

**San José State University**  
**Computer Science Department**  
**CS156, Introduction to Artificial Intelligence, Fall 2021**

**Term project**

## Objective:

Select an interesting AI problem and develop a machine learning solution for it.

## Details:

Select and download a dataset of interest. This dataset will be used to train your prediction model and to solve some interesting AI problem. You can browse the following sources for free datasets but are not limited to them:

- Kaggle datasets (<https://www.kaggle.com/datasets>)
- UC Irvine ML dataset repository (<https://archive.ics.uci.edu/ml/index.php>)
- Physionet (<https://www.physionet.org/about/database/>)
- Scikitlearn datasets (<https://scikit-learn.org/stable/datasets/index.html>)
  - For your project you **cannot** use one of the datasets we have used in class so far this semester

You can use any number of other sources for publicly available data for your dataset of choice.

In the dataset you selected and downloaded select the independent variables and the dependent variable. Remember that independent variables are those you are going to use to train your solution model. The dependent variable is often called the response variable and it is the variable your solution model will be predicting. Notice what types of variables you are dealing with. Are your independent variables all numeric, all categorical, a mix of the two? Does your data consist of free text (an NLP problem)? Is the dependent/response variable numeric or categorical? This will guide you into establishing if you are dealing with a regression or a classification problem.

Based on the information you observed about the dataset, make a decision about the type of ML algorithm you want to use. If you would like to compare several ML algorithms then make a decision about which one you would like to use and which hyperparameters you are going to use.

Use the project submission template Jupyter notebook named TermProjectSubmissionTemplate.ipynb to write all the code and produce visualizations for this project.

You will use the power point presentation file named FinalProjectPosterTemplate.pptx to create a poster presentation for your project for the final submission. This template is only a suggestion.

## Term project

Feel free to change the style and section placements as you see needed. Feel free to add sections as well. Your poster, however, must contain the following sections:

- Introduction/Background
- Dataset description
- Methodology
- Analysis and results
- Summary/Conclusions
- Key References

Please make sure that your final project poster presentation is complete and clear and is visually pleasing. Imagine you have to present this poster at a conference or display it at work on a wall. Make sure the text is clear, with no spelling or major grammatical mistakes. Make sure the plots/figures and tables in your poster are clearly labeled and numbered, with brief description of what the figure/table represents (e.g. “This confusion matrix shows the true vs. predicted observation counts in each class, with overall accuracy of 90%.”).

Save/print your poster as pdf for your final submission.

## Submission:

Two submissions during the semester will be made for this project:

- Preliminary report on the chosen problem and selected dataset (word document or pdf of your preliminary report)
- Final poster presentation and Jupyter notebook for the project (poster presentation pdf file, Jupyter notebook .ipynb file, and Jupyter notebook pdf file)

For each of the submissions (preliminary and final), email your submission to me at [Yulia.Newton@sjsu.edu](mailto:Yulia.Newton@sjsu.edu) and the grader (Gursimran Singh) at [gursimransingh@sjsu.edu](mailto:gursimransingh@sjsu.edu). Make sure to email this submission by 11:59pm on the due date listed in Canvas. Your sent email is the proof of submission. The subject of the email should say “CS156 term project”. In the body of the email list your name as it appears on the class roster and your student ID. Attach to this email both the pdf of your Jupyter notebook, which contains the solution for this homework assignment, as well as the notebook itself (the notebook file with .ipynb extension). Make sure to submit both files, otherwise the submission will not be considered complete.

## Grading:

A total of 35 points are possible for this homework assignment:

- 10 points for the submitted and complete preliminary report
- 10 points for the submitted project code
- 15 points for a complete, clear, and visually presentable poster presentation