Major Project

you can chose any dataset online and problem related to classification, regression or GAN. You have to submit a report (pdf format) and code (google colab notebook. ipynb format).

Steps:

1. Find a dataset online (see the "Where to Find Datasets" section below)
2. Understand and describe the modeling objective clearly
    1. What type of data is it? (images, text, audio, etc.)
    2. What type of problem is it? (regression, classification, generative modeling, etc.)
3. Clean the data if required and perform exploratory analysis (plot graphs, ask questions)
4. Modeling
    1. Define a model (network architecture)
    2. Pick some hyperparameters
    3. Train the model
    4. Make predictions on samples
    5. Evaluate the test dataset
    6. Save the model weights
    7. Record the metrics (loss, accuracy per epochs)
    8. Try different hyperparameters & regularization
5. Conclusions - summarize your learning & identify opportunities for future work
6. Publish and submit your Jupyter notebook
7. Write a report to describe your experiments and summarize your work.

Notebook:

* code should be in .ipynb (notebook format)

* All the plots and results should be plotted in proper format and should be visible (inculde title, xlabel, ylabel and legend)

* use markdown or text cell to explain code in jupyter notebook itself

* explain your model also.

* use multiline comment and single line comments to explain working of any function you include or defined

Report

* in pdf format

* should explain about the dataset, type of data and what is your problem and how you are solving it.

* should include image or table or plot

* model architecture

* reference and links of blog, paper and code you have used.

* codes with explanation

* output accuracy and loss etc.

* if you used multiple model and please explain comparison

List of dataset (you can chose dataset from other source also, please take care of size of dataset. Large dataset may not work in colab)

Indian stocks data

- https://nsepy.xyz/
- https://nsetools.readthedocs.io/en/latest/usage.html
- https://www.kaggle.com/rohanrao/nifty50-stock-market-data

Indian Air Quality Data

- https://www.kaggle.com/rohanrao/air-quality-data-in-india

Indian Covid-19 Dataset

- https://api.covid19india.org/

World Covid-19 Dataset

- https://www.kaggle.com/imdevskp/corona-virus-report

USA Covid-19 Dataset

- https://www.kaggle.com/sudalairajkumar/covid19-in-usa

Megapixels Dataset for Face Detection, GANs, Human Localization

- https://megapixels.cc/datasets/ (Contains 7 different datasets)

Agriculture based dataset

- https://www.kaggle.com/srinivas1/agricuture-crops-production-in-india
- https://www.kaggle.com/unitednations/global-food-agriculture-statistics
- https://www.kaggle.com/kianwee/agricultural-raw-material-prices-19902020
- https://www.kaggle.com/jmullan/agricultural-land-values-19972017

India Digital Payments UPI

- https://www.kaggle.com/lazycipher/upi-usage-statistics-aug16-to-feb20

India Consumption of LPG

- https://community.data.gov.in/domestic-consumption-of-liquefied-petroleum-gas-from-2011-12-to-2017-18/

India Import/Export Crude OIl

- https://community.data.gov.in/total-import-v-s-export-of-crude-oil-petroleum-products-by-india-from-2011-12-to-2017-18/

US Unemployment Rate Data

- https://www.kaggle.com/jayrav13/unemployment-by-county-us

India Road accident Data

- https://community.data.gov.in/statistics-of-road-accidents-in-india/

Data science Jobs Data

- https://www.kaggle.com/sl6149/data-scientist-job-market-in-the-us
- https://www.kaggle.com/jonatancr/data-science-jobs-around-the-world
- https://www.kaggle.com/rkb0023/glassdoor-data-science-jobs

H1-b Visa Data

- https://www.kaggle.com/nsharan/h-1b-visa

Donald Trump’s Tweets

- https://www.kaggle.com/austinreese/trump-tweets

Hilary Clinton and Trump’s Tweets

- https://www.kaggle.com/benhamner/clinton-trump-tweets

Asteroid Dataset

- https://www.kaggle.com/sakhawat18/asteroid-dataset

Solar flares Data

- https://www.kaggle.com/khsamaha/solar-flares-rhessi

Human face generation GANs

- https://www.kaggle.com/arnaud58/flickrfaceshq-dataset-ffhq

F-1 Race Data

- https://www.kaggle.com/cjgdev/formula-1-race-data-19502017

Automobile Insurance

- https://www.kaggle.com/aashishjhamtani/automobile-insurance

PUBG

- https://www.kaggle.com/skihikingkevin/pubg-match-deaths?

CS GO

- https://www.kaggle.com/mateusdmachado/csgo-professional-matches
- https://www.kaggle.com/skihikingkevin/csgo-matchmaking-damage

Dota 2

- https://www.kaggle.com/devinanzelmo/dota-2-matches

Cricket

- https://www.kaggle.com/nowke9/ipldata
- https://www.kaggle.com/jaykay12/odi-cricket-matches-19712017

Basketball

- https://www.kaggle.com/ncaa/ncaa-basketball
- https://www.kaggle.com/drgilermo/nba-players-stats

Football

- https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017
- https://www.kaggle.com/abecklas/fifa-world-cup
- https://www.kaggle.com/egadharmawan/uefa-champion-league-final-all-season-19552019

Where to Find Datasets

General sources:

- https://www.kaggle.com/datasets (use the opendatasets library for downloading datasets)
- https://course.fast.ai/datasets
- https://github.com/ChristosChristofidis/awesome-deep-learning#datasets
- https://www.kaggle.com/competitions (check the "Completed" tab)
- https://www.analyticsvidhya.com/blog/2018/03/comprehensive-collection-deep-learning-datasets/
- https://lionbridge.ai/datasets/top-10-image-classification-datasets-for-machine-learning/
- https://archive.ics.uci.edu/ml/index.php
- https://github.com/awesomedata/awesome-public-datasets
- https://datasetsearch.research.google.com/