

## Final Project Requirement and Datasets

The assignment is worth 45% of your final grade.

### Why?

This project is to explore an interesting machine learning problem of your choice in the context of a real-world data set. Your class project must be about new things you have done this semester; you can't use results you have developed in previous semesters.

We hope from Mid-term project you have learned how to explore dataset in the wild and apply machine learning to gain valuable insights. In this final project, you can team up with other students and explore ideas the team agrees upon. It works like this:

1. You form a team of no more than 3 students.
2. In phase 1, the team agrees upon an idea and submit one-page proposal for review. The proposal shall have all team members' name and NEU.
3. In phase 2, your team will implement the idea with our feedback. We encourage you to try new ideas, as we will grade on your effort that you can demonstrate. If you want to go for a Kaggle competition, go for it!

### Requirement

- No deep learning. I know it is fun, but you shall learn to walk first.
- Projects can be done individually, or in teams of no more than 3 students. For a two- or three-person group, group members are responsible for dividing up the work equally and making sure that each member contributes.
- If you are having trouble writing a proposal, feel free to consult with a TA or the instructor. Once you submit your proposal, we will give you feedback. Of course, the final responsibility to define and execute an interesting piece of work is yours.

### The Problems Given to You

Easy Idea	Intermediate Ideas
1. Cartoonify Image with Machine Learning	1. Music Genre Classification Machine Learning Project
2. Iris Flowers Classification Project	2. Bitcoin Price Predictor Project
3. Emojify – Create your own emoji with Python	3. Uber Data Analysis Project
4. Loan Prediction using Machine Learning	4. Personality Prediction Project
5. Housing Prices Prediction Project	5. Handwritten Character Recognition
6. MNIST Digit Classification Machine Learning Project	6. Xbox Game Prediction Project
7. Stock Price Prediction using Machine Learning	7. Credit Card Fraud Detection Project
8. Titanic Survival Project	8. Sign Language Recognition
9. Wine Quality Test Project	9. Barbie with Brains Project

10. Fake News Detection Project	10. Customer Segmentation
---------------------------------	---------------------------

From <https://data-flair.training/blogs/machine-learning-project-ideas/>

You can find other ideas and dataset by web search:

<https://www.google.com/search?q=machine+learning+research+project+for+beginners>

## What you need to submit

### Phase 1: Project Proposal

Read our feedback on proposal. If you still don't get how to write a proposal, consult our TA.

**Page limit:** Proposals should be one page maximum.

Include the following information:

- Project title
- Data set
- Project idea. This should be approximately one paragraph.
- Software you will need to write.
- Articles to read. Include 1-3 relevant articles including scientific journal papers, weblogs or Kaggle reports. You will probably want to read at least one of them before submitting your proposal.

### Phase 2: Project report, presentation, and Code

The report should be a **5-6 pages** report in the form of a [NeurIPS](#) paper. You can also use the word document provided by instructor. It consists of background, method, experiment, conclusion, and references. The sections on the experiments and conclusions will have whatever results you have obtained. If you expect to continue working on it after the course (say Kaggle competition), you can have a discussion section which covers you plan or results you hope to obtain.

The presentation is a 5-min recorded video given by one member on behalf of the entire team.

The code shall be runnable and can product the charts, figures, numbers used in the report.

Grading scheme for the project:

- 20% for Data acquisition and Exploratory Data Analysis.
- 40% for proposed method and experiments conducted to obtain insights from the data.
- 20% on Data Visualization including graphs and charts.
- 10% on oral presentation. Is it clear and concise? Does it cover all the important findings?
- 10% on code. Is it runnable? Can it produce the same charts as shown in the report?