Building Machine Learning Systems with TensorFlow.js #2

Asst.Prof.Dr.Supachai Vorapojpisut

Thammasat University

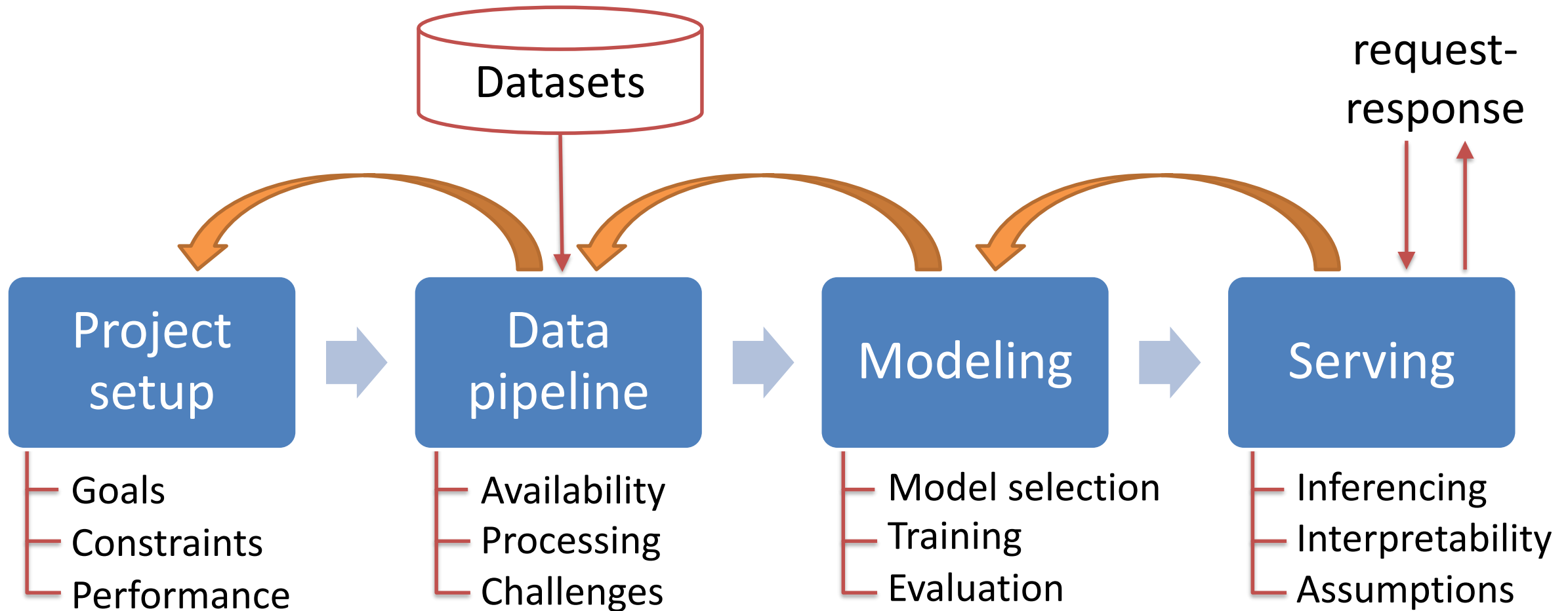https://qrgo.page.link/tMfJr

Outline

Jan 7, 2020

- Working with data
- Machine learning concepts
- Time-series prediction
- ML coding
  - jest
  - TensorFlow.js

Jan 8, 2020

- Machine learning system design
- Model management
- SpatioTemporal data analytics

Machine learning system design

Datasets

request-response

Project setup	Data pipeline	Modeling	Serving
— Goals	— Availability	— Model selection	— Inferencing
— Constraints	— Processing	— Training	— Interpretability
— Performance	— Challenges	— Evaluation	— Assumptions

Using trained models

The neural network features are listed as follows:
1. Total server IT load [kW]
2. Total Campus Core Network Room (CCNR) IT load [kW]
3. Total number of process water pumps (PWP) running
4. Mean PWP variable frequency drive (VFD) speed [%]
5. Total number of condenser water pumps (CWP) running
6. Mean CWP variable frequency drive (VFD) speed [%]
7. Total number of cooling towers running
8. Mean cooling tower leaving water temperature (LWT) setpoint [F]
9. Total number of chillers running
10. Total number of drycoolers running
11. Total number of chilled water injection pumps running
12. Mean chilled water injection pump setpoint temperature [F]
13. Mean heat exchanger approach temperature [F]
14. Outside air wet bulb (WB) temperature [F]
15. Outside air dry bulb (DB) temperature [F]
16. Outside air enthalpy [kJ/kg]
17. Outside air relative humidity (RH) [%]
18. Outdoor wind speed [mph]
19. Outdoor wind direction [deg]
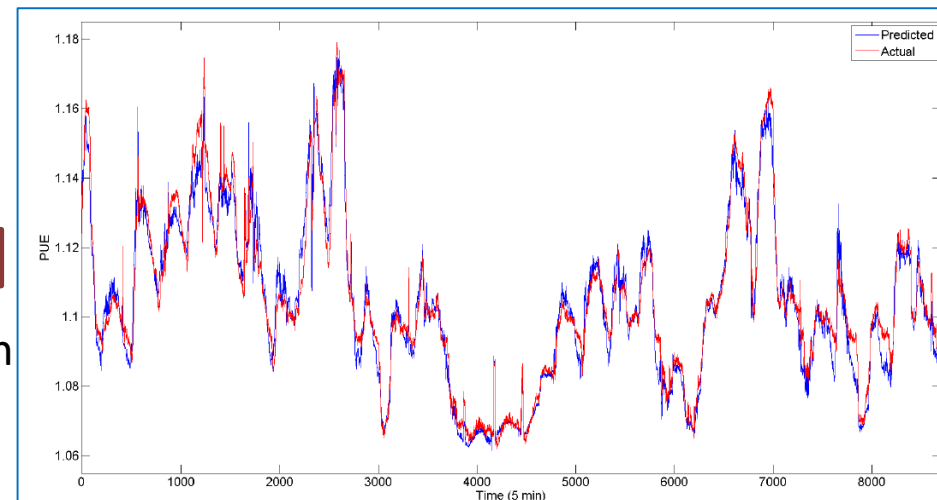
Neural Network
Training Workflow

Regression | Classification | Clustering | Prediction

What-if Analysis

Power Usage Effectiveness

What-if analysis

Optimization

DeepMind AI Reduces Google Data Centre Cooling Bill by 40%

From smartphone assistants to image recognition and translation, machine learning already helps us in our everyday lives. But it can also help us to tackle some of the world's most challenging physical problems -- such as energy consumption. Large-scale commercial and industrial systems like data centres consume a lot of energy, and while much has been done to stem the growth of energy use, there remains a lot more to do given the world's increasing need for computing power.

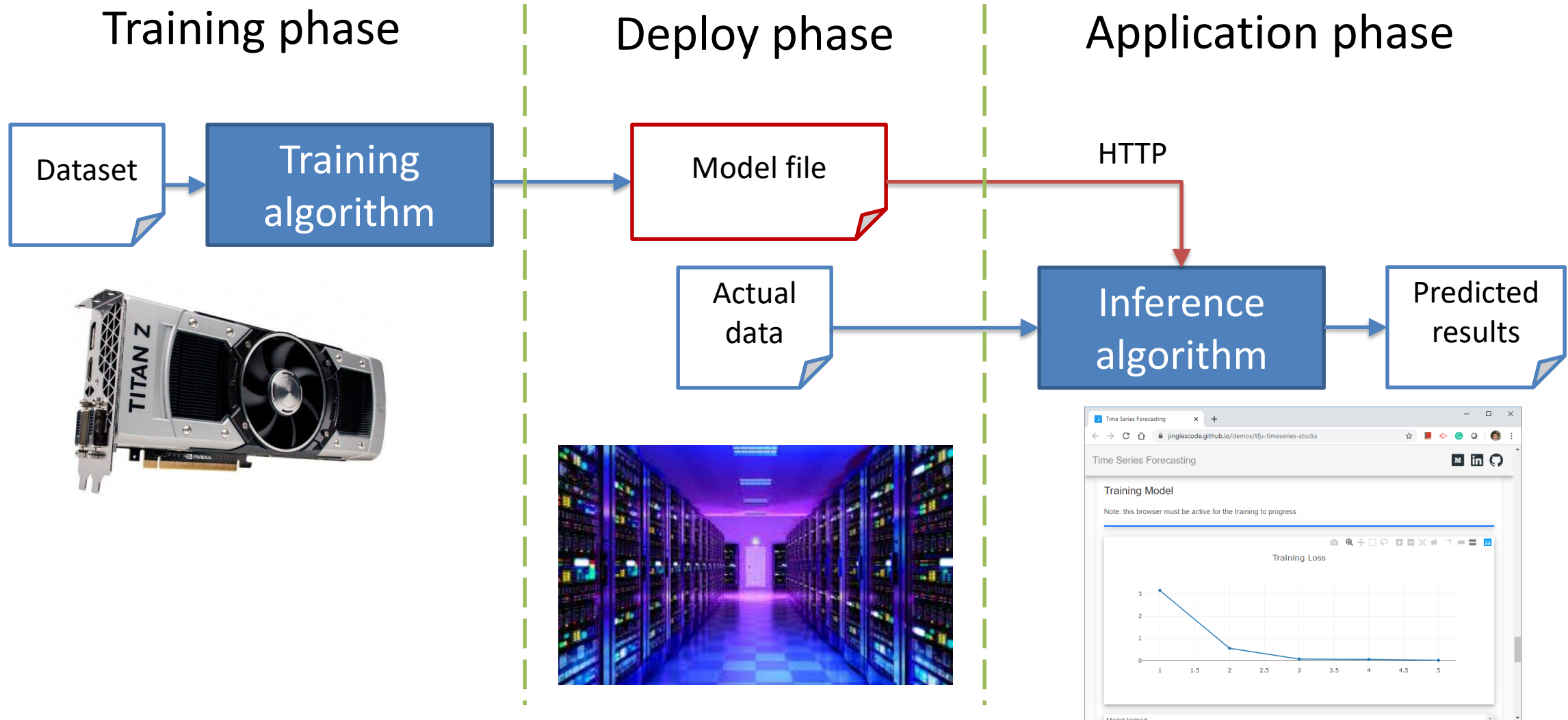# Model management

Training phase | Deploy phase | Application phase

Dataset → Training algorithm → Model file → HTTP → Inference algorithm → Predicted results

Actual data → Inference algorithm

TensorFlow.js: server to browser

```javascript
const tf = require('@tensorflow/tfjs-node');

datasetObj = await createDataset();
model = createModel();
dataset = datasetObj.dataset.shuffle();
trainDataset = dataset.take();
validationDataset = dataset.skip();

await model.fitDataset();
result = model.predict();
await model.save();
```

```html
<html>
  <head>
    <script src="js/@tensorflow/tfjs@latest">
    </script>
  </head>
  <body>
    <h4>Tiny TFJS example<hr/></h4>
    <div id="micro-out-div">Training...</div>
    <script src="./index.js"> </script>
  </body>
</html>
```
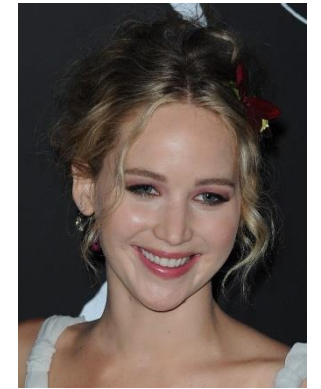
SpatioTemporal properties

Diane Keaton

Michelle Pfeiffer

Julia Roberts

Liv Tyler

Jennifer Lawrence

70s   80s   90s   2000   year

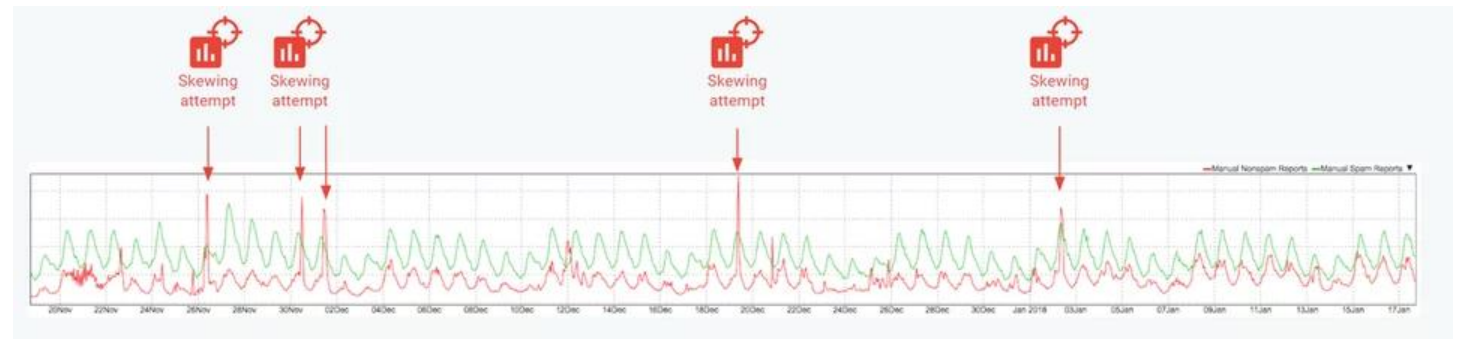# AI attacks

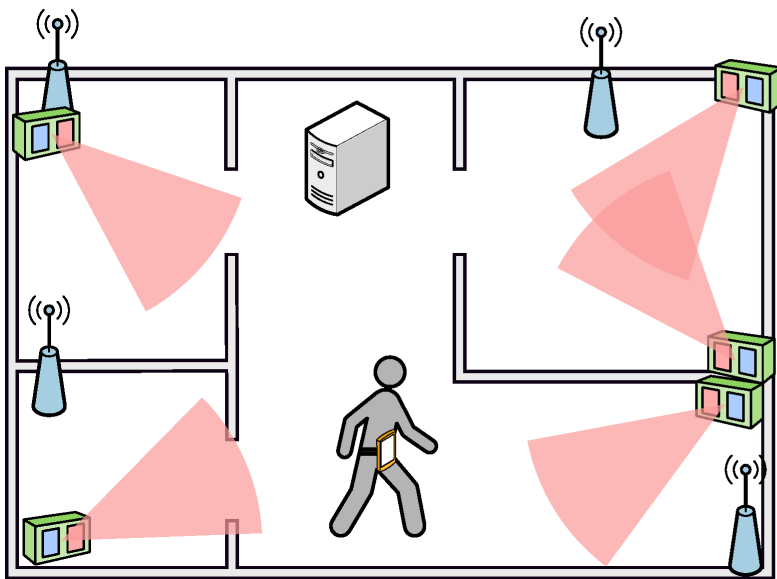Adversarial attacks: face detection

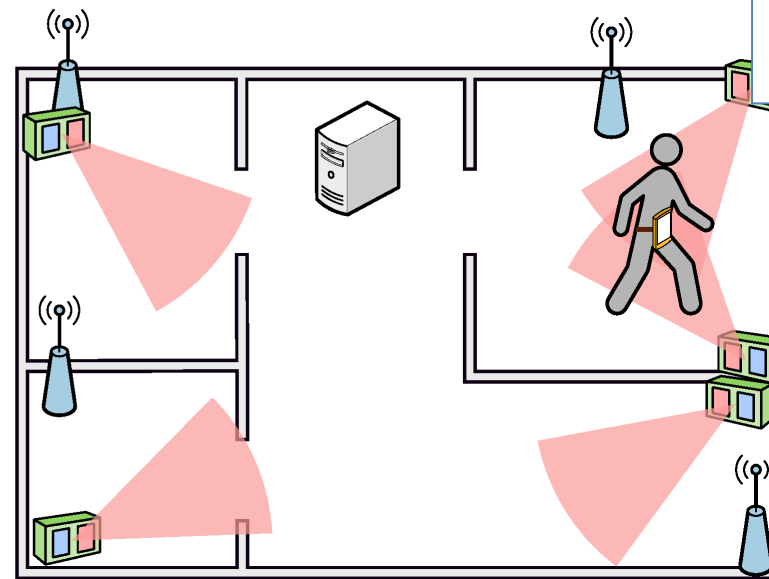(a)   (b)   (c)   (d)

Data poisoning attacks: email spam filter

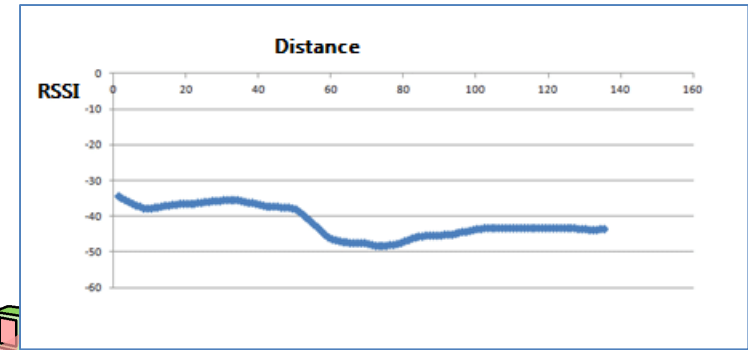https://towardsdatascience.com/breaking-neural-networks-with-adversarial-attacks-f4290a9a45aa

https://elie.net/blog/ai/attacks-against-machine-learning-an-overview

Using BLE RSSI for proximity detector

(a)   (b)
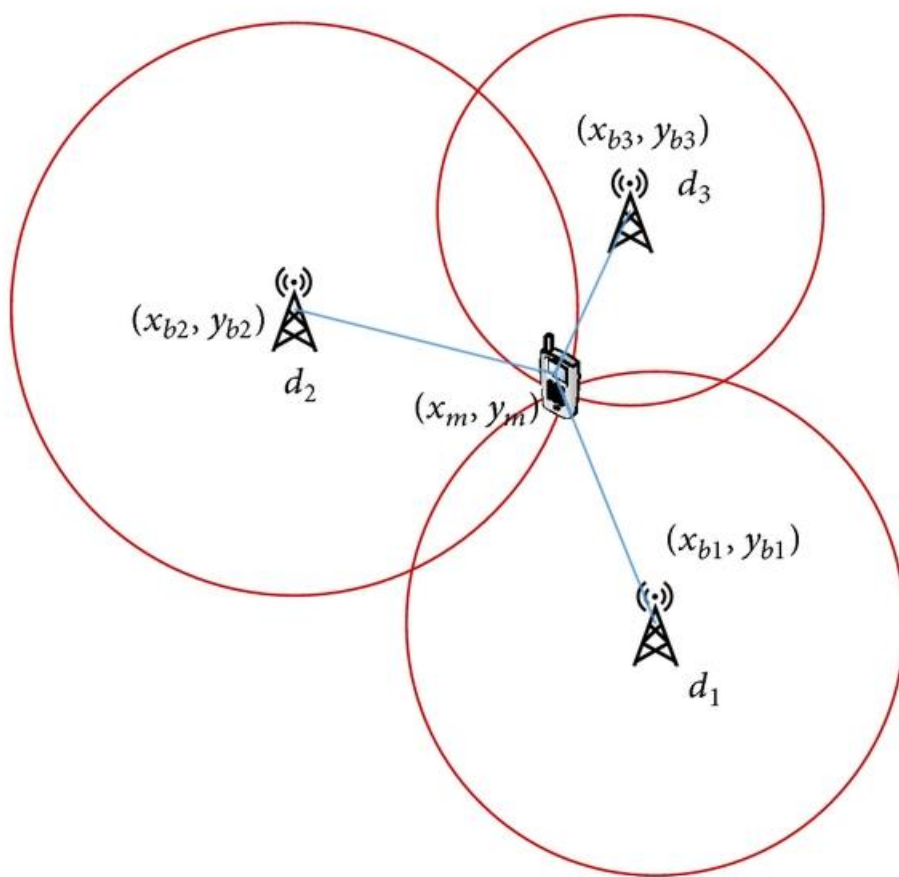
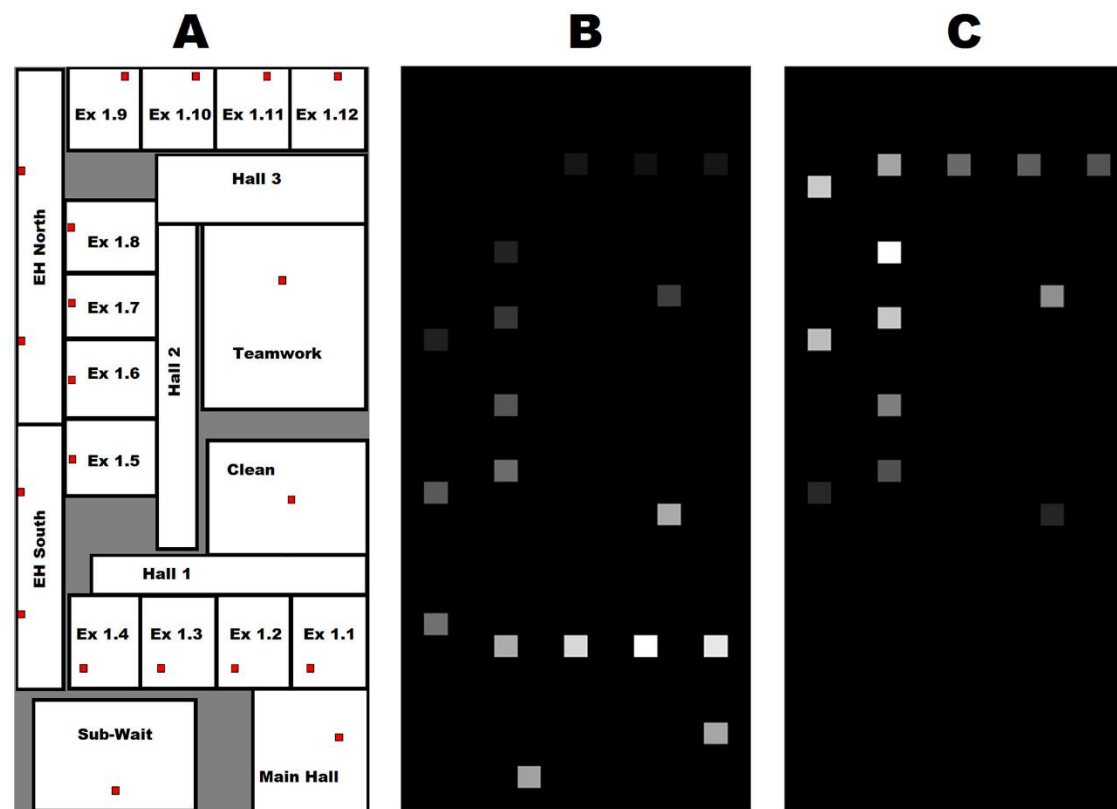BLE anchor   ranging sensor   tag   system controller

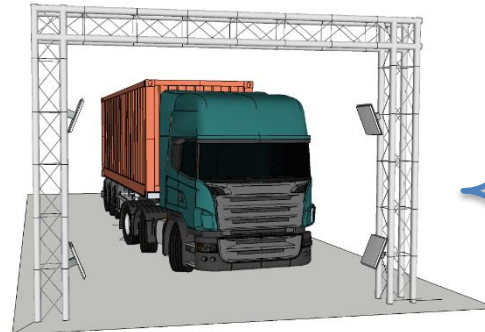# Classic vs NN approaches

RSSI triangulation method

Deep NN method
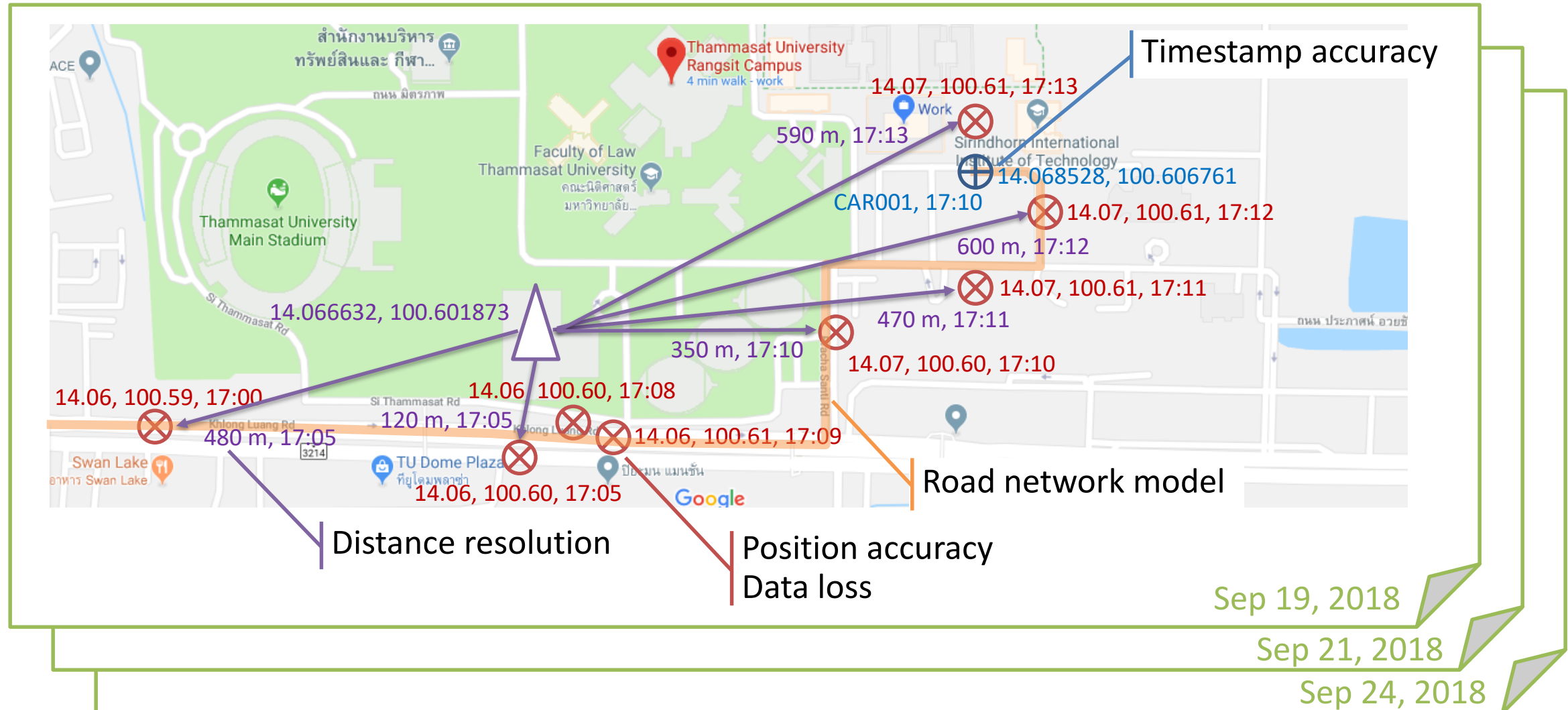
Spatiotemporal data

Accuracy, frequency, coverage

Data model

ID-based positioning
(lat, lng, time), …
(x, y, time), …

Location-based positioning
At X,Y with timestamp 12:00
(X, Y, t, p), …

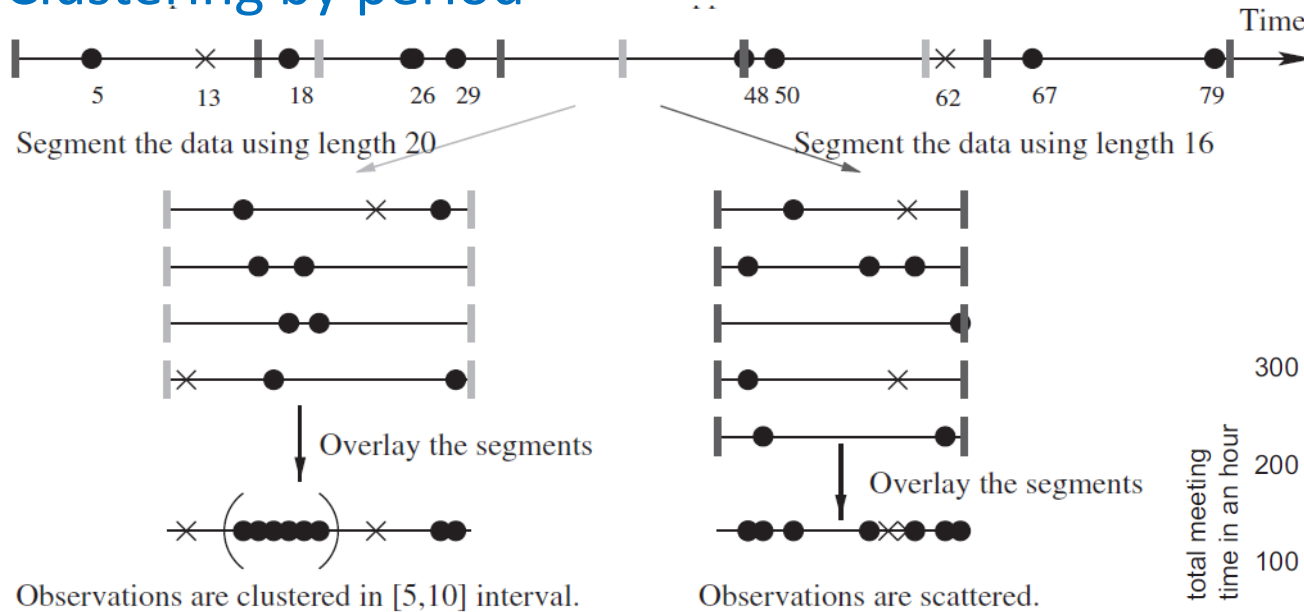"Spatiotemporal Pattern Mining: Algorithms and Applications", Frequent Pattern Mining

Spatiotemporal data

Timestamp accuracy

14.07, 100.61, 17:13

590 m, 17:13

14.068528, 100.606761

CAR001, 17:10

14.07, 100.61, 17:12

600 m, 17:12

14.07, 100.61, 17:11

470 m, 17:11

14.066632, 100.601873

350 m, 17:10

14.07, 100.60, 17:10

14.06, 100.59, 17:00

14.06, 100.60, 17:08

480 m, 17:05

120 m, 17:05

14.06, 100.61, 17:09

Road network model

14.06, 100.60, 17:05

Position accuracy
Data loss

Distance resolution

Sep 19, 2018

Sep 21, 2018

Sep 24, 2018

Spatiotemporal data mining

Clustering by period

Clustering by semantic

Clustering by zones

Segment the data using length 20

Segment the data using length 16

Overlay the segments

Observations are clustered in [5,10] interval.

Overlay the segments

Observations are scattered.

a non−friend pair
a friend pair

total meeting time in an hour

days of the week

day 1
day 2
day 3

a   am object's movement

b   a set of predefined regious
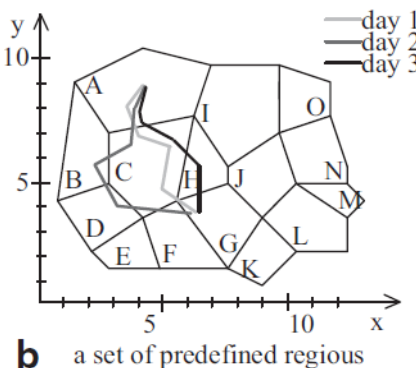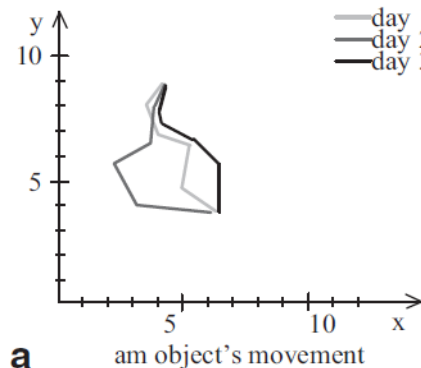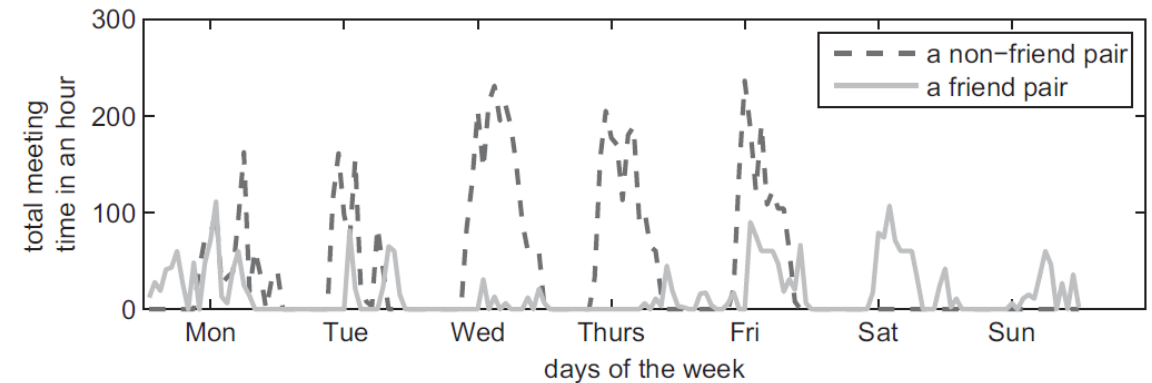
events sequence:
A A C C C G | A A C B D G | A A A C H G

some partial periodic patterns:
support(AA***G) = 3
support(AAC**G) = 2
support(AA*C*G) = 2

c   event-based patterns