Project Proposal
Group Effort

SEIS 763-03

Spring 2025

1 Goals

- Identify and describe dataset for term project.

- Identify preliminary project goals and hypothesis.

2 Proposal Guidelines

The goal of the proposal is not to come up with an exact script and follow it throughout. Instead, consider this a framework and deviations from it are not and even expected.

Consider the following when writing the project proposal:

- **Dataset Size:** What data are being used - where do the data come from, and what are some characteristics (size, missing values, continuous vs. categorical).

  You should have a minimum of `3000` instances.

  Note: Do not use the *wine dataset*. Available on `Kaggle` and other resources.

  Note: For the purpose of this class you *can* use Kaggle datasets. If your aim is to use this work on your resume - widely used datasets *may* not be the best idea.

- **Task Type:** Classification v/s Regression.

- **Class Imbalance:** If you are doing classification, see whether the classes are appropriately balanced.

  If applicable, come up with a strategy to overcome the class imbalance.

  Avoid datasets that are very highly imbalanced (more that approximately `75/25` imbalance).

- **Attributes:** It is preferred to have a dataset that has more numerical attributes compared to the categorical ones.

  Avoid picking a dataset that has too many categorical attributes.

- **Exploration:** Before finalizing a dataset, identify a few candidates and perform exploratory data analysis on them. Look for any anomalies, skew, missing values, outliers, etc.

  You can include those exploration results in your project proposal.

  If the dataset requires too much pre-processing, it may not be the best fit for the project.

- **Hypothesis:** What is the question(s) of interest - be specific.

  You should avoid generic questions like *"I want to look for patterns in stock prices"*.

  Devise specific question(s) you can answer with the data.

- **ML Techniques:** What methods do you plan to use?

  Note that this might change and that we have yet to cover many methods in class.

3   Submission Guidelines

- Any one person from the team can submit the project proposal. Submit a `pdf` file.

- Naming convention: `group-num--project_proposal.pdf`.

- Eg. Group 1, `group-1--project_proposal.pdf`