# Term project

# ML workflow

**Problem scoping**

**Experimentation**

**Deployment (not covered in this course)**

# Problem scoping

**Define business problem**

- What are you trying to solve?

- What is the business value of solving this problem? Why is it important?

**Define ML problem**

- Type of ML problem: **supervised learning** (regression or classification)

- Model: linear regression, logistic regression, neural network, random forest, etc.

- Data: data sources, features, labels

- Goal: performance metrics, performance target

# Experimentation

**Collect and process data**

**Build and train model**

**Evaluate model**

- How to improve model based on diagnosis?

- more data (new data), feature selection and engineering, regularization, model selection, etc.

**Repeat**

# Term project: Build supervised ML model

Follow the ML workflow

You can choose any dataset

You can use any ML library (scikit-learn is recommended)

You can use any ML models covered in the course

# Presentation format

**Date: December 2nd and 4th**

**Time: 20 minutes per group**

1. presenting group: 15 minutes presentation
2. discussant group: 3 minutes discussion (**one-page discussion slide**)
3. presenting group: 2 minutes response and Q&A from the audience

# Project deliverables

**Proposal**

**Presentation slides (November 29th)**

**Discussion slides (December 1st)**

**Final report (December 4th)**