

# Project I. traffic prediction with graph neural network

The goal of this project is to apply all material learned in INF367 and successfully apply geometric deep learning methods. Please read the description carefully!

This project is a compulsory part of the course. This project contributes 25% to the final grade. The grade will be based on good choice of methods, correctness of answers, clarity of code and thoroughness and clarity of reporting.

The data for this project comes from Statens vegvesen. There are two data files, one about the different traffic stations and one about the recorded traffic. The data about the traffic stations includes the columns id, name, county, municipality, latitude and longitude. The traffic data contains the columns station\_id, station\_name, time.from, time.to and volume. Our goal is to predict the volume one hour before it happens.

## Requirements

You will build a machine learning model to predict how many vehicles will drive at different places an hour ahead of time. The system takes as its input the current date and time and currently recorded traffic, and will output the predicted number of vehicles in one hour. The work will consist of four parts:

- Data preparation and exploratory data analysis (10 pts): Describe the data.
- Baseline model (10 pts): A machine learning model that does not take into account the spatial positions of the different places, but takes into account the position, month, day of week, time of day and the traffic in the hour before.
- Graph neural network models (20 pts): Test different graph neural network models that use node features, edge features and global features to predict the traffic in an hour.
- Final model (5 pts): A description of the final model with estimated generalization RMSE.
- Report (5 pts): A clear report that justifies all choices, explains the result with well-described figures.

## Deadline

The project has to be submitted before Sunday, 13.03, 23.59. Deliver at MittUIB.no/assignments. Late assignments will not be considered.

## Deliverables

For the final submission, please provide one zip file that contains all relevant parts of the project including a README.md file that gives an overview of all files included. Alternatively, you can provide a link to a github/gitlab repository that includes all relevant files.

In addition to packages from the standard library, you may use the following python packages: `numpy`, `pandas`, `scipy`, `sklearn`, `matplotlib`, `seaborn`, `plotly`, `geopy`, `pytorch_geometric`, `torch`. If you use any other packages we will not be able to run your app and you will fail the project (unless discussed with the instructors ahead of time).

Code should be documented and tricks (e.g. to avoid division by zero, to make sure it takes finite time to run, etc.) should be reported. The rationale behind all steps in the code should be clear from the report.

Model selection and evaluation are important parts of the project and will be assessed accordingly.

NOTE: This project is a learning experience. If we see that you have copied your answers from online resources, you will get 0 points.