Department of Computing

CS 471: Machine Learning (3+1)

Class: BSCS12-A

Lab 14: Open Lab

Date: 24-12-2024

Time: 2:00 PM – 4:50 PM

Instructor: Prof. Dr. Muhammad Moazam Fraz

Lab Engineer: Ms. Iram Tariq Bhatti

Course Learning Outcomes (CLOs)

No.	Course Learning Outcomes	Graduate Attributes (CS)	BT Level	Teaching & Learning Methods	Assessment Methods
CLO-1	Explain theoretical concepts behind various machine learning algorithms and techniques	GA-2	C-2	Active/Blended	Quizzes, Assignments, ESE, MSE
CLO-2	Analyse a given problem and select the most suitable supervise / unsupervised machine learning algorithm for the task.	GA-3	C-4	Active/Blended	Quizzes, Assignments, ESE, MSE
CLO-3	Implement various machine learning algorithms using modern software libraries for a range of applications.	GA-5	P-3	Cooperative Learning	Labs, Project
CLO-4	Contribute effectively in a team to complete a given task	GA-6	A-2	Cooperative Learning	Labs, Project

Lab 14: Open Lab

Learning Outcome

CLO 2: Analyse a given problem and select the most suitable supervise / unsupervised machine learning algorithm for the task.

CLO 3: Implement various machine learning algorithms using modern software libraries for a range of applications.

CLO 4: Contribute effectively in a team to complete a given task

Tools / Software Requirements

Python (>= 3.8), Google Colab, pandas, numpy, matplotlib, seaborn, scikit-learn

Contents

1) Lab Objective

In this lab, you will be analyzing a real-world graph dataset of a global flight network. The goal is to uncover patterns and trends that could improve air traffic management, predict delays, or improve route planning. You will be applying machine learning and graph analysis techniques, using your own understanding and creativity to derive meaningful conclusions.

2) Dataset Details

We are going to use the **OpenFlights Airline Route Dataset**, which represents the global network of flights. You can structure your data as follows:

Nodes: Airports

Edges: Direct routes between airports

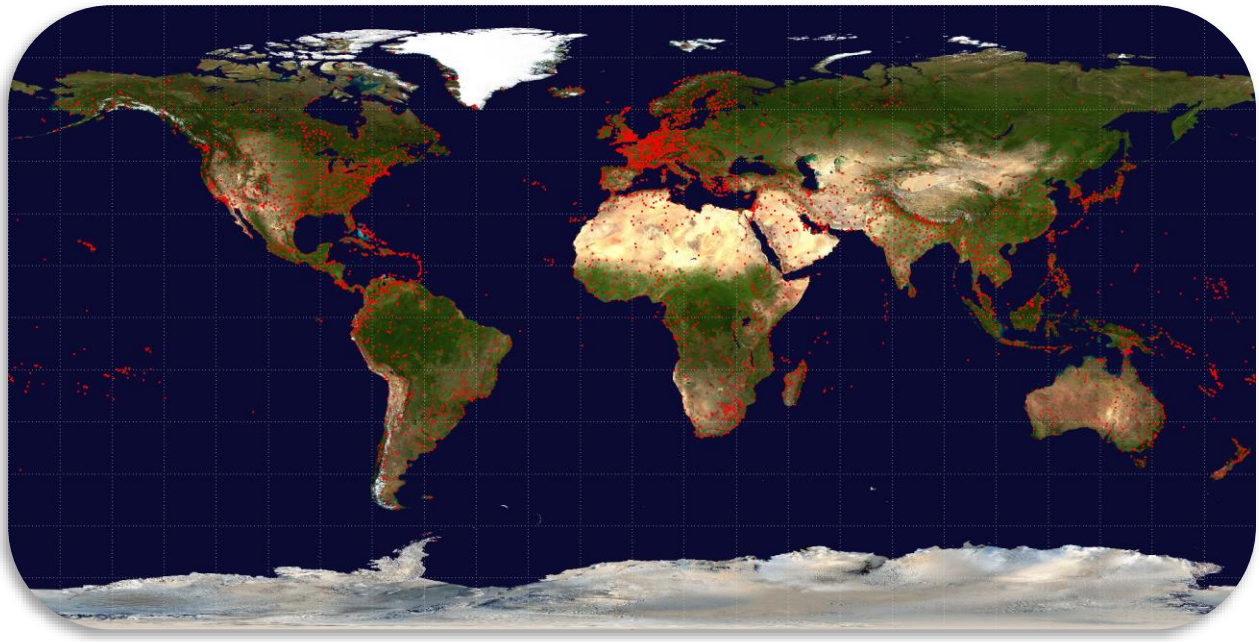
Node Attributes:

- Airport Name, City, Country, Latitude / Longitude

Edge Attributes:

- Airline, Distance, Source Airport, Destination Airport

Airport Database



Route Database

3) Defining Your Objective

Choose an objective for your analysis. Examples include:

- **Hub Identification -** Identifying the most critical hubs in the global air network using centrality measures.
- **Route Optimization -** Suggesting new routes to reduce overall travel time or distance.
- **Anomaly Detection -** Detecting underutilized airports or unusual traffic patterns.
- **Network Visualization –** Creating an interactive map of the network to visualize hubs, routes, or delays.

Note: Feel free to propose your own unique objective.

4) Implementing Your Analysis

Follow these steps using Python and graph analysis tools:

- Load and Explore the Dataset

Download and preprocess the dataset. Inspect key properties of the graph like number of edges, nodes, connected components, density, etc.

- Apply Graph Analysis

- Centrality Measures (Calculate degree centrality, betweenness centrality, or eigenvector centrality to identify important airports.)
- Community Detection Apply algorithms like Louvain or Girvan-Newman to find communities in the network.)
- Route Analysis (Evaluate shortest paths or create alternative routes between major hubs.)

- Model Predictions

Use machine learning to train a predictive model that supports your objectives based on route and airport characteristics. Feel free to use scikit-learn, PyTorch Geometric or any other library / resource.

- Visualization

Create visualizations using **Plotly** (for interactive world maps with nodes and edges), **Geopandas** (for geospatial analysis), **Graphistry** (to visualize graph clusters and key routes).

5) Deliverables

By the end of the lab, you are required to submit the following:

1. **Jupyter Notebook** – With code, visualizations, and detailed commentary on your findings.
2. **Summary Report** – Key extractions, techniques used, and implications of your analysis.

3. **One Slide (Optional)** – A summary for presentation purposes.

6) Reflection Questions

- How can the information and insights extracted from this dataset be used to optimize air travel?
- What challenges did you face when analyzing graph-based data, and how did you overcome them?
- How does the global structure of the flight network reveal economic or geographic trends? Did you notice any geographical, political, or economic factors influencing the graph structure?
- What patterns or trends did you observe in the dataset that were surprising or counterintuitive?
- Which graph metrics or algorithms were the most useful for your analysis, and why?
- How could airlines or airport authorities use your analysis to make operational decisions?
- Which visualizations were the most effective for communicating your findings?
- What additional data or features could make your visualizations more informative?