An Introduction to Supervised Machine Learning

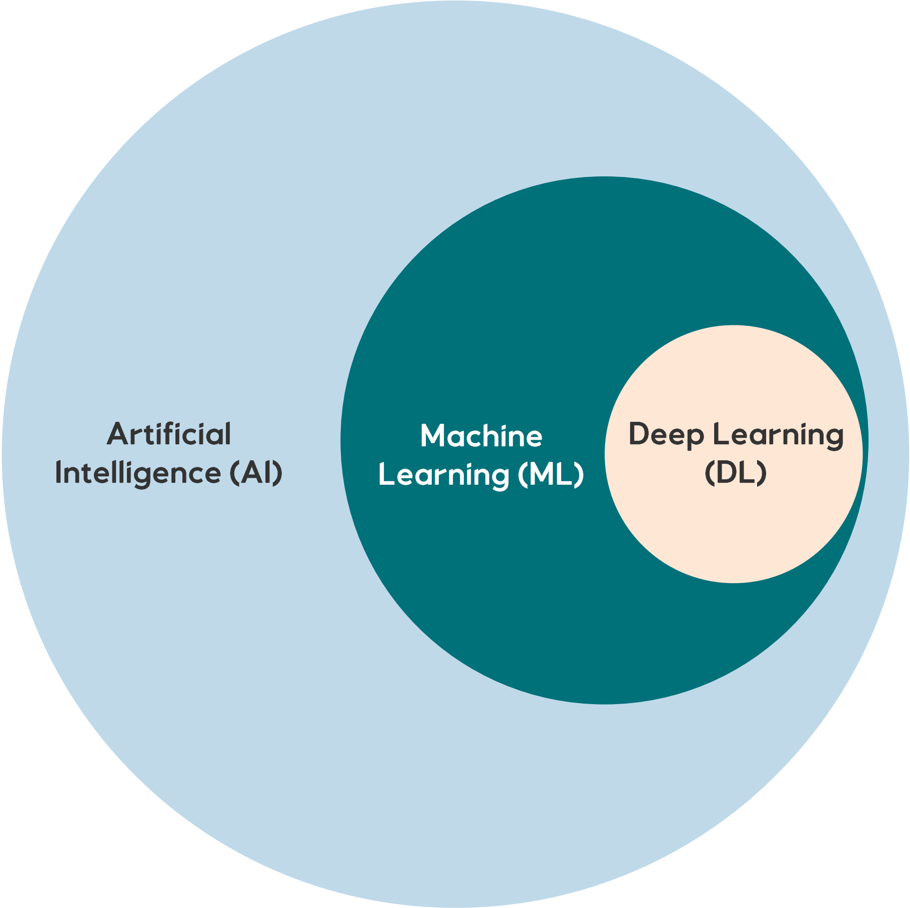
FORCE Data Analytics, Machine Learning and data centric workflows Network Group

24th May 2023

A Brief Explanation about Machine Learning

Artificial Intelligence - *Machines mimicking cognitive abilities / human functions.*

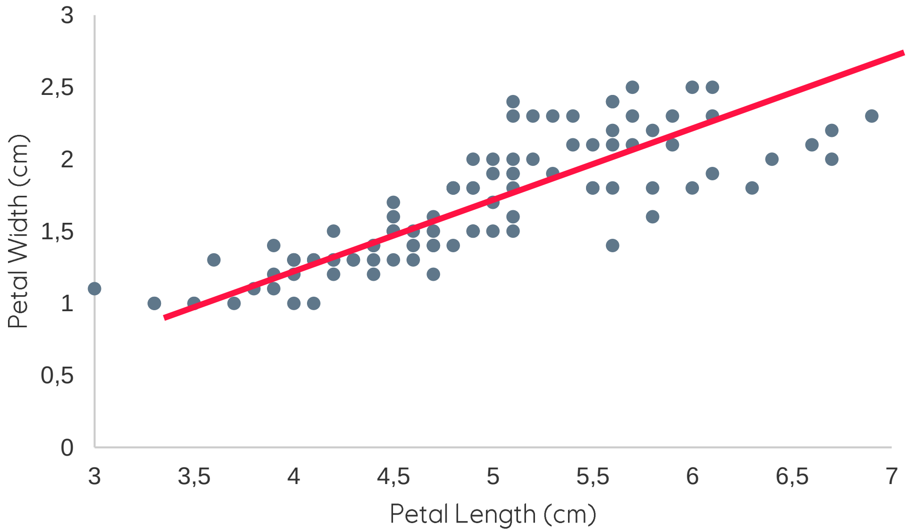
Machine Learning – *A subset of AI that uses statistical methods to enable machines to improve with experience without being explicitly programmed.*



What makes ML 'supervised'?

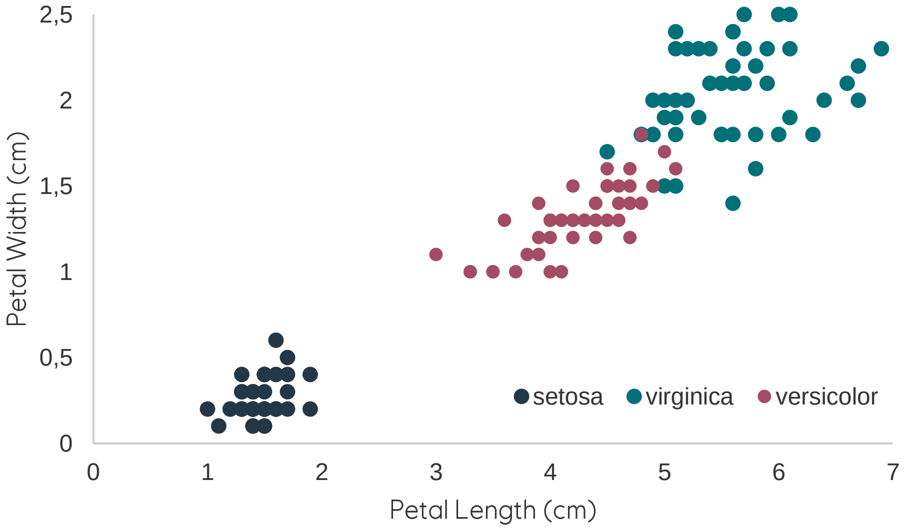
Supervised Machine Learning – *A subset of ML that uses labelled datasets to train algorithms that classify data or predict outcomes.*

Regression



- Mapping inputs to **continuous** output
- E.g., Predicting the width of a petal given its length

Classification



- Mapping inputs to output **category**
- E.g., Classifying plant species given its length and width

What Data is Needed for Supervised ML?

Input data needs to have a labelled output(s)

- Applies to both regression and classification tasks
- If there are no labels, then the problem becomes **un**supervised

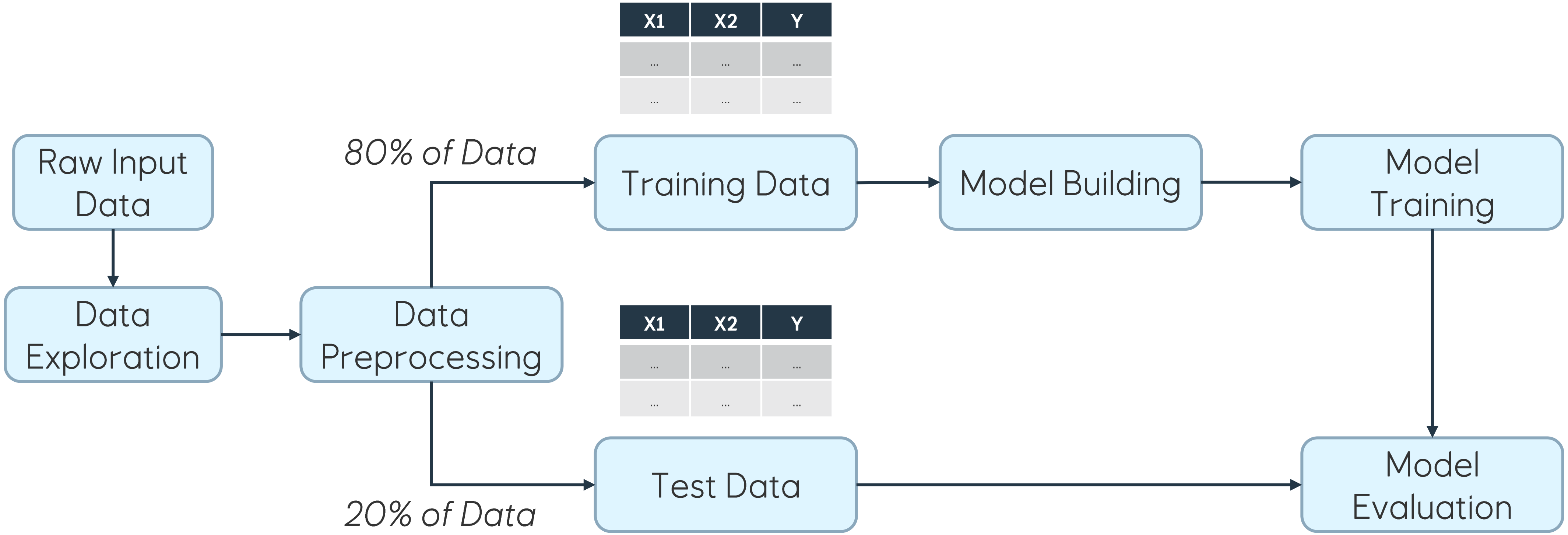
Inputs / Features				Output
Sepal Length (cm)	**Sepal Width (cm)**	**Petal Length (cm)**	**Petal Width (cm)**	**Species**
5.1	3.5	1.4	0.2	setosa
4.9	3	1.4	0.2	setosa
6.5	2.8	4.6	1.5	versicolor
5.7	2.8	4.5	1.3	versicolor
6.5	3.2	5.1	2	virginica
6.4	2.7	5.3	1.9	virginica

Versicolor

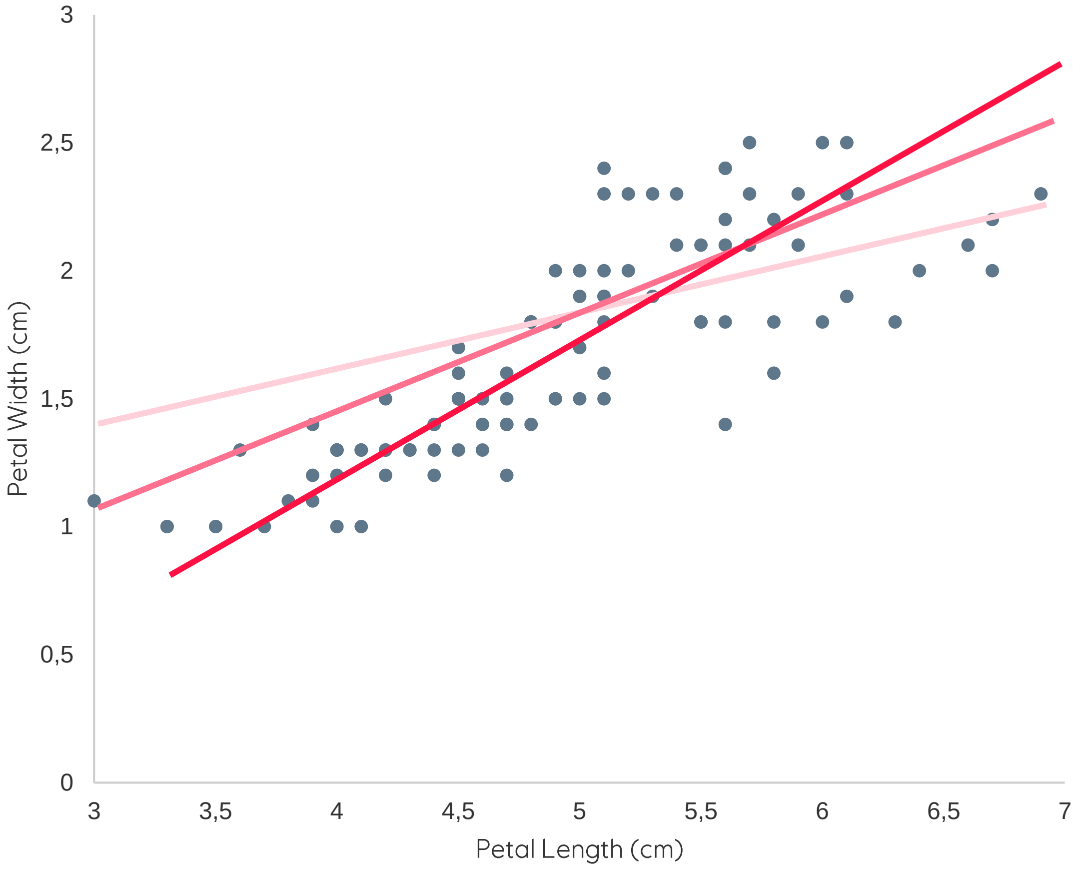
Setosa

Virginica

A Supervised Machine Learning Workflow



How Does A Machine Learning Model 'Learn'?



Goal: Find a best fit line through the data points

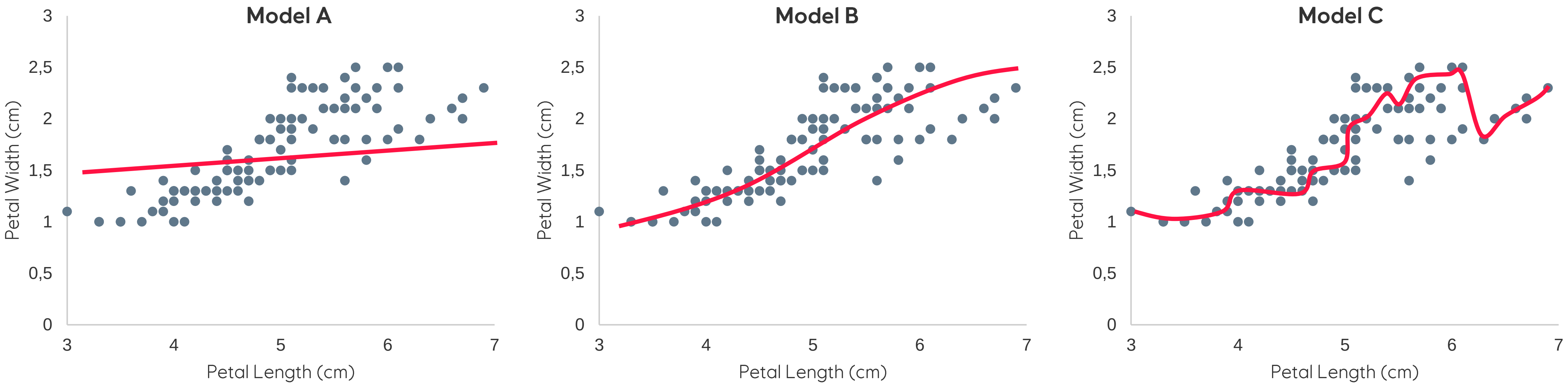
- Line is represented by the **slope** and **y intercept**

Parameters to optimize!

- Draw line and calculate the error between the line and data points
- Progressively update line parameters in the direction that minimizes the loss

Evaluating Model Performance

After model training, its important to evaluate performance:



Underfitting

- Poor performance on training data
- Unable to understand input-output relationships

Balanced

- Good performance on training and test data
- Generalize on unseen examples

Overfitting

- Poor performance on testing data
- Unable to generalize on unseen examples

Factors affecting performance: *Data Quantity, Representative Features, Algorithm Choice, Training Time, …*

The Value & Challenges of Supervised Machine Learning

Utilizes prior experience

Saves defining complex rules

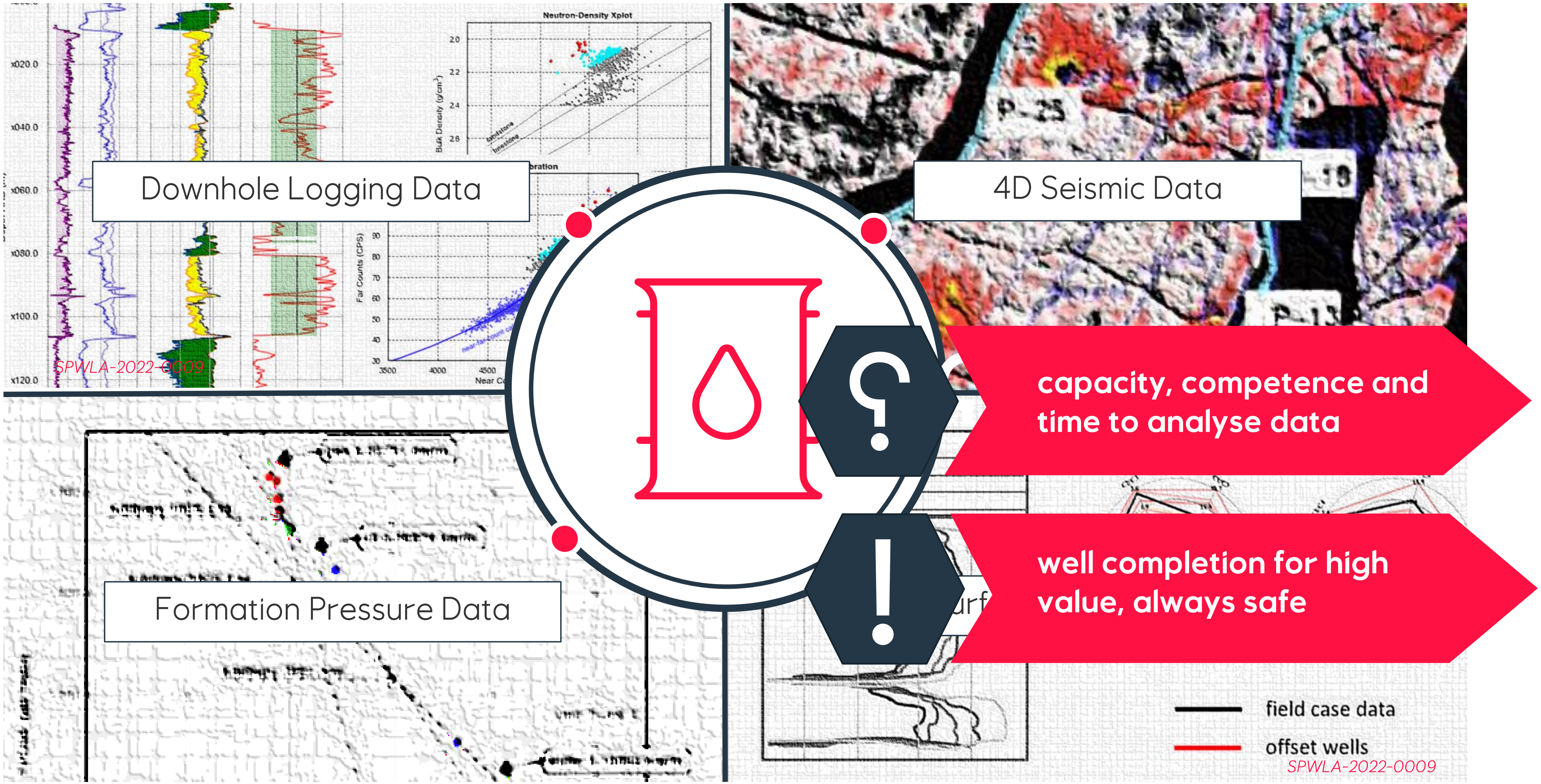
Better understand relationships between inputs-outputs

Data Preparation & Algorithm Selection

Lack of representative training data

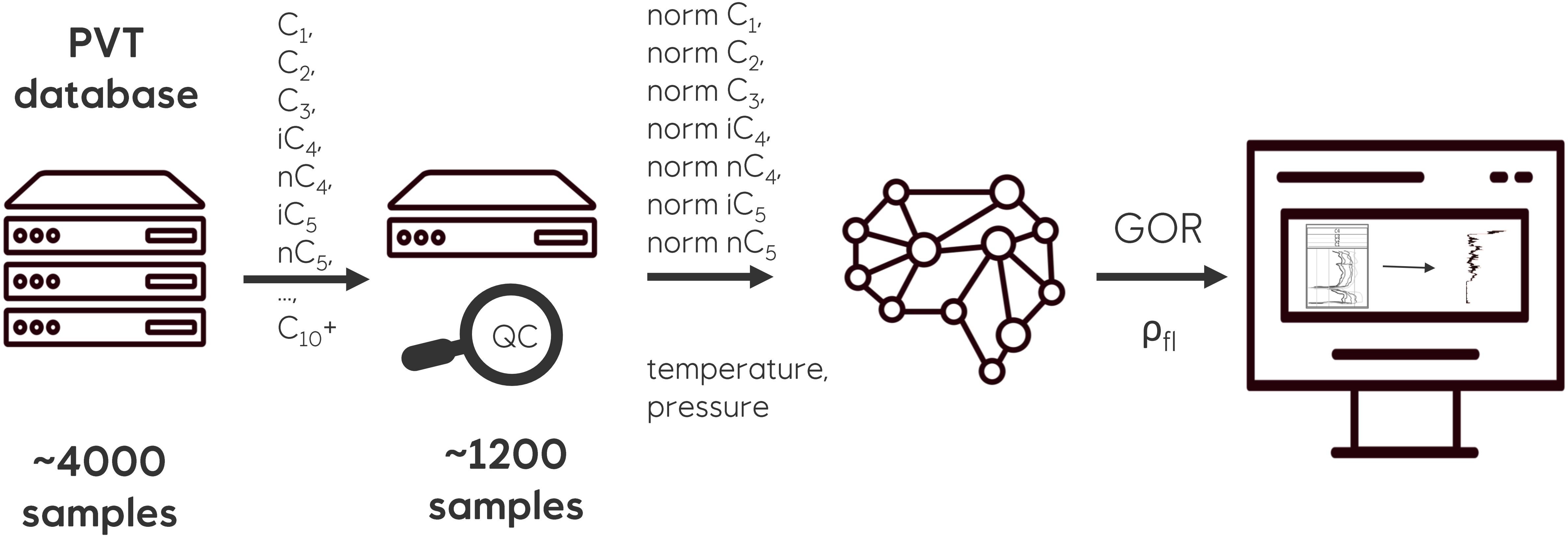
Over/Under-fitting

equinor



Downhole Logging Data
4D Seismic Data
Formation Pressure Data
capacity, competence and time to analyse data
well completion for high value, always safe
Neutron-Density Xplot
field case data
offset wells
SPWLA-2022-0009
SPWLA-2022-0009

Reservoir Fluid Property Prediction – Formation Gas Analysis



SPE-201635, SPE-205842-MS,
DOI: 10.30632/SPWLA-2022-0009

Real-time case: Fluid Identification

- Data acquisition strategy
- Field Case 1: Exploration example to optimize fluid sampling
- Field Case 2: Development well from depleted reservoir to optimize real-time decisions

SPWLA 63rd Annual Logging Symposium, June 10-15, 2022
DOI: 10.30632/SPWLA-2022-0009

REAL-TIME FLUID IDENTIFICATION FROM INTEGRATING ADVANCED MUD GAS AND PETROPHYSICAL LOGS

Margarete Kopal, Gulnar Yerkinkyzy, Marianne Therese Nygård,
Alexandra Cely, Frode Ungar, Sandrine Donnadieu and Tao Yang, Equinor ASA, Norway

Copyright 2022, held jointly by the Society of Petrophysicists and Well Log Analysts (SPWLA) and the submitting authors.
This paper was prepared for presentation at the SPWLA 63rd Annual Logging Symposium held in Stavanger, Norway, June 10-15, 2022.

ABSTRACT

Advanced mud gas logging has been used in the oil industry for about 25 years. However, it has been challenging to predict reservoir fluid properties quantitatively (e.g., gas oil ratio – GOR) from only the advanced mud gas data (AMG) while drilling. Yang et al. proposed the first accurate GOR r... tive ...od... n 2019 ... advanced surface ... lea... orith... S...

question of whether it encounters any injection gas. We applied the new approach to several production wells and obtained satisfying result. The latest information from the predictive GOR model solved many puzzles in petrophysical interpretations.

This paper presents a new approach for reservoir fluid identification by integrating advanced mud gas data and petrophysical logs while drilling. This new approach makes real-time operational adjustments possible based on reservoir fluid identification along the well. The business potential is significant for accurately mapping ...esources for in-f... w... boosting pr... tabili... and ...vering carb...

Added Value to Exploration and Production Wells

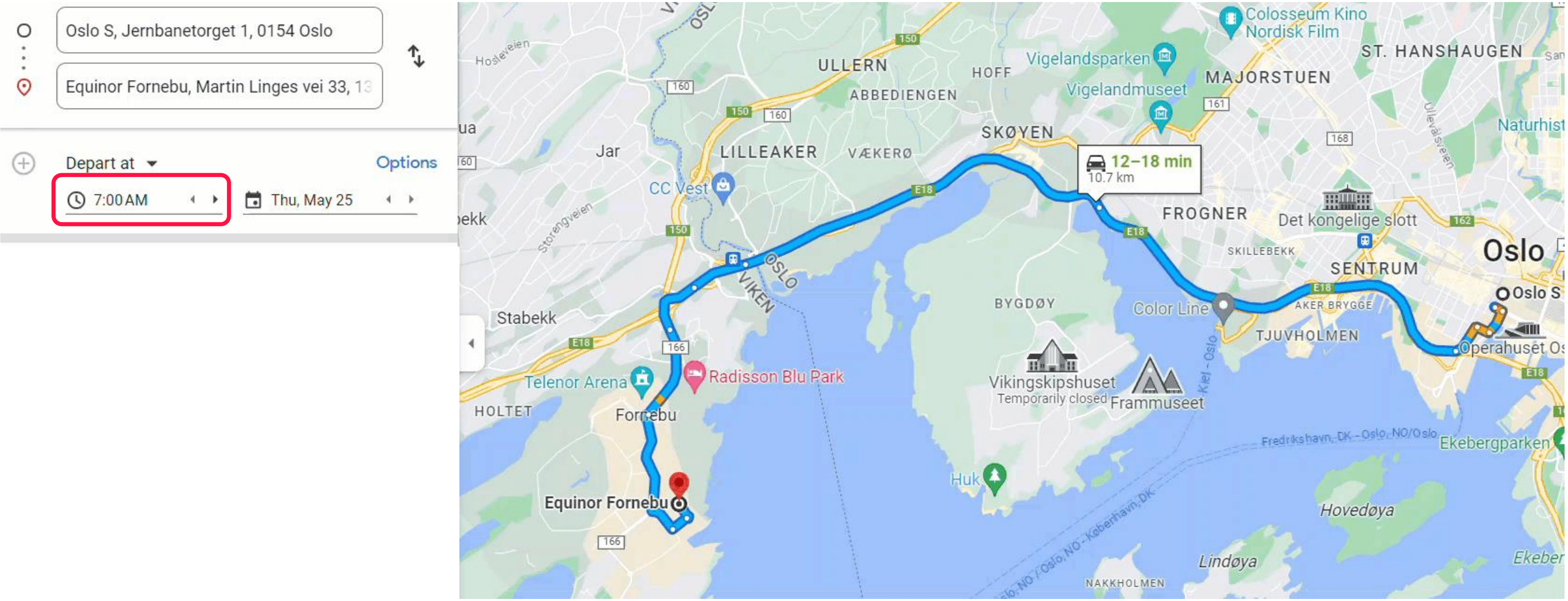
- Run Quality-Control of Surface Logs
- Compare Composition with Database
- Predict Reservoir Fluid Properties
- Adjust Sampling Program in Real-Time
- Decide on Completion for Optimized Oil Production

Supervised Machine Learning: A daily life example

equinor

equinor

Predicting Driving Travel Time

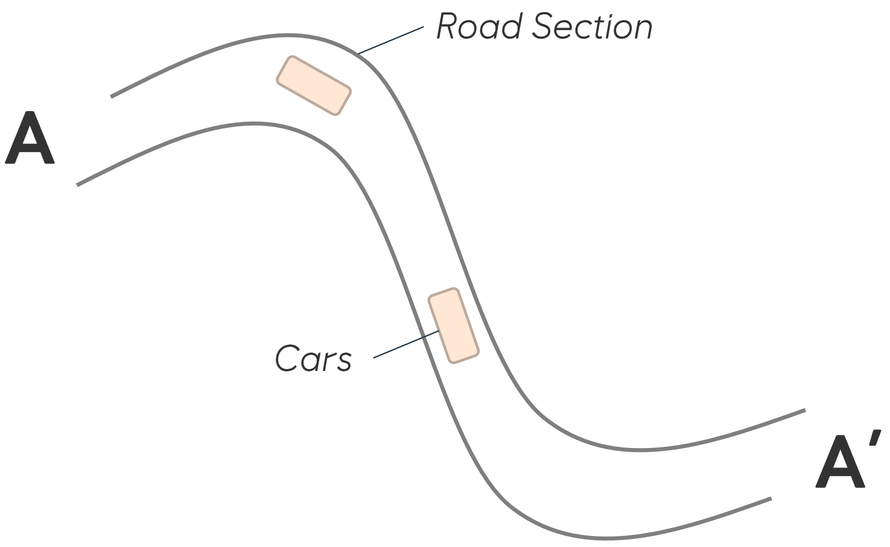


Source: Google Maps

Framing The Problem

Goal:
Improve the accuracy of Expected Time of Arrival (ETA) for road routes

Simplified Concept



What Data Do We Have?

Inputs

- Time (i.e., hour)
- Location / Road Section
- Weather Conditions

Output

- GPS Location (i.e., to calculate speed/travel time)

Any other relevant inputs?

Early Intuition About The Problem

Some hypotheses we might have about travel time:

↑ *Weekday mornings/evening*

↑ *Locations close to urban centers*

↑ *Heavy rain/snow*

↓ *Weekends*

↓ *Summer vacation period*

Why not use these hypotheses as rules?

- We want to detect non-obvious patterns in the data
- Our dataset may contain complex relationships (i.e., interaction between weather, location and time)
- We want our model to learn, and update based using new data without manual intervention

Model Workflow | Data

An example of what raw data could look like:

	Inputs		Output
Time	**Weather**	**Road Type**	**A-A′ Travel Time (minutes)**
06:00	Sun	Highway	5.5
09:00	Rain	Highway	12
12:00	Sun	Highway	7
…	…	…	…

- Numerical data (e.g., travel time)
- Categorical data (e.g., weather)
- Cyclical features (e.g., time)

Thoughts on data preprocessing?

Model Workflow | Splitting Data

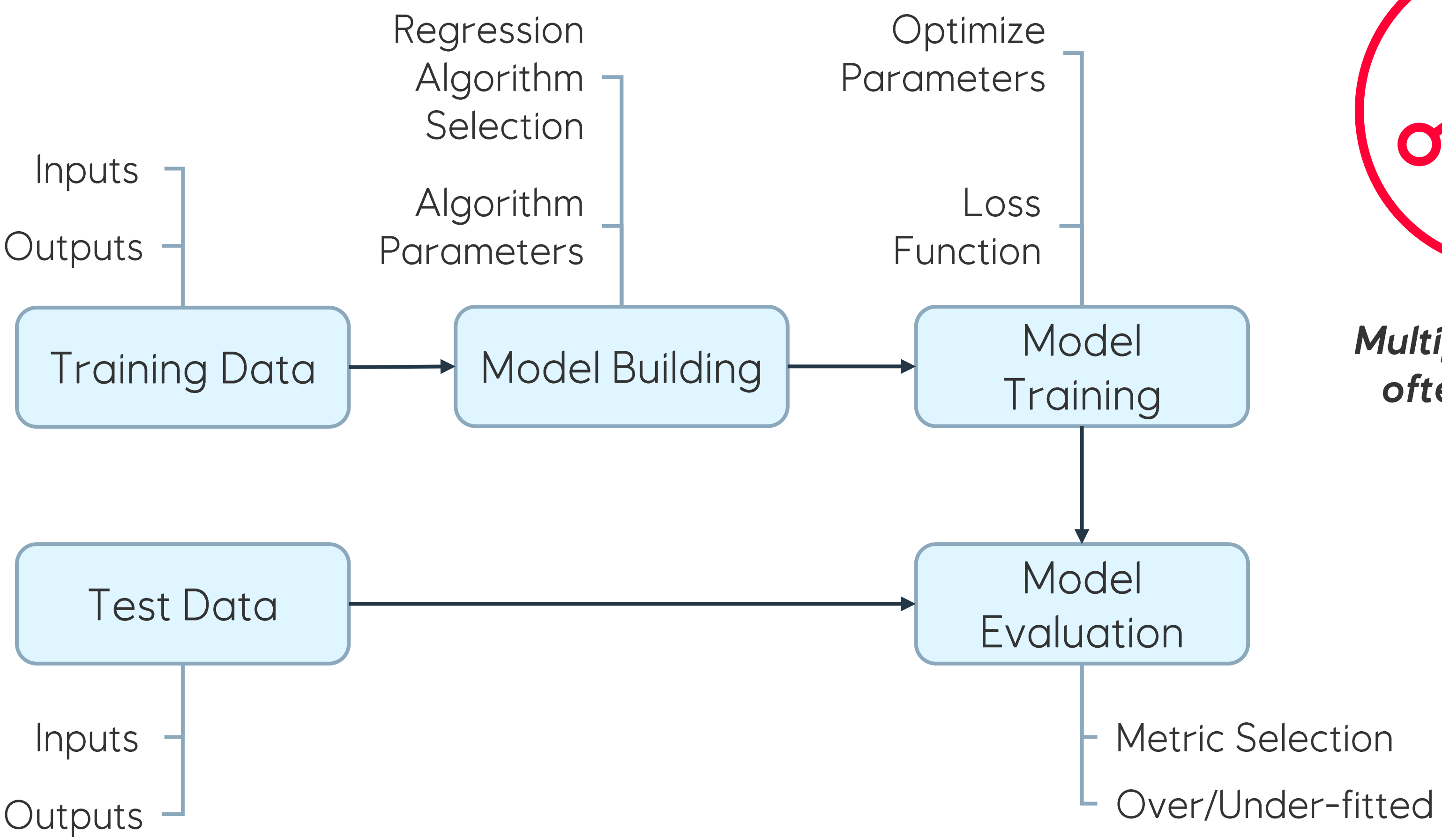
Training Dataset (80% of total):

Hour_sin	Hour_cos	Sun	Rain	Highway	Single Lane	A-A′ Travel Time (minutes)
0	6.12e-17	1	0	1	0	5.5
7.07e-01	-7.07e-01	0	1	0	1	12
...	...	...	...	...	...	...

Test Dataset (20% of total):

Hour_sin	Hour_cos	Sun	Rain	Highway	Single Lane	A-A′ Travel Time (minutes)
1.22e-16	0	1	0	1	0	7
...	...	...	...	...	...	...

Model Workflow | Building, Training & Evaluation



Multiple Iterations often required!

An introduction to Supervised Machine Learning

Elliot Humphrey & Margarete Kopal

© Equinor ASA

This presentation, including the contents and arrangement of the contents of each individual page or the collection of the pages, is owned by Equinor. Copyright to all material including, but not limited to, written material, photographs, drawings, images, tables and data remains the property of Equinor. All rights reserved. Any other use, reproduction, translation, adaption, arrangement, alteration, distribution or storage of this presentation, in whole or in part, without the prior written permission of Equinor is prohibited. The information contained in this presentation may not be accurate, up to date or applicable to the circumstances of any particular case, despite our efforts. Equinor cannot accept any liability for any inaccuracies or omissions.