Course Project

Due: April 27

Assignment submission policy.
1.Submit the results in a pdf file.
2.All assignments should be submitted by the due date. There will be 20% penalty per day for late submissions. No grade will be given to homework submitted 5 days after the due date.
3. Each group prepares a final report as a team. Collaboration within a group is encouraged. Each group member must upload a copy of the final report individually.

Each group selects one task from the following 3 tasks and develops machine learning models to solve it.

A. COVID-19 Detection (Dataset link)

Corona - COVID19 virus affects the respiratory system of healthy individual & Chest X-Ray is one of the important imaging methods to identify the corona virus.

With the Chest X - Ray dataset, develop ML Models to classify the X Rays of Healthy vs Pneumonia (Corona) affected patients. The input and output of the ML model are:
Input: X-Ray Image
Output: Normal/ Abnormal

B. Traffic Prediction (Dataset link)

The Smart Mobility and Traffic Optimization Dataset integrates data from cyber-physical networks (CPNs) and social networks (SNs) to enhance intelligent traffic management and smart mobility solutions. It includes real-time traffic patterns, vehicle telemetry, ride-sharing demand, public transport efficiency, social media sentiment, and environmental factors.

You will train machine learning models for traffic congestion prediction based on this dataset. The input and output of the ML model are:
Input: Traffic features: Vehicle count, speed, road occupancy, and traffic light status. You can also add other features in the dataset.
Output: Traffic Congestion Level: Categorized as Low, Medium, or High, based on traffic density, speed, and road occupancy.

C. Stock Market Prediction (Dataset link))

This dataset contains historical daily prices for all tickers currently trading on NASDAQ. The up to date list is available from nasdaqtrader.com. The

historic data is retrieved from Yahoo finance via yfinance python package.

You will train machine learning models for stock price prediction based on this dataset. Inputs and outputs of ML models are given below. Input and output of ML models are given below.
Input: Historical stock price.
Output: Future stock price.

Final Report Requirements

Each group needs to submit the original codes and a final report. The final report should be **at least 3 pages** written in Word Template or Latex Template for ACM SIG conferences. The final report should include the following **sections**

- Group members and Individual contributions
- Introduction and problem description
- Literature review (Summarize the existing studies regarding the selected topic.)
- Machine learning models, methods, or algorithms
- Experiment Results (Report training loss, testing accuracy, and other metrics which can evaluate the performance of the implementation)
- Conclusion (Summarize the key contribution)

Grading criteria

The final report will be graded in terms of the following aspects:

- Integrity (20 pts): Include all the sections metioned above (10 pts). Meet the page requirement (at least 3 pages in Word Template or Latex Template) (10pts).
- Clarity (20 pts): The introduction section has clear problem description (10 pts). The "ML models, methods, or algorithms" section has clear definition of ML models and training/testing methods (10 pts).
- Literature review (20 pts): Has literature review section (10 pts). Summarize 3 or more related papers or studies. (10 pts)
- Results (30 pts): Clearly show the simulation results via tables, figures, or other forms (10 pts). Add 5 pts if one of the following results is added.
  (1) Implement an advanced ML model (DNN, CNN, RNN, transformer, or LLM, etc);
  (2) Plot the training loss trajectory (How does the training loss change with No. of iterations?);

(3) Explain whether overfitting or underfitting issues exist and discuss how to solve these issues.
(4) Ablation studies showing the effects of different hyper-parameters (Effects of the width/ depth of neural networks, the step size of gradient descent, or the size of the training dataset, etc.);
(5) Computational complexity analysis (No. FLOP operations, Storage occupancy, or Training/Testing time etc.);
(6) Compare different methods and explain why one method is better than another one.
Total score does not exceed 30 pts.

- Individual Contribution (10 pts): Group members who made the major contribution get 10 pts (Implement one of the major parts of the experimental results, or Write at least 50% of the report); Members who made the assistance contribution get 5 pts.