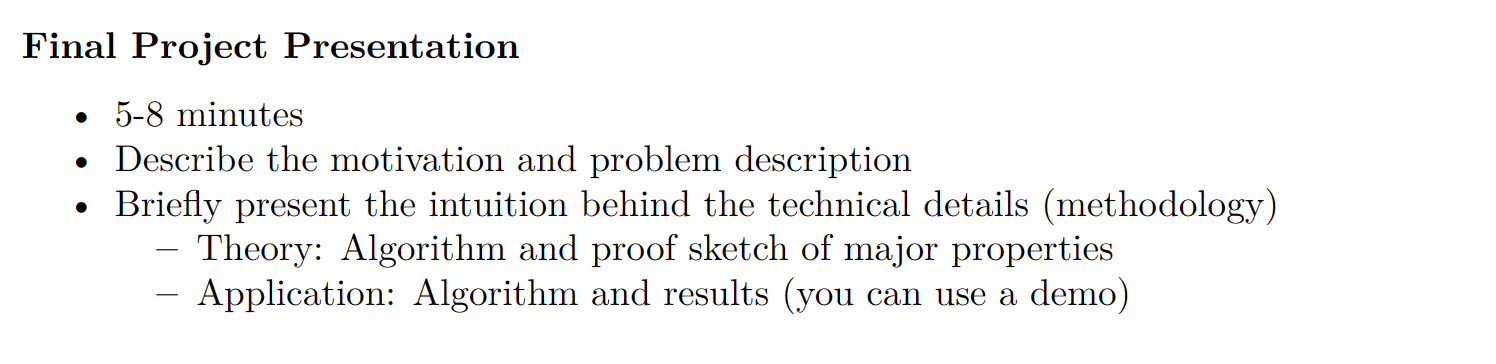
****

**Coding:**

1. **Dataset (training and test data set) : Bo-Data Processing**
2. **ML model selection: training and test : JD-Project Analyzing & May**
3. **Evaluation: ?**
4. **Final Report**

**JD-Project Analyzing: Next Thursday 29th Mar**

1. Create the Intermediate Project Report, then send it to you next Monday
2. You guys review it, then update/modify it
3. Last, Bo re-format it
4. Submit by next Thursday
5. Tentative Final project presentation discussion

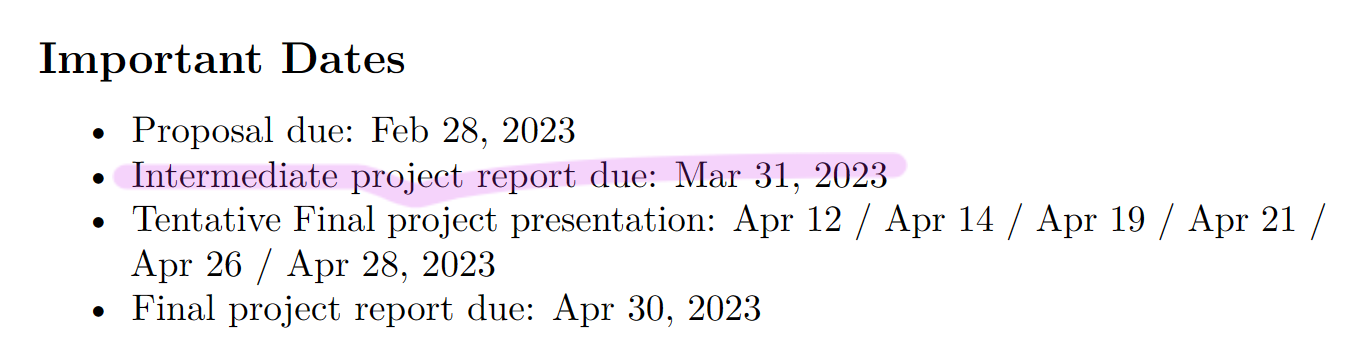
****

1. Project schedule discussion

* **Coding, Demo**
* **Report**

**Important Dates**

• ~~Proposal due: Feb 28, 2023 DONE~~  
• Intermediate project report due: Mar 31, 2023 🡪 Processing, plan to finish 30 Mar  
• Tentative Final project presentation: Apr 12 / Apr 14 / Apr 19 / Apr 21 / Apr 26 / Apr 28 🡪 Prepare



**Goal:**

1. What is the most frequently (rarely) bought recently? (Calculate it by hour, day, week, weekend, etc.)

2. Adjust stock

1) Add more items to the stock for popular items

2) Sell unpopular and will be expired items ASAP

3. **Recommend items** (High percentage of bought and Low percentage of bought) to the customers

**TODO:**

The due time of project proposal submission is 11:59 pm, Wednesday, Mar 1st.

1. **Format proposal: till Morning Sunday 26th Feb**
2. **Team Review & Finalize the proposal: Sunday, Monday**
3. **Submission: Monday, Tuesday 28th Feb**

**Preliminary plan (please add some description or anything you want to add by Sunday/Monday)**

**JD:**

1. **Project Analyzing**
2. **Data Preprocessing**
3. **Model Selecting**

**Bo:**

1. **Model Training**
2. **Model Testing**
3. **Model Evaluation**
4. **Report Writing**

**May:**

**Reference & Tool (Mine) & Re-format Finalize by next Monday/Tuesday**

**TBD – Copy & Paste**

# 5th Discussion: 29 Mar 2023 Wed

1. Project Schedule
2. Project Analysis & Understanding Today (Mar 29)
3. Data Preprocessing: Apr 3
4. Model Training: Apr 6
5. Model Evaluation: Apr 13
6. Parameter Tuning: Apr 18
7. Making Predictions: Apr 24
8. Final Project Report Writing: Apr 28 (Due: May 1st)
9. Intermediate Project Report Review & Finalize
10. Project Introduction From JD
11. Source code
12. QA
13. ~~Set Up next Discussion & Due days to us~~
14. ~~Update the intermediate report tomorrow to Bo, and then refine & reformat it~~
15. ~~Share schedule spreadsheet~~

# 6th Discussion: 7 April 2023 Fri

1. Project Schedule
2. ~~Project Analysis & Understanding (Mar 29)~~
3. ~~Coding Environment Setup (R -> Python): Tue 4-Apr~~
4. **Bo: Data Preprocessing: Waiting Contact TA/Professor**
5. **JD: Model Training:**
6. Model Evaluation:
7. Parameter Tuning:
8. **May:** Making Predictions:
9. Final Project Report Writing: Apr 28 (Due: May 1st)
10. **Presentation Prepare Fri 14-Apr**
11. PPT: 10 mins (2.5 mins/person): 3 parts

1.1 JD: Read PPT (Project Details **Model**)

1.2 Bo: Read PPT (**Data** processing)

1.3 May: Project schedule: What have we done so far & Remaining works

**Describe the motivation and problem description**

1. Modify/Update PPT 🡪 May
2. Graphs/Demo Pick Up (Recording Pre-Watch)
3. QA & Next Meeting

**Wed 12-Apr 14:00 – 15:00**

# 7th Discussion: 12 April 2023 Wednesday

1. Project Schedule confirmation
2. ~~Project Analysis & Understanding (Mar 29)~~
3. ~~Coding Environment Setup (R -> Python): Tue 4-Apr~~
4. **Bo: Data Preprocessing: Waiting Contact TA/Professor**
5. **JD: Model Training:**
6. Model Evaluation:
7. Parameter Tuning:
8. **May:** Making Predictions:
9. Final Project Report Writing: Apr 28 (Due: May 1st)
10. **Presentation task assignment: This Friday**
11. PPT: 10 mins (2.5 mins/person): 3 parts

1.1 JD: Read PPT (Project Details **Model**)

1.2 Bo: Read PPT (**Data** processing)

1.3 May: Project schedule: What have we done so far & Remaining works

**Describe the motivation and problem description**

1. Graphs/Demo Pick Up (Recording Pre-Watch)
2. QA & Next Meeting

**8th Discussion**

**Wed 12-Apr 14:00 – 15:00**

# 8th Discussion: ~~21 April~~ 24 April 2023 Monday

1. **Final Report:**
   1. **Content**
   2. **Plot Chart etc.**
2. **Model**
3. **Schedule: Due 1st May**
   1. **First Version: Today**
   2. **Fill in your parts**
   3. **26 April Wednesday Professional Version**
   4. **29 / 30 April Submit**
4. **Upload related documents and code to GitHub**