Data challenge: Flights prediction

Mathurin Massias, Inria

Key dates

- kickoff today
- end of challenge registration Monday 13/12
- EMINES students: 5 days of class next week
- submission period ends mid January

Submissions

- you have access to a observation/target training dataset
- a public test set is released without ground truth
- one submission allowed per day
- EMINES students: you must submit the code for the best submission, it will be trained and tested on hidden data
- be careful about not overfitting the test set...

Inria

Context

A single unnamed airline company

- 100-200 airplanes with crew
- numerous connected cities in the US
- some fixed cost by flight, indep. of passenger number

Questions

- how to dynamically allocate planes among possible routes?
- how many passengers travel at a given date?
- how many people book and don't show up?

Inria

Available data

- date
- number of passengers (through transformation)
- departure and arrival airports
- how early the tickets are booked

This is probably not enough and can be complemented by external data

Observations

- data is inherently spatio-temporal
- train test split is based on a cutoff date
- time series approach?
- graph approach?
- limited number of submissions: be careful

Process

- explore different ML models
- data processing, external datasets
- tune hyperparameters

Inria

Logistical details

- teams of 1-3 (EMINES: 2)
- discord channel to form teams
- and ask questions publicly

There is no prize, so focus on learning, testing your ideas... be nice to others and have fun

Inria