

Project Report

Group Effort

SEIS 763-03

Spring 2025

Note that this project handout only provides broad requirements and guidelines while enabling sufficient flexibility in the course project.

This is not meant to be fully prescriptive, each team should consult with me on a regular basis ensure that the work is relevant and deserving of an appropriate grade.

You have the flexibility to make your own design decisions, but clearly document each decision and its rationale, and include these in the final project report.

1 Project Specifics

Consider the following when writing the project proposal:

- The class project will involve the application of Machine Learning techniques discussed in class to a dataset of your choice.

Numerous data sets available online. Pick something that you will enjoy working on. The only requirement is that the dataset needs to have at least **3000** instances before class balancing.

- Consider your dataset carefully. Even if you downloaded it, you should look for information about it.

How were they collected? What are the data quality issues? Are there biases inherent in the person or technique that collected this data? How might this impact subsequent conclusions?

A good report will dive into biases that may exist or data quality issues.

- Formulate questions that you would like to answer about this data set. What is the dependent variable? What are the predictors?
- Implement your analysis using machine learning techniques. These should have some relation to what we have learned in the class.

Are you doing a regression or classification task? Are there interesting visualizations to do? How will you evaluate the performance of your model, or choose between competing models?

2 Project Proposal

Your proposal should be approximately 1 page long, single spaced. The purpose of the proposal is to make sure that you are on the right track and to give me enough information so that I can give you useful feedback.

Details about the proposal will be available on [Canvas](#).

This is a proposal, and I expect that your question and your approaches will likely change as the class progresses.

3 Project Report

The project report is the most crucial component of your term project. Ensure that you document your experiments, analyses, design decisions, and results throughout the semester so that you can compile them into a relevant report.

- Write a report summarizing your data, your question of interest, and your findings.
- Reference other existing work which has analyzed your data or addressed similar topics. If you borrowed an existing approach, ensure that you cite it.
- The report should at minimum include specification of the problem (based on application objectives), data collection, data preparation, machine learning and statistical techniques used, interpretation of results, and conclusions. More information is detailed in the next section.

The report should be formatted based on the [IEEE Conference](#) format and 6-8 pages long (without any code).

- [IEEE Word Format](#)
- [IEEE L^AT_EXformat](#)

4 Evaluation Criteria

4.1 Experiments, Results, and Discussion

This component will be evaluated based on the size of the dataset, whether extensive experiments were performed, convincing results obtained, etc.

Although this is not an exhaustive list and will change based on your dataset, most experiments and results should contain:

- Data pre-processing steps (data splitting, scaling, imputing, etc)
- Class balancing (if you are doing classification).
- Start with all features of the data. If any features are eliminated manually, provide suitable justification.
- Use feature selection (backward elimination or forward selection).

- Feature engineering steps (if any).
- Dimensionality Reduction:
 - PCA, LDA, and Kernel PCA (for classification).
 - PCA and Kernel PCA (for regression).
- You should perform k-fold cross validation for all the experiments and hyperparameter tuning.
- Comparative analysis on different models with discussions on difference in performance.
- Present relevant results and metrics in tabular and/or graphical form.
- Discussion and analysis of results.

4.2 Presentation

Preferably in-person.

- You will be presenting to the class explaining the dataset, models built, and results obtained.
- The entire team needs to be present for the presentation.
- *Ideally*, all members must present a portion of the work. Each member must be able to answer questions related to the project.
- Presentation will be a maximum of 20 minutes including Q&A.
- Do not include code in the presentation.

4.3 Project Report

In addition to the above, overall quality of the project report will be used as criteria for evaluation. The report should be a stand alone document with all the required information.

I would highly recommend reading a few machine learning papers to understand general formats and writing practices. Pay special attention to how results are presented and discussed.

4.4 Code

You should also include your code and dataset. An executable `Jupyter` notebook must be included and I should be able to run it end to end.

Ensure that the notebook is well documented and commented.

5 Final Submission

- Upload your final submission as a zip file with naming convention : `group-num--project_report.zip`
- Your zip file must contain the following at the same directory level :
 - Project report in `pdf` format : `project_report.pdf`
 - An executable Jupyter Notebook of your submission : `notebook.ipynb`
 - `HTML` export of Jupyter Notebook for viewing in browser : `notebook.html`
 - Raw data used for your analysis within a subfolder called `./data`.
 - Any helper code should go in a `./utils` folder.

Here is an example submission `group-1--project_report.zip` with folder structure, for **Group 1**. Unzipping the submission file would yield the following directory structure.

```
group-1--project_report
├── project_report.pdf
├── notebook.ipynb
├── notebook.html
├── utils
│   ├── python_file1.py
│   └── python_file2.py
├── data
│   ├── data_file1.extension
│   └── data_file2.extension
```