

Manufacturing Data Science Spring 2025

Project Instruction

Purpose: Apply the methodologies of manufacturing data science or operations research to the "real problem" and support the decision-making process. **Project “MUST” be related to Manufacturing or Service Systems.** The public dataset is also encouraged and please find the dataset you are confident of **interpreting** the result well.

Proposal: Please just share your idea with us about what your team would like to do. Your team should upload **1-2 page PPT proposal** to NTU COOL by **5pm, May 8, 2025**, and give a **1-minute talk** on **May 9, 2025** in class.

Guideline: This project has 25% in your final grade. This is a team-work project (Individual project is NOT allowed). Each team has **4-5** members (please send your team members with NAME and ID to TA 李郁恩, r12725031@ntu.edu.tw or upload to NTU COOL by **5pm, Apr. 11, 2025**). The project should follow the project instruction and write a brief project report (around 10 pages including figure and table) and give a 10-min presentation (around 12 slides) including 30 seconds video demo (資訊系統介面操作與分析結果視覺化) and 3-min Q&A on **June 6, 2025 in class (ALL students should join the course full-time from “8am-1pm” on June 6)**. We will follow the grading policy “前一組問後一組問題，以及同儕互評(其他組別評分去頭去尾後取中位數)”. The report violating the project instruction is not accepted. Please “zip” project report (eg. Word), slides (eg. PPT), **dataset** (eg. csv/txt), code (eg. Python), video (eg. mp4) with file name “MDS_Final_Project_NAME1_NAME2_NAME3_NAME4_NAME5.zip” to NTU COOL before **5pm, June 10, 2025**. Late project report is not accepted.

Open dataset:

1. 政府資料開放平台 (<https://data.gov.tw/>)
2. Kaggle (<https://www.kaggle.com/datasets>)
3. KDnuggets (<http://www.kdnuggets.com/datasets/index.html>)
4. UCI Machine Learning Repository: Data Sets (<https://archive.ics.uci.edu/ml/datasets.html>)
5. Others...

You can use the real dataset; however, do NOT use the confidential dataset, or bear your own consequences. (請勿使用任何會違反公司保密協議之資料).

Content and Format:

1. Title

Give a title to your project work, eg: Process Diagnosis and Troubleshooting in the High-Tech Industry, Demand/Price Forecasting and Capacity Planning in the Petrochemical Industry, Scheduling Optimization in the Fastener Manufacturer, Queueing System Improvement in McDonald, etc.

2. Background and Motivation

2.1 Motivation: Describe your motivation and why you choose this topic. Why this problem/decision is significant to our society or important to us.

2.2 Background: Describe the problem background or context. eg: how/why the process diagnosis problem benefits us? how the scheduling optimization change the society?

2.3 Problem Definition: Give one or two sentences to define your problem clearly. eg: This study uses data science technique to identify the key variables and classify the good and defective products through addressing data imbalance problem. This study applies genetic algorithm to find the best scheduling under the machine capability constraint.

3. Methodology

3.1 Give a research framework to describe your analysis flow. The framework may include (but not necessary) data preparation (eg. missing value Imputation, outlier detection, data transformation), feature selection (eg. principle component analysis, stepwise selection, clustering), prediction model (eg. regression, decision tree, neural network), pattern interpretation and prediction accuracy, decision optimization (eg. linear programming), and insights and conclusion.

3.2 What is the method you choose to analyze your work? eg: association rule, decision tree, clustering analysis, neural networks, queueing theory, etc. Why you choose this method? Is this method fit your problem? (Method Justification)

4. Data Collection and Analysis Result (Video demo in the class presentation)

4.1 Data Collection: Describe the data source and how your get this data.

4.2 Analysis: Use the method introduced in Section 3 to analyze your work. Show your calculation/procedure in details (eg: Python/package). Please use **table or figure** to illustrate your numerical result if possible (i.e., **Visualization**).

4.3 Interpretation: what's the result or the best alternative you suggest? What's the prediction accuracy? How to support your decision-making based on the prediction results? What's the **risks** you face when your prediction is inaccurate? What's the **insight or story** behind the pattern you found? (IMPORTANT)

5. Conclusion

Summarize and guide the future research of your project work.

Finally, please remember this project is to let you know how to “implement” the knowledge you learn in class to real-world setting. **The original or innovative idea is encouraged.** You will learn how to find an **interesting and significant** problem from our society, and rebuild the real problem to the quantifiable and analyzable problem in this project. Show your best work and good luck.

Thank you for joining MDS class and wish you have a wonderful Summer Break 2025!

大家辛苦了~ 加油! 希望是個很充實的一學期~