



BUSS2505: Machine Learning

Course Project Guideline

Spring 2025

Instructor: Chengzhang Li (李成璋)
E-mail: cz.li@sjtu.edu.cn

The course project is intended to achieve one of key learning objectives of this course. That is, to prepare you to apply machine learning algorithms to analyze the datasets from the real-world problems and applications, and develop your skills to extract information, make predictions/decisions, and generate managerial insights using the available data.

Topics

Each team should pick a project topic and apply the techniques that you learn in this course to solve those real-world tasks. You will find an application setting that interests you or is related to your major, and explore how to best perform the prediction tasks with machine learning algorithms. You can use any Python library to complete the course project. Here are some directions that you can further explore to enhance the contribution of your project:

- *Application-centric*: the problem is new and challenging. In this case, you need to define the problem, collect and clean the data. Moreover, you may generate some insights from the learned models, and provide implications for practitioners.
- *Algorithm-centric*: the problem can be a classical and standard task. You may think about designing your own learning algorithm, or propose a novel variant of existing ones. The proposed ones could be more effective, i.e., better prediction outcomes, or more efficient, i.e., less running time.
- *Implementation-centric*: You don't implement the machine learning algorithms with the libraries such as SK-Learn, TensorFlow, PyTorch, etc. and prepare the code from scratch to build and learn the machine learning models.

Project teams

Student groups of 5 people will be formed. Although the preferences of individual students will be honored where possible, the instructor reserves the right to alter the team membership by adding or deleting team members to assure that all teams have roughly the same number of people. Students who are unable to form a team of an appropriate size should contact the instructor who will assure that all students are members of a team.



Key Elements

Through this semester-long project, you will work on the following elements one by one to complete the project.

- Problem description
- Data collection & preparation
- Feature engineering
- Model formulation
- Algorithm implementation
- Model evaluation
- Implications and insights

Informal Project Reports

To make sure that you start with a *relevant* and *tractable* topic and do not get stuck during the project, you will submit to two informal reports possibly followed by online or offline discussions.

- **Project proposal** (1 page): you should briefly describe the problem that you try to solve, the data that you need to collect or the dataset that you plan to work on, the machine learning algorithms to use, the experiments that you plan to conduct.
- **Mid-term report** (1-2 pages): describe the status of your project, you may want to report the progress, the challenges that you meet, the preliminary results if any, what you plan to do the next step.

Evaluation

The project evaluation is based on the **final project report**, the **final presentation**, and the **peer evaluation**.

- A complete report that includes all the elements of your project. The main text should be **no more than 15 pages**. You can include additional numerical results, tables, references, etc. in the appendix. You need to present the title of your project and the team members' names on the cover of your report.
- Each team will have 15 min to present their course project and 3 min Q&A. The presentation can be delivered by the representative(s) or all members. All team members should be prepared to answer the questions.
- The inter-group peer evaluation and intra-group peer evaluation forms will be distributed.

Disclosure policy on Generative AI

Students are permitted to use generative AI tools (e.g., ChatGPT, DeepSeek) as part of their course projects. However, any use of such tools must be clearly disclosed in the project submission. Students should specify which tools were used, how they were applied, and to what extent the output was directly incorporated into the final work.



Datasets

Here are some popular online resources for datasets. You can also crawl the data by yourself or download the dataset from other open-source websites.

- UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets.php>
- Kaggle datasets: <https://www.kaggle.com/datasets>