

CS 412 - Machine Learning

Project Assignment — Fall 2021-2022

Group Formation: Thursday, December 17, 2021, 23:55

Due Date: Sunday, January 9, 2022, 23:55

Late Submission: Late submissions will **not** be accepted.

1 Goal

The purpose of the project is to increase your knowledge about machine learning and get hands-on practical experience.

2 Projects

You have a choice of 2 ML projects that involves building a classification or regression model for one of the given problems. There will be two competitions for the two projects, from which you will choose one.

2.1 Project I - Car Price Prediction

Try to predict the selling price of a car, given 12 features. The data consists of information of 60,000 cars and 12 features such as make, model, age,

Features:

- ID (Unique id number for a given sample)
- brand
- model
- year (year of production)
- transmission
- mileage (how many miles has this car been used)
- fuelType
- mpg (fuel consumption; miles per gallon)
- engineSize
- tax
- **Price (target label)**

Kaggle Link: We have setup a Kaggle competition for this task at the following [link](#).

2.2 Project II - Digit Prediction on the Wild

You have worked on the MNIST digit dataset, now build Machine Learning models to recognize the SVHN(street view house numbers) dataset. You will work on the MNIST-like 32x32 images extracted from house number sequences (format 2 version of the dataset), in this project. But in this task, in addition to the center digit that you need to recognize, there may be other digits to either side.



Figure 1: Sample images from the data

Kaggle Link: We have setup a Kaggle competition for this task at the following [link](#).

3 Kaggle Evaluation

- Performance metrics will be Coefficient of determination for Car Price Prediction and Accuracy for Digit Prediction.
- You can submit your model predictions for the *test set* to Kaggle in order to observe the performance of your model.
- Your model's performance will be calculated based on 50% of the test set and will be shown in the "public leaderboard" (will be shared with you to guide you to good models). Your models' performance on the segregated portion (the remaining 50% of the test set) will be calculated in the "private leaderboard" (will not be shared).

- Each group can do at most 3 submissions per day. This will form the basis of the Public Leader board where you can see your standing until January 9, 2022.
- Your Kaggle score/rank will be based on the average of your results on these the Public and Private leaderboards.

4 Terms and Rules

- You will work in groups of 2 to 5. Please indicate your groups in the Google [sheet](#). The teams on Kaggle must be the same as on the sign up sheet.
- Only one person for each team will form a team on Kaggle site and invite the team members to join.
- Name your team as follows: "SU_GroupX", e.g. if you are team 11, your team name should be "SU_Group11".
- You may use any approach, but all work must be yours, except possibly for generic (ImageNet), pre-trained models etc.
- You will have to submit your submissions until (January 9, 23:59) and late submissions will not be accepted.
- You are expected to use Python 3 notebooks and submit only the .ipynb or .py codes, no other format will be accepted.
- Any violation of the above rules will result in large penalty points.

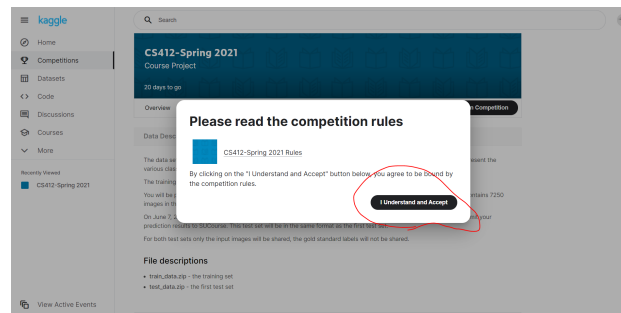
5 What to Submit

- Kaggle submission (see below),
- A report as a [pdf](#) file, which should be a 2 or 3-page report explaining your approach and findings and should be submitted along with your code and *test set* predictions, by the project closing time.
- You can either use Colab Notebooks or Jupyter Notebooks if you wish to work locally on your computers, however it is recommended for you to use Google colab as you can see and track your teamwork more easily.
- If you use Google colab Notebooks, your report should contain the link for all Google colab notebooks that you worked on for this project. If you use Jupyter Notebooks, you should submit all them on the SuCourse as .ipynb files.

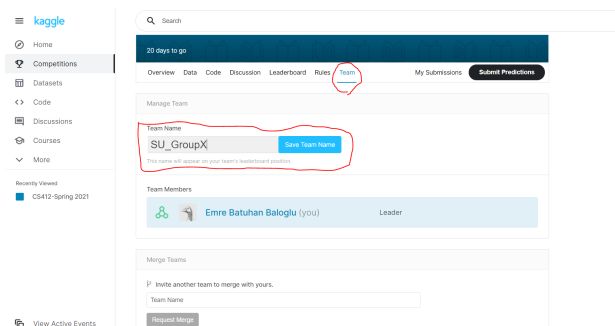
- Your code should be clear and well commented, outputs should also be run, for us to check your work.

6 Kaggle Submission Instructions

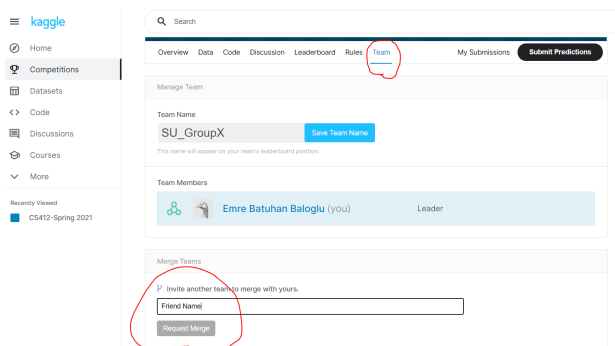
- Join one of the competitions by clicking the button and accept the competition rules (all group members should do this):



- One of the group members should go to the "team" tab and rename your team as instructed.

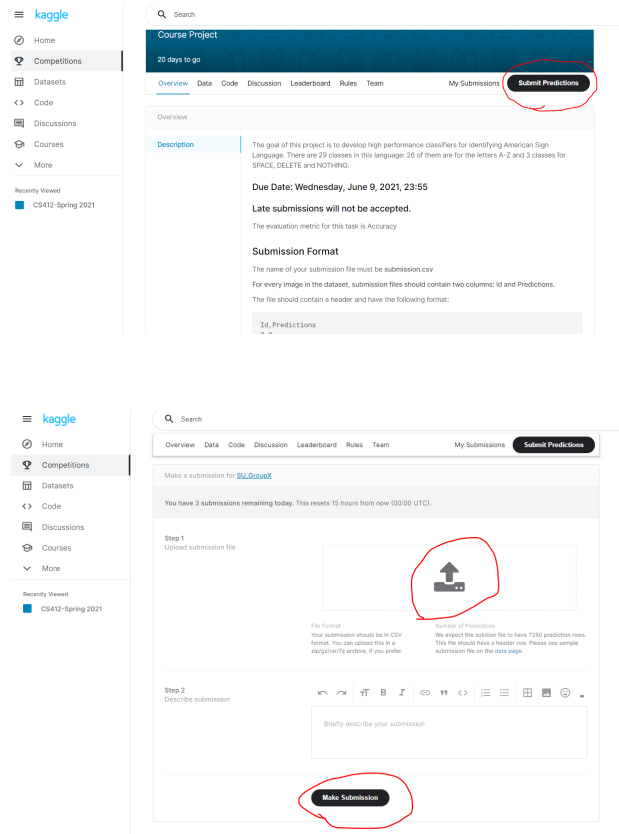


- That member should invite other group members using the "Merge Teams" section in the teams tab.



- After training your model, you should submit the predictions of the model for the test set that is provided at that time (*test set* before January 9, 2022.)

CS 412 - Machine Learning: Project Assignment — Fall 2021-2022



- Your submission file **must** be a .csv file, having the same number of predictions as the size of the test set. Name your file as "submission.csv". Also you must name the columns as "Id" and "Prediction" for Digit Prediction on the Wild competition. You must name the columns as "ID" and "price" for Car Price Prediction competition. (both are case sensitive)

	A	B
1	Id	Prediction
2		0 X
3		1 X
4		2 X
5		3 X
6		4 X
7		5 X
8		6 X
9		7 X
10		8 X
11		9 X
12		10 X
13		11 X
14		12 X
15		13 X
16		14 X
17		15 X
18		16 X
19		17 X
20		18 X
21		19 X
22		20 X
23		21 X
24		22 X
25		23 X
26		24 X

7 Grading

You will have a small quiz at the end about your project about the work done in your project. Your grade will be based on the performance (Kaggle score/rank) you obtain (50%), your report (30%) and your quiz score (20%).