Machine learning task for TDT4173 (Modern Machine Learning in Practice)

Objective

Given AIS data from 1st January to 7th May 2024, predict the future positions of vessels at given timestamps for five days into the future. Students will develop predictive models that account for various factors such as congestion, port calls, and other events affecting the vessels' journeys.

Task Description

- Given an AIS dataset containing the positions of 689 vessels per time from January to May 2024, predict the next positions of a select 216 of these vessels at their given timestamps.

- One must utilize the AIS dataset; the vessels, schedules and ports datasets are optional.

Dataset Overview

- The required datasets are the following:

  ○ ais_train.csv: Contains the AIS data for training. This contains the positions of 689 vessels. The dataset was sampled every 20 minutes, but the timestamps for each vessel are irregular.

  ○ ais_test.csv: Contains AIS data without longitudes and latitudes for testing for 216 vessels that will be used to evaluate the students. 50% of this dataset will be used for the public leaderboard score, and the rest is used for the private leaderboard.

- The optional datasets are:

  ○ schedules_to_may_2024.csv: Contains the planned arrival destinations and time as communicated from the shipping lines for a select 252 vessels.

  ○ vessels: Contains some information about each vessel.

  ○ ports.csv: Contains information about the ports referenced in scheduels_to_may_2024.csv.

Evaluation

- The evaluation metric is the weighted average of the mean geodetic distance in killometer (using geopy.distance.geodesic) between the prediction and the ground truth point for each vessel. The average is weighted per day into the future. The weights are as follows:

  - ○ 0 - 1 day: 0.3
  - ○ 1 - 2 days: 0.25
  - ○ 2 - 3 days: 0.2
  - ○ 3 - 4 days: 0.15
  - ○ 4 - 5 days: 0.1