

Please find all instructions needed to complete your assignment below

General guidelines

1. All coding must be done in python 3 in a jupyter notebook.
2. All code from external sources should be cited.
3. Please explain your thought process clearly using comments.
4. Final Submission will be a jupyter notebook in .ipynb format and .html format.
5. You may use any tools, github repos, python libraries etc but please make sure to cite them clearly.
6. You may use any computational resource for completing these tasks.
7. Please ensure that your notebook runs before submitting. Also indicate the conda and python package versions used as a requirement.txt file or within the notebook itself.
8. We do not expect you to spend more than 3-4 hours on this assignment.
9. You may use free tier google cloud, azure or colab notebooks for doing this assignment. Kaggle also offers free gpu.
10. The datasets used for the assignment are open source and plenty of work on them could be found online. We basically would like to see how you approach the tasks below.
11. All trained models should be attached with the submission
12. A single Zip file submission with all notebooks, models should be sent to the contact below. You may use the same email below for any clarifications.

What we are looking for ?

1. Simplicity and innovation.
2. Data exploration and explanation.
3. Accuracy and results matter less as compared to your approach.
4. Explain the metrics that you use clearly.
5. Clearly explain your results.
6. Coding skill.
7. Follow coding standards (pep-8).
8. Explain your models or why you decided to use them.

Datasets

You are provided with 2 datasets

- A) Credit card fraud detection dataset (Classification Task)
- B) Stanford Car Datasets

Task 1

The credit card fraud detection dataset [Standard kaggle dataset]. You may reference existing kaggle notebooks. We are looking for how you approach this problem rather than achieving top accuracy.

<https://www.kaggle.com/mlg-ulb/creditcardfraud?select=creditcard.csv>

Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise. More description of the dataset could be found in the kaggle link above.

Task description

- a) **Train a classifier to predict if a transaction is a fraudulent transaction or not ?**
- b) **How would you deploy this model in production? (show example or link to an approach that you would like to follow)**
- c) **If you had more time, how could you improve this model? (No code)**

Task2

You are provided with the Stanford Car Dataset. “The *Cars* dataset contains 16,185 images of 196 classes of cars. The data is split into 8,144 training images and 8,041 testing images, where each class has been split roughly in a 50-50 split. Classes are typically at the level of *Make, Model, Year*, e.g. 2012 Tesla Model S or 2012 BMW M3 coupe”

https://ai.stanford.edu/~jkrause/cars/car_dataset.html].

The dataset is separated as train and test. Each sub-folder in the train/test folder represents a class and includes all images.

Anno_train.csv and Anno_test.csv have the data in the below format

Image file, 4 bounding boxes coords and last being class number

00001.jpg 30 52 246 147 181

Task description

- a) Train an image classifier to predict the Make model year of a car in an image. (code)
- b) Briefly also explain how you could deploy your model in a production environment. (No code)
- c) If you had more time, how could you improve this model? (No code)
- d) Use any existing pretrained models to draw bounding boxes for all the cars in the Image (we do not recommend training from scratch for this task)



All submissions should be made as per the guidelines on page 1. Please submit the models along with the submissions.