



NTNU



Norwegian University of  
Science and Technology

# Project Presentation

TDT4173 Machine Learning  
Fall 2023

Ruslan Khalitov  
PhD Candidate, Department of  
Computer Science

# Outline

- Purpose and task overview
- Aneo
- Data and task details
- Evaluation
- Virtual Teams
- What should be delivered?
- Grading

# Purpose

Learn how to solve real-world problems using machine learning models.

## Learning goals:

- Data manipulation
- Feature engineering
- ML model training lifecycle
- Close the gap between «toy» and real ML problems
- Solving ML competitions

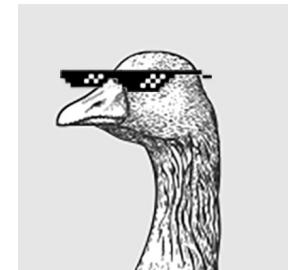
# The task

- Accurately solve a prediction problem
- Real-world data
- You will compete with Virtual Teams (VTs) we've created
- Better score → More VTs defeated → higher grade

# Kaggle

# kaggle

- The competition is hosted on Kaggle InClass
- Kaggle is the largest platform for data science competitions
- Why Kaggle?
  - You can check solutions automatically
  - Allows private InClass competitions for students
  - User-friendly interface
  - World-known platform



A close-up, profile shot of a woman's face and hair. Her hair is dark brown and appears to be blowing in the wind, with some strands catching the light and appearing golden. The background is a soft, out-of-focus gradient from light blue to white.

ANEO

# About Aneo

## Norway's second largest wind power company



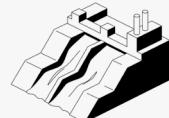
928 GWh annual production

## Leading national centre of excellence for energy management of renewable production



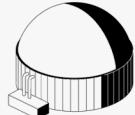
6TWh under management, 1 TWh of which is for third parties

## Owns 18,98% of TrønderEnergy's hydropower production



346 GWh annual production

## Ownership in leading Nordic biogas company

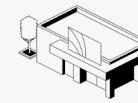


Substantial ownership share in Scandinavian Biogas Fuels (listed in Stockholm)

## Innovative downstream services



**1 Aneo Mobility**  
EV charging solutions  
for housing associations  
and commercial  
customers



**2 Aneo Retail**  
Delivers and manages  
sustainable energy  
infrastructure for retail



**3 Aneo Build**  
Delivers and manages  
charging containers for  
electrification of  
building and  
construction



**4 Aneo Industry**  
Energy saving and  
zero emission  
solutions for process  
industry



**5 Aneo Real Estate**  
Solar energy solutions for  
commercial property

# Artificial Intelligence at Aneo

**Strong capabilities in AI** are the core of Aneo's digital platform.

We leverage the value of data and AI through increased **quality and efficiency** in the company's core processes and activities.

With a **highly skilled internal AI team**, combined with domain experts we enable process improvement throughout the entire value chain.

**Internal AI team** develops energy forecasting, optimization, trading and control systems



14

Employees

6

PhDs

4

PhD  
students

# AI internships and master projects at Aneo

## Relevant topics

- Probabilistic time-series forecasting of energy consumption and production
- Condition-based monitoring for wind turbines
- Intelligent control of buildings with model predictive control and reinforcement learning
- Physics-informed machine learning for heating and ventilation modeling



### Summer Internship 2024 - AI and Product Development

Aneo

[Apply for position](#)

Aneo is a renewable energy company with headquarters in Trondheim, established in 2022 by TrønderEnergi and HitecVision. Aneo invests heavily in wind power, solar power, electrification and energy efficiency in the Nordic region.

Artificial intelligence (AI) and software development play key roles in the necessary renewable energy transition due to the increasing need for automation and data-driven decision making. Therefore, Aneo has an AI and product development department dedicated to developing and maintaining systems for forecasting, optimization, simulation, and control of energy infrastructure including wind turbines, hydropower plants, construction sites, EV chargers and buildings. The department is a cross-disciplinary and international team, with strong backgrounds in machine learning, operations research, and software development. The team works full stack, from understanding business needs and data analysis in Python to deploying AI systems in production, setting up necessary DevOps and cloud infrastructure, and creating user interfaces.

As a summer intern at AI and Product Development, you will work in groups of 2-3 students on real-world problems related to the topics mentioned above, in close collaboration with both other students and full-time employees.

All summer students at Aneo will participate in our summer student program, which includes introductions to the power industry and social events throughout the summer. Additional social events are organized by WorkationTrd, where you also meet interns from other companies in Trondheim.

Application de...	12.11.2023	Positions:	6	Webcruiter ID:	4705302760
Employer:	Aneo	Full-time/part...	Full-time	Social sharing :	<a href="#">Twitter</a> <a href="#">LinkedIn</a> <a href="#">Facebook</a> <a href="#">Email</a>
Town/city:	TRONDHEIM	Employment t...	Summer job		
Title:	Summer Internship 2024 - AI and Product Development	Percentage of ...	100		

<https://201607.webcruiter.no/Main/Recruit/Public/4705302760>

# Outline

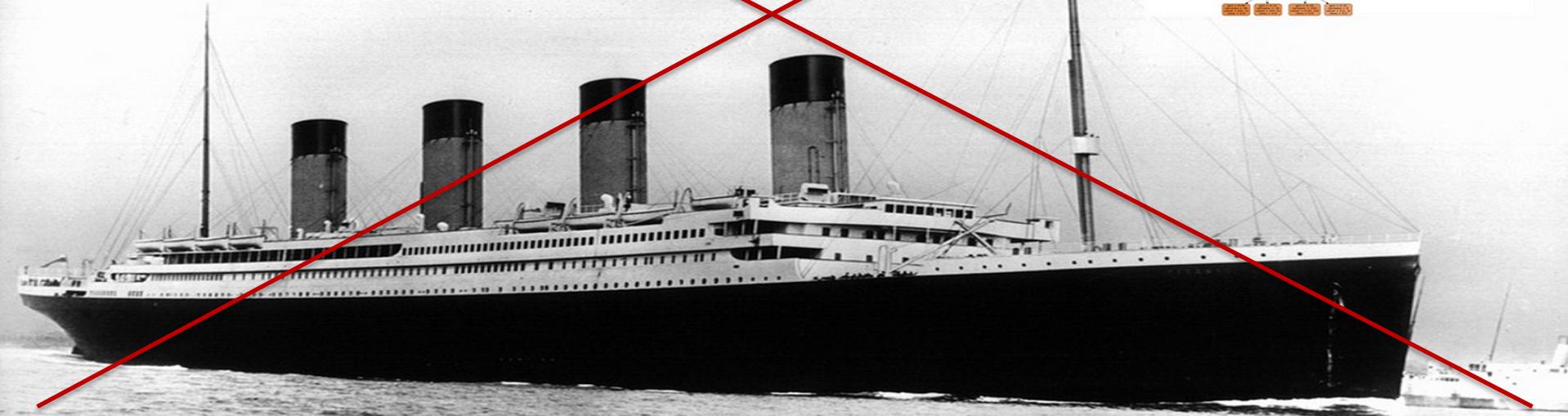
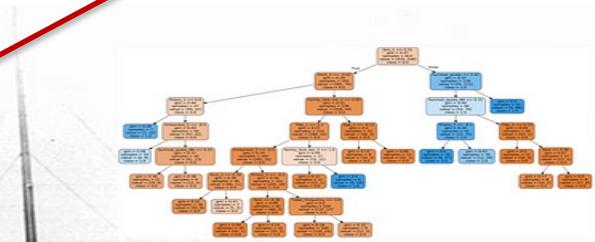
- Purpose and task overview
- Aneo
- Data and task details
- Evaluation
- Virtual Teams
- What should be delivered?
- Grading

# Task and data details

Task: Predict survival of Titanic passengers

## Machine Learning with the Titanic Dataset

An end-to-end guide to predict the Survival of Titanic passenger



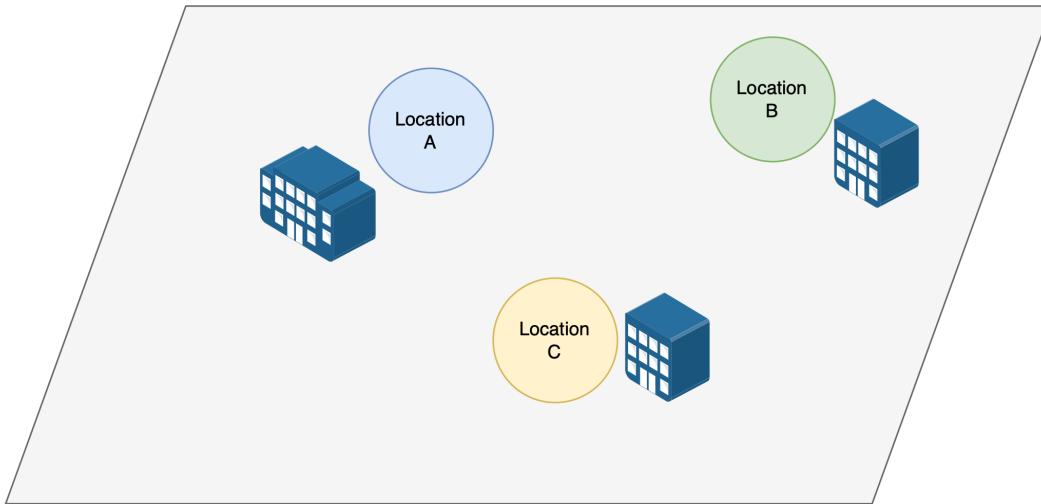
# Task and data details

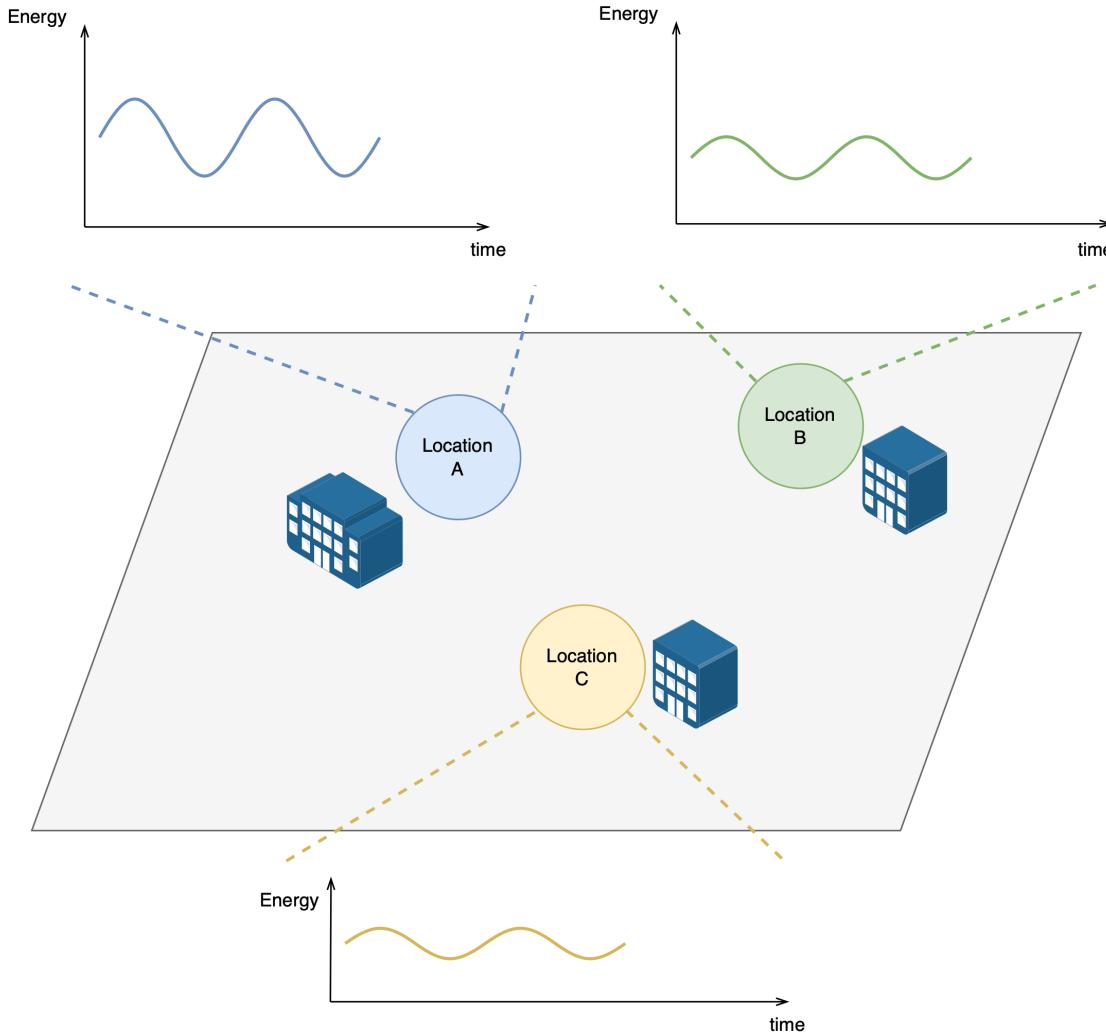
**Task:**  
**build a model that can  
predict solar energy production  
for every hour of the next day**

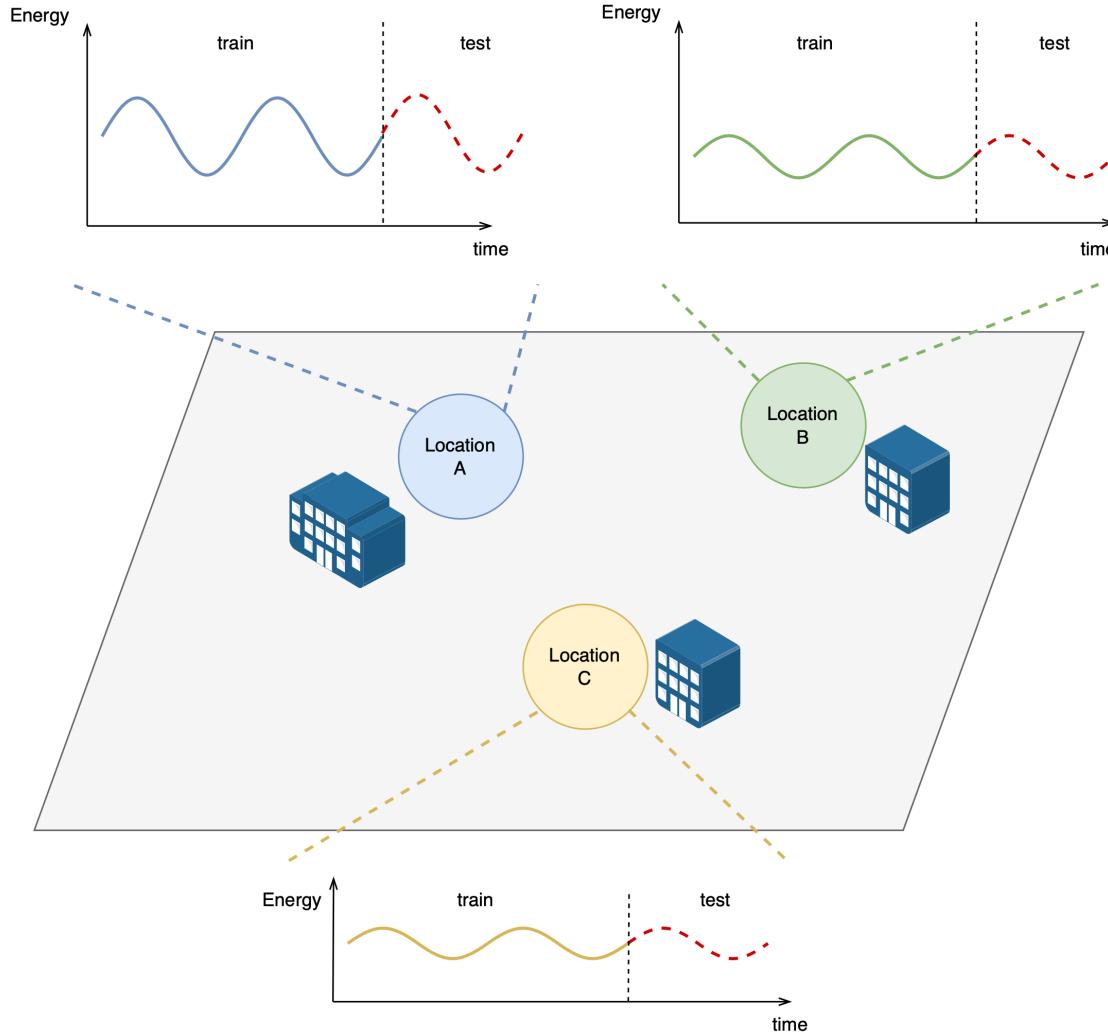
**Problem type:**  
**time-series forecasting,  
regression problem**

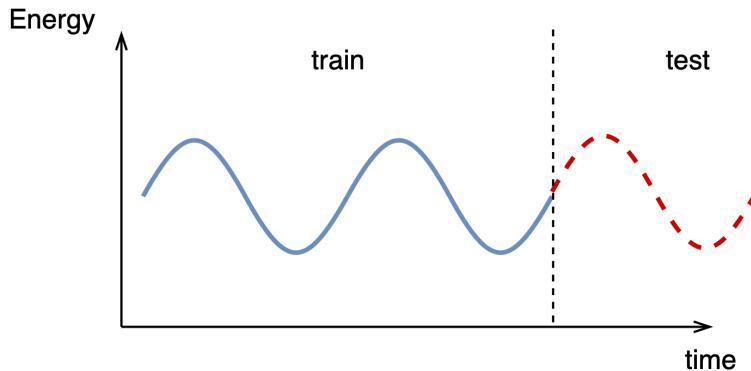


**Close to realistic  
production setup!**







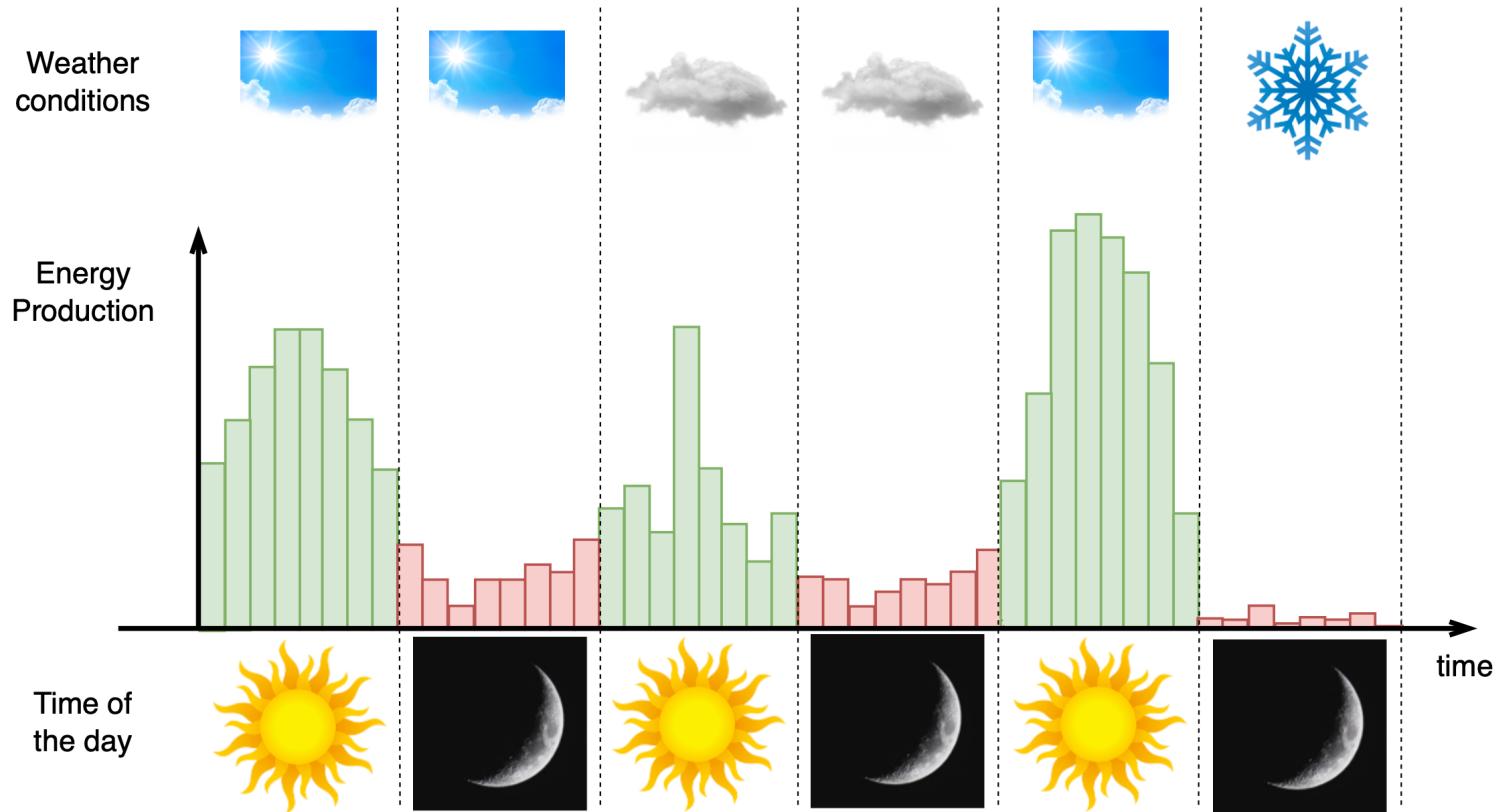


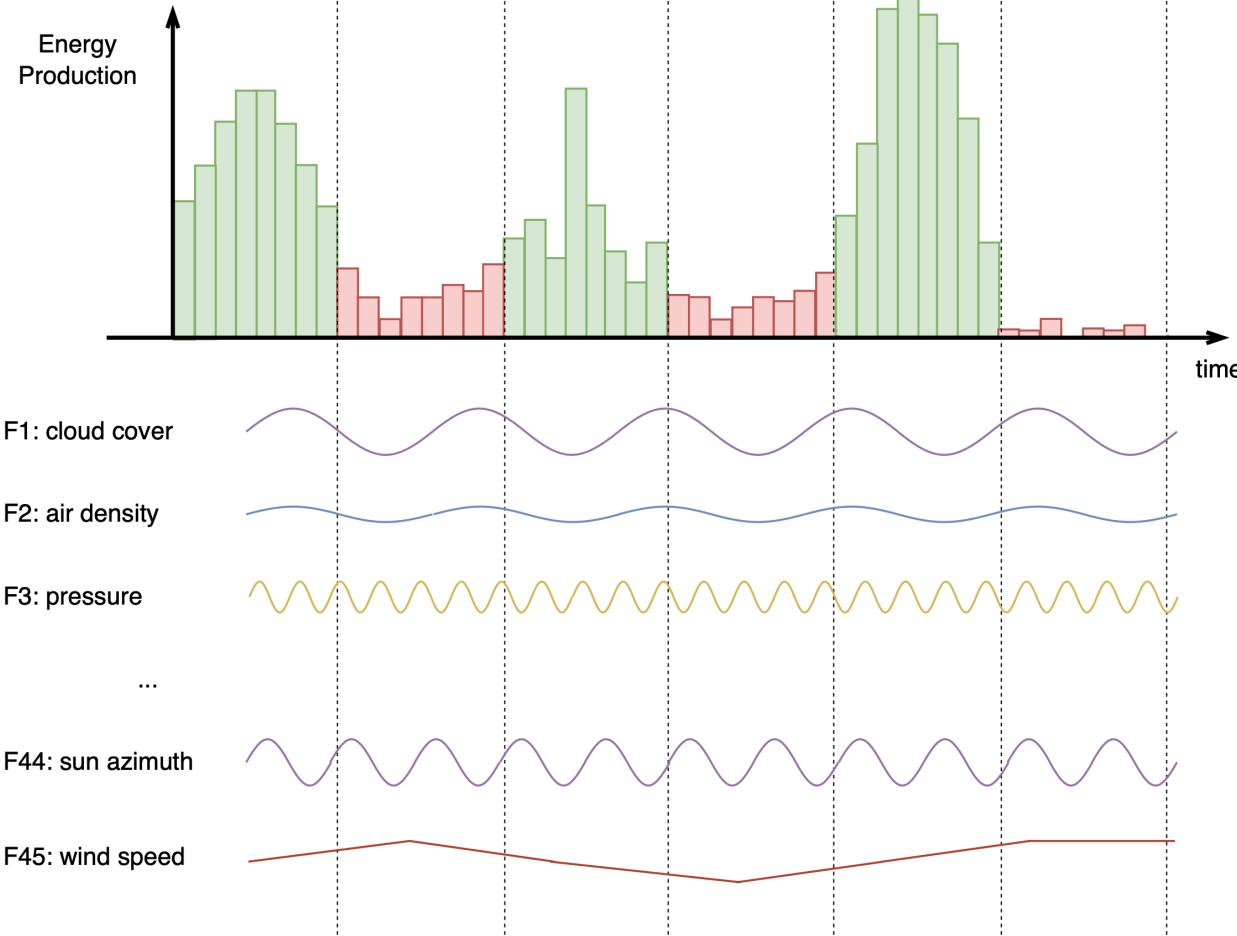
Standard TS forecasting:

Test period can be predicted accurately using only historical data.

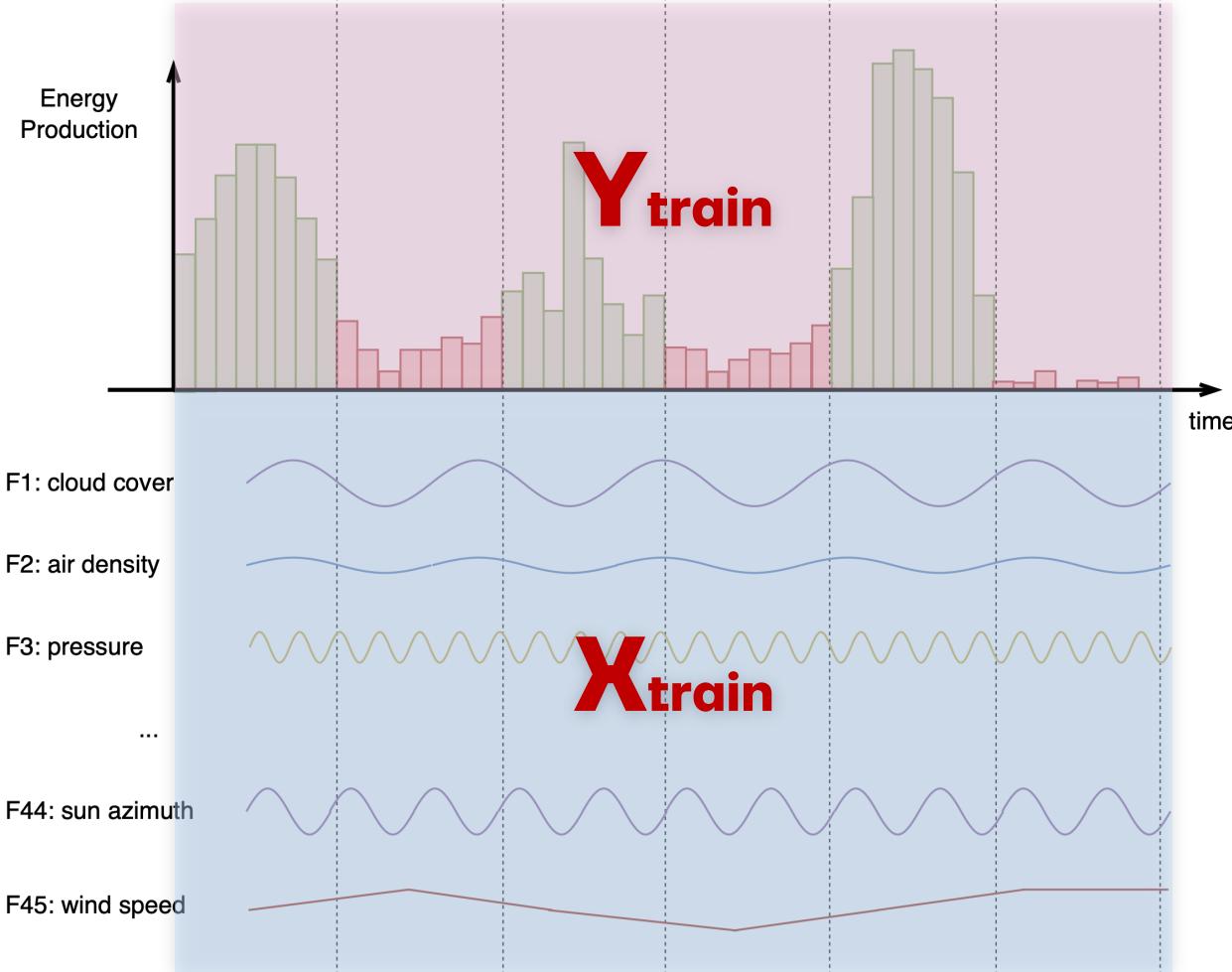
Our task is not that straightforward.

# Solar energy is highly dependent on weather conditions and time

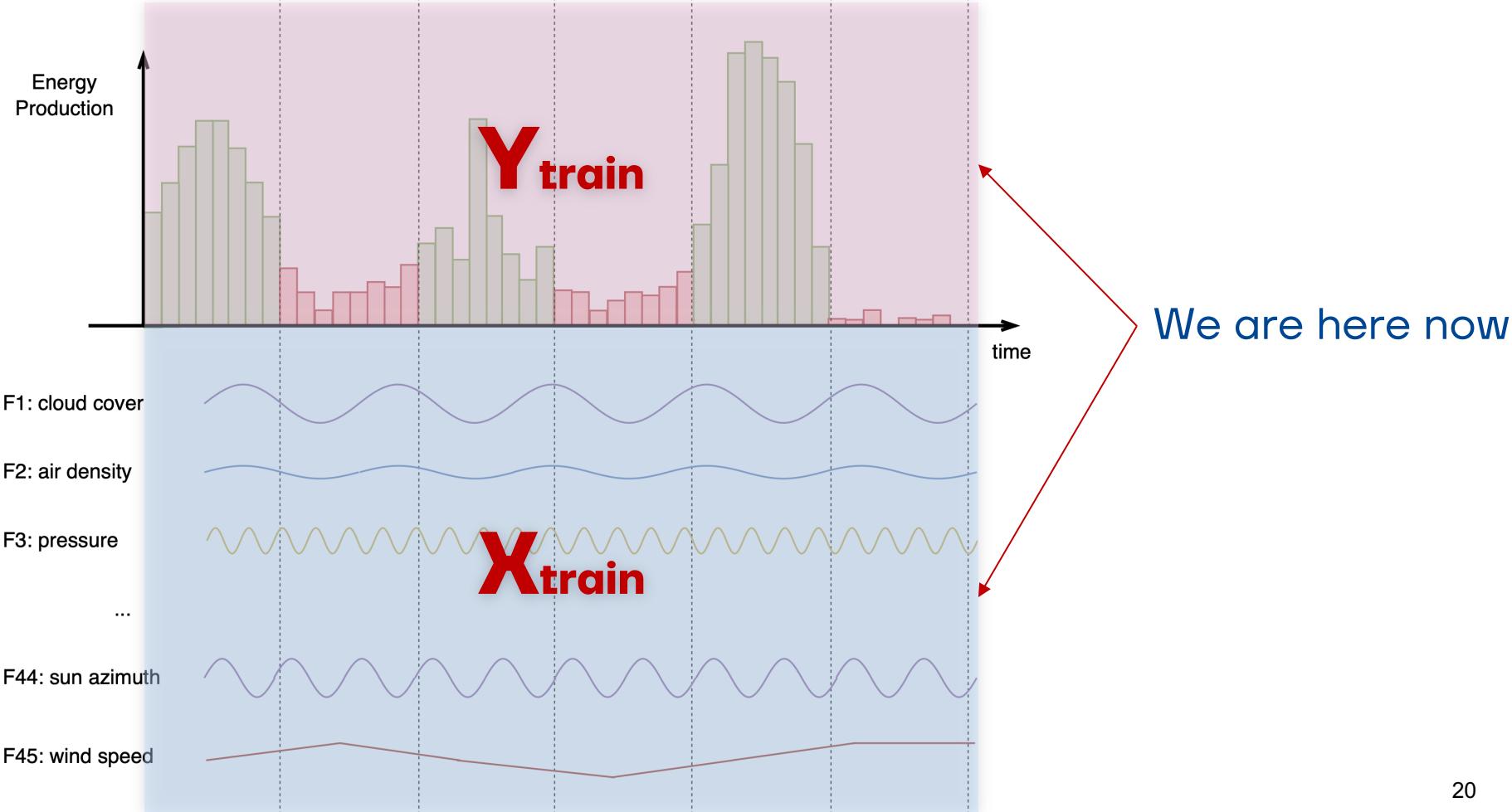


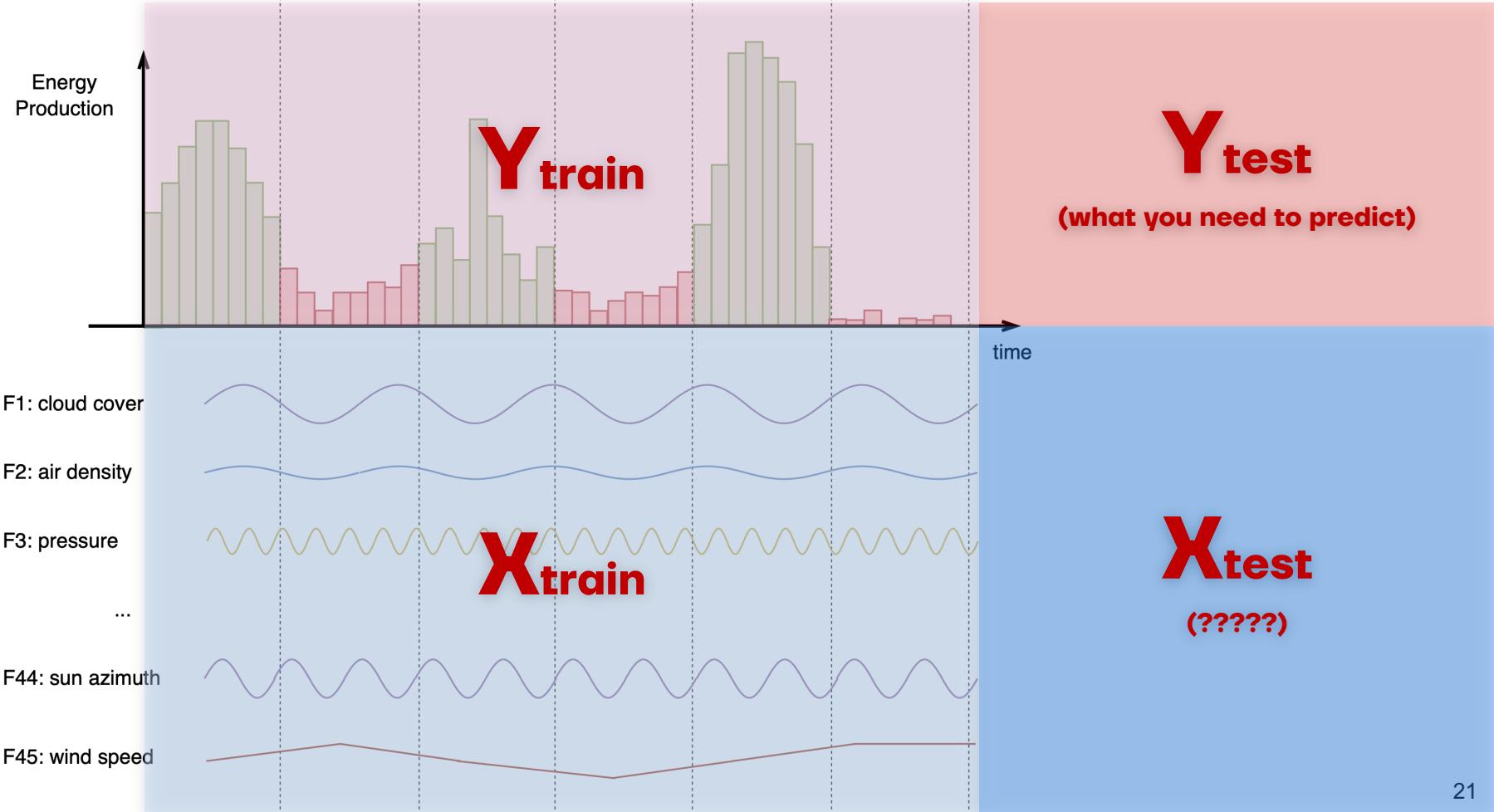


We give you 45  
weather related  
features observed  
historically



We give you 45  
weather related  
features observed  
historically





**X<sub>test</sub>**

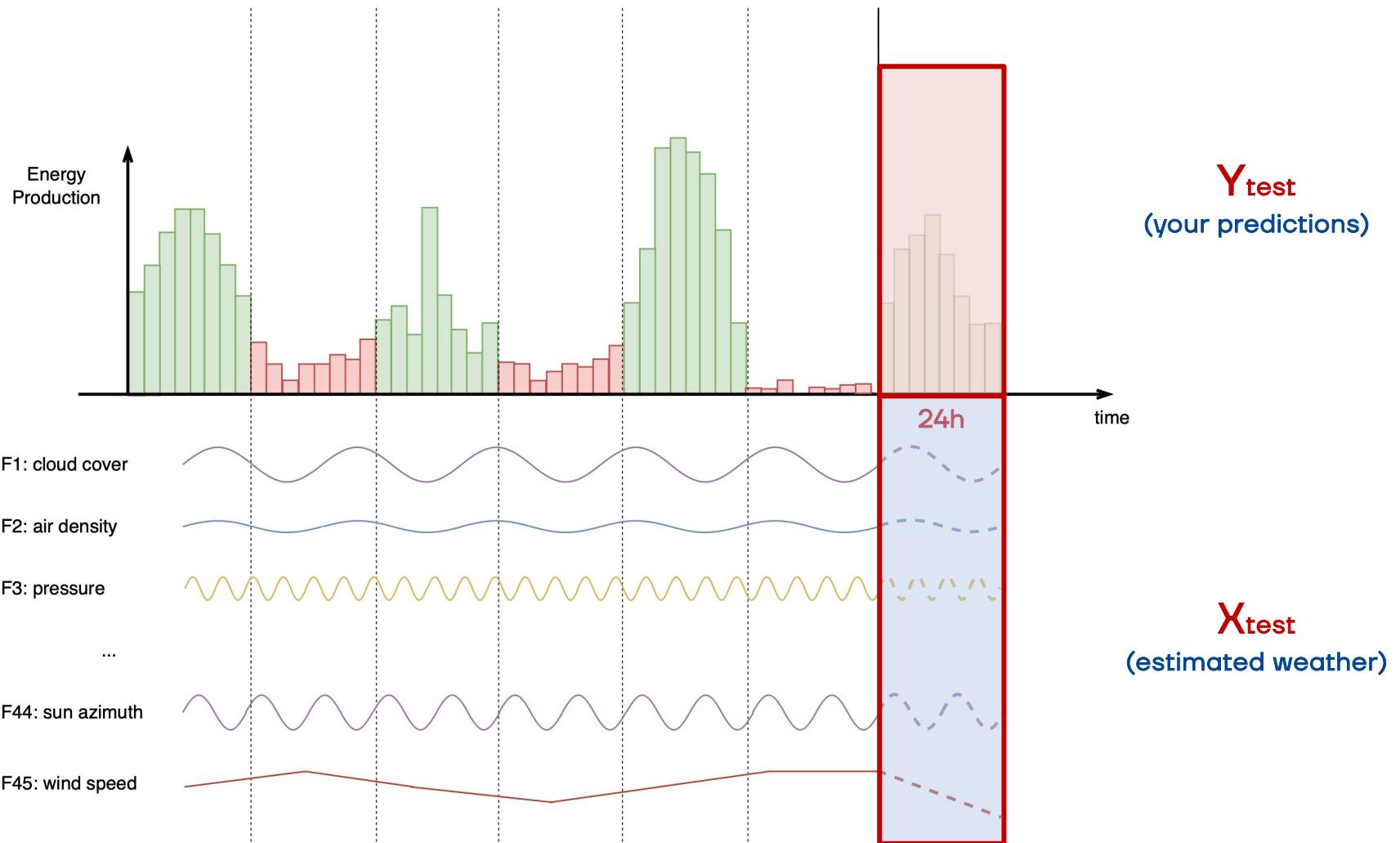
(don't know the future)

The problem #1:  
We do not know future weather

But we can use **weather predictions!**

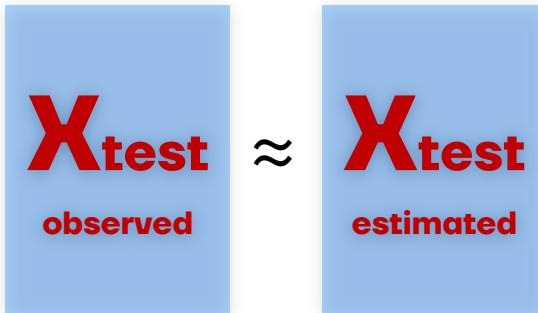
Every day a meteorologist sends us his weather predictions for the next day (24 h) for all 45 features.

Your system is supposed to take these predictions as **X<sub>test</sub>** and use it to predict solar energy production **Y<sub>test</sub>**.





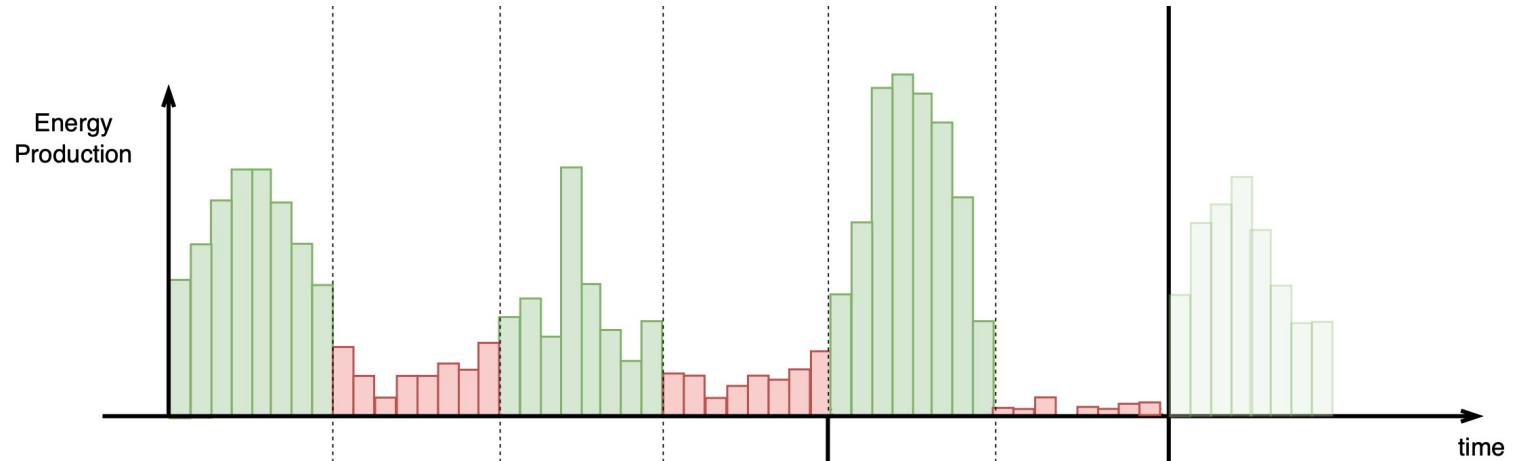
Weather  
predictions?



The problem #2:  
Weather predictions differ from real  
observed weather

If you train your model on **real weather**  
but predict future using **estimated weather**  
there will be a risk of an error.

Your model must **adapt to the distribution difference between real weather and estimated weather.**



F1: cloud cover

F2: air density

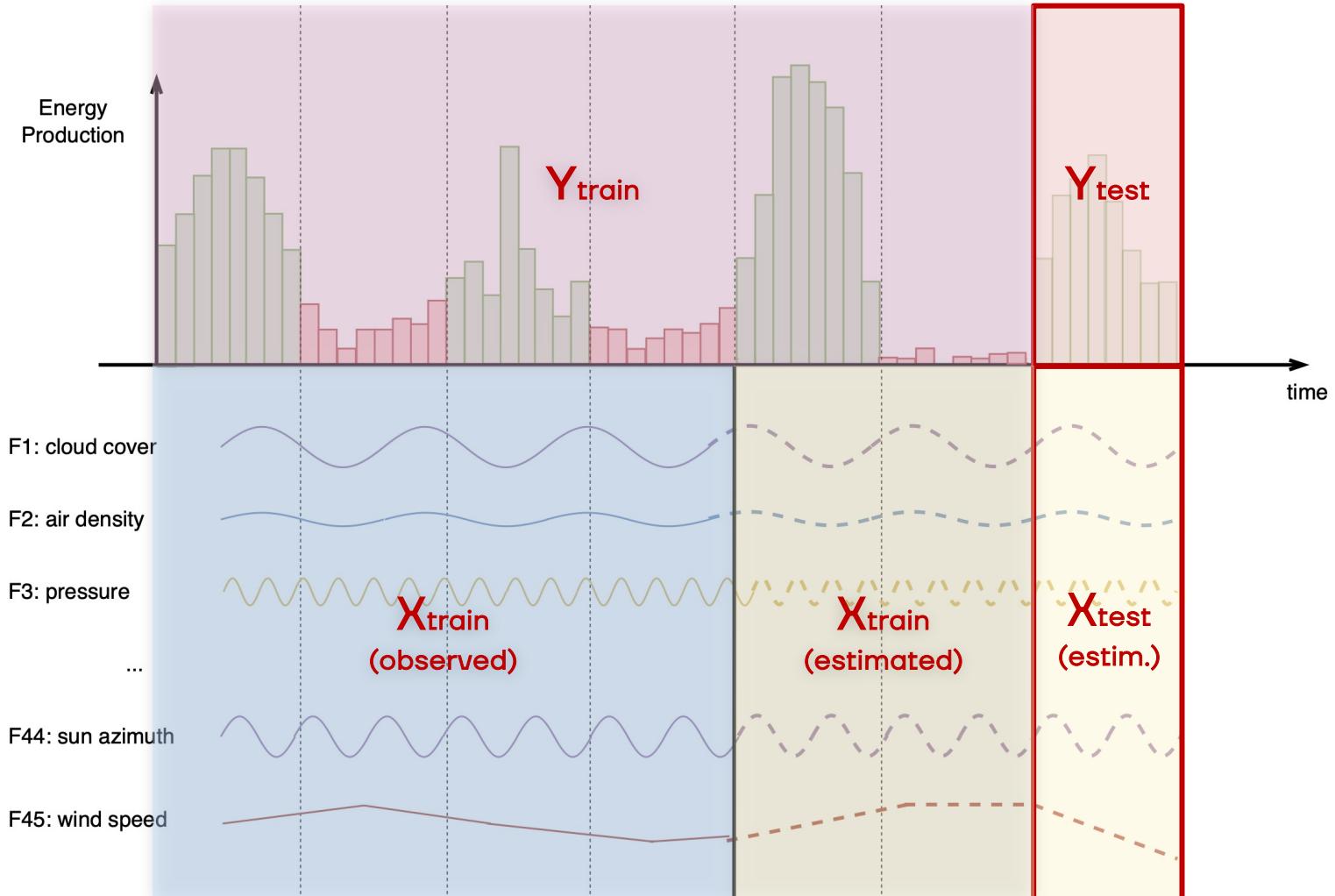
F3: pressure

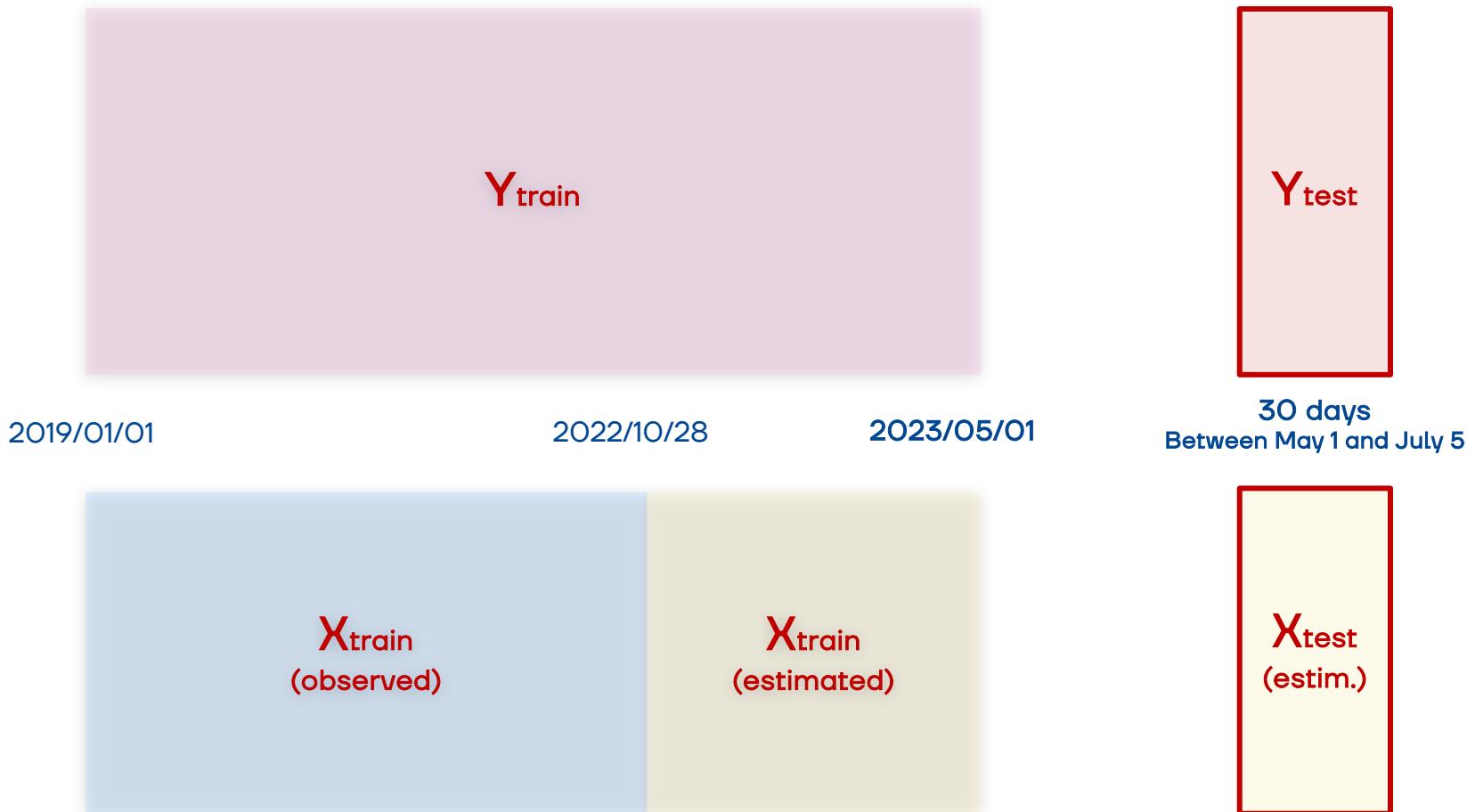
...

F44: sun azimuth

F45: wind speed





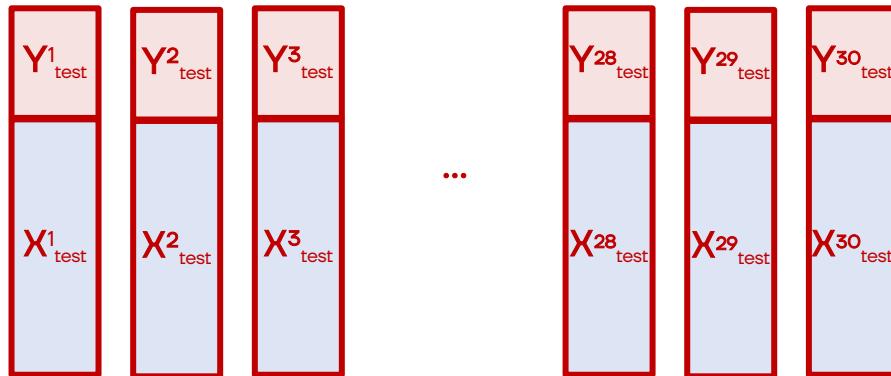


# Why 30 days???

You said 24 h prediction!!!

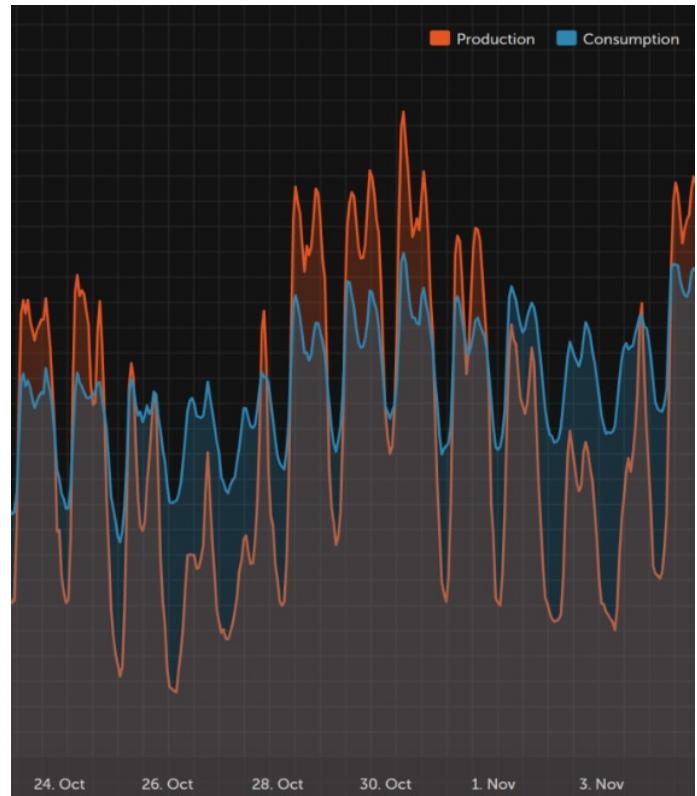
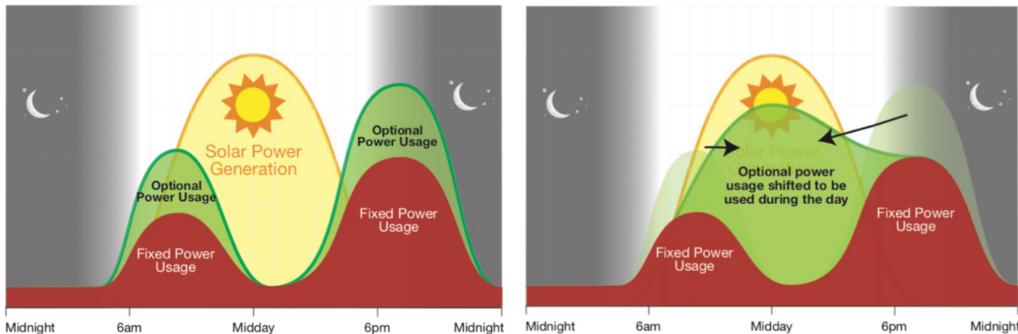


**We will imitate everyday predictions for 30 days.  
Imagine you trained a model and let it work without  
your supervision.**



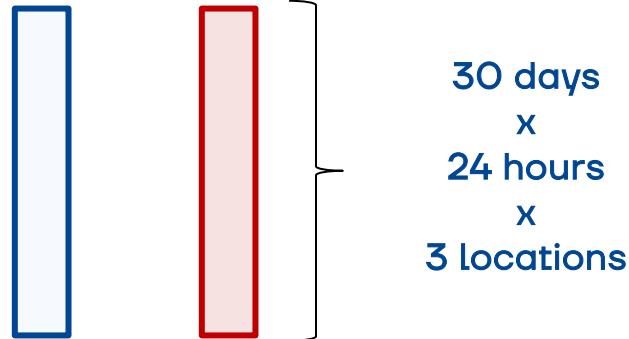
# Why solar energy prediction is important for Aneo?

- Aneo aims to develop up to 2 TWh of solar power by 2030
- Forecasts are required to balance production and consumption, keeping the grid frequency stable and electronic devices working properly
- We can use more solar energy instead of fossil fuels by moving consumption to hours with a lot of solar energy



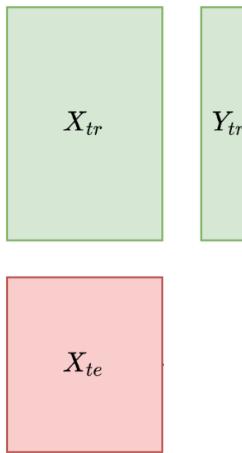
# Evaluation metric

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

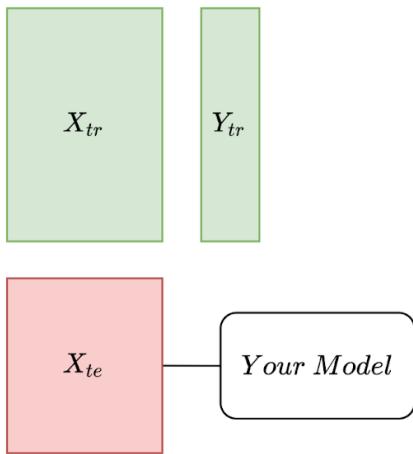


Mean Absolute Error

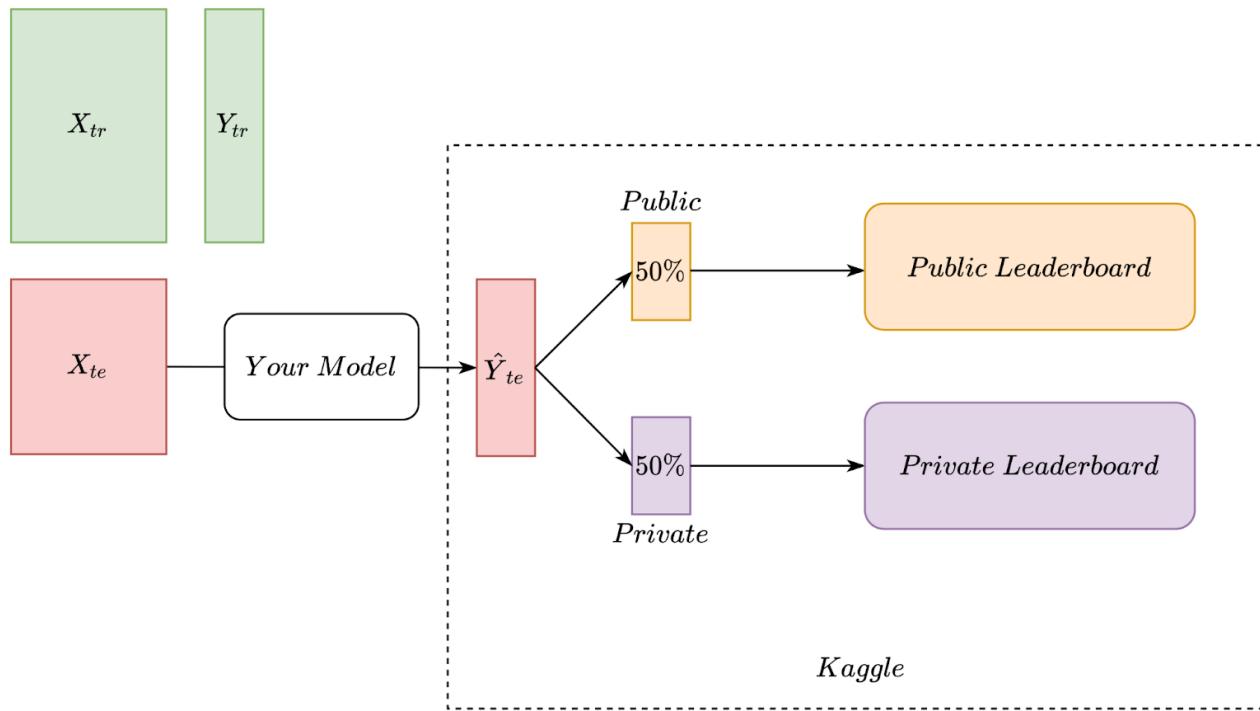
# Evaluation. Public/Private



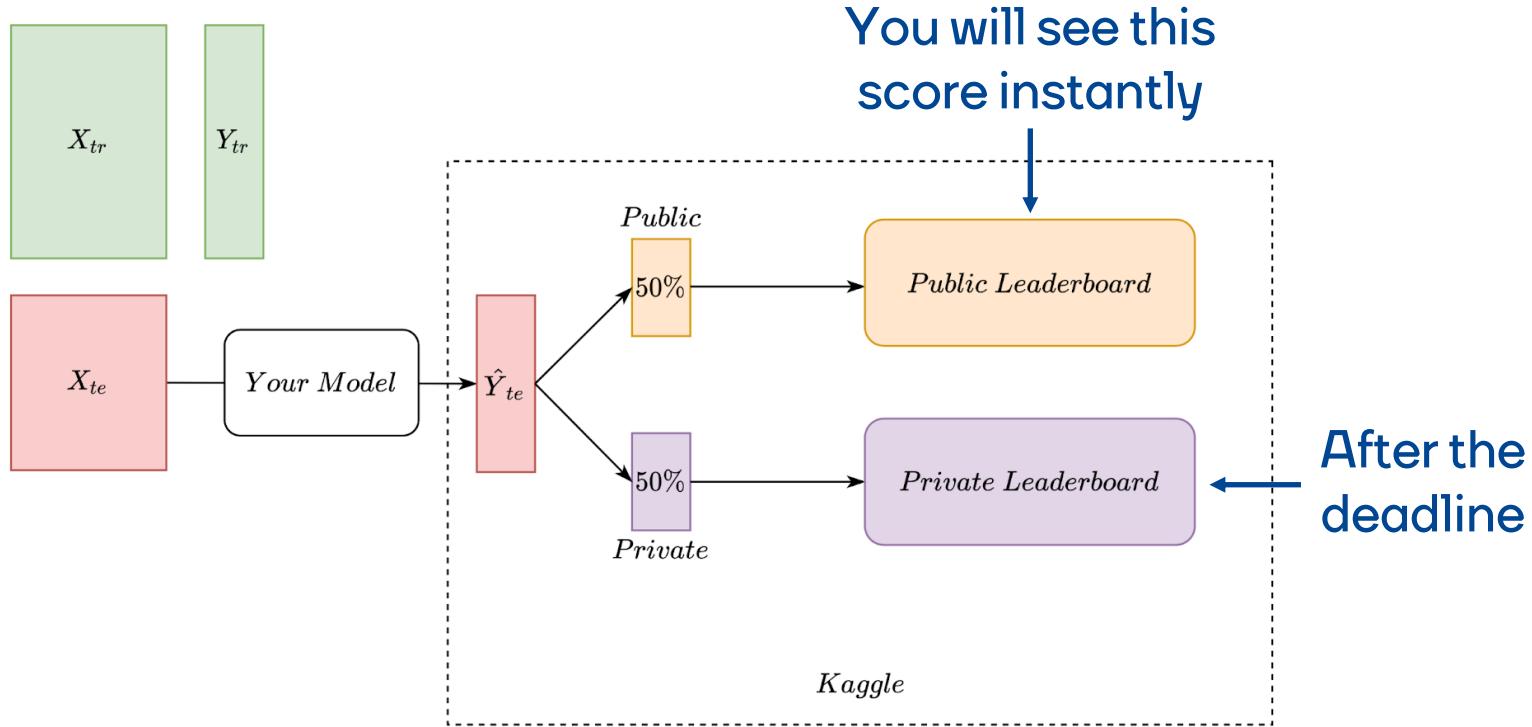
# Evaluation. Public/Private



# Evaluation. Public/Private



# Evaluation. Public/Private



# Evaluation. Public/Private

- You don't know the exact Public/Private split
- You can see your **estimated score** on the test set instantly
- You can submit **max 5 times** per day
- **The final score of your solution is your Private score**
- Before the competition end you have to select 2 solutions:  
Tip: select diverse solutions

# Evaluation. Public/Private

- You have extra information from the public leaderboard (50%)
- **The best public score does not necessarily correspond to the best private score**
- **Do not overfit on the public leaderboard score**

This leaderboard is calculated with approximately 49% of the test data. The final results will be based on the other 51%, so the final standings may be different.

#	Team	Members	Score	Entries	Last	Solution
1	Group 32	  	0.63613	127	10mo	
2	Group 91		0.64380	52	10mo	
3	[49] Team Tøffel	  	0.64704	138	10mo	
4	[1] Do u even overfit bro?	  	0.64801	4	10mo	
5	Group 37	 	0.65213	61	10mo	
6	Gruppe 5	  	0.65304	114	10mo	
7	[24]outsourced sweatshop	  	0.65607	19	10mo	
8	Group 9	  	0.66323	55	10mo	
9	[VT] Quan Chi [Group 19]	  	0.66586	84	10mo	
10	[64]Feature overengineering	  	0.66849	48	10mo	

The private leaderboard is calculated with approximately 51% of the test data.  
 This competition has completed. This leaderboard reflects the final standings.

#	△	Team	Members	Score	Entries	Last	Solution
1	—	Group 32		0.65205	127	10mo	
2	▲ 2	[1] Do u even overfit bro?		0.65990	4	10mo	
3	▼ 1	Group 91		0.66028	52	10mo	
4	▼ 1	[49] Team Tøffel		0.66237	138	10mo	
5	—	Group 37		0.66246	61	10mo	
6	▲ 1	[24]outsourced sweatshop		0.66890	19	10mo	
7	▼ 1	Gruppe 5		0.66911	114	10mo	
8	▲ 3	erasmus_40		0.67964	122	10mo	
9	▼ 1	Group 9		0.67981	55	10mo	
10	▼ 1	[VT] Quan Chi [Group 19]		0.68006	84	10mo	

[Public Leaderboard](#)[Private Leaderboard](#)

This leaderboard is calculated with approximately 50% of the test data.

[↓ Raw Data](#)[⟳ Refresh](#)

The final results will be based on the other 50%, so the final standings may be different.

■ In the money ■ Gold ■ Silver ■ Bronze

#	Team Name	Notebook	Team Members	Score ⓘ	Entries	Last
1	X5 (for cheating)			4.45806	120	2y
2	[ods.ai] Anton Chikin			4.67581	2	2y
3	In			4.81140	10	2y
4	VyD			5.00859	15	2y
5	Born to sleep	</> Fork of Fork of Me...		5.17518	50	2y
6	puppy play	</> albumentations		5.27418	14	2y
7	Chris Deotte			5.41850	21	2y
8	Karsten Tiemann			5.53078	50	2y
9	Madao			5.60532	9	2y
10	Siavash			5.75288	34	2y

[Public Leaderboard](#)[Private Leaderboard](#)

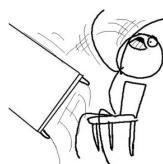
This competition's private leaderboard was evaluated by a private metric. This competition has closed and is no longer accepting submissions.

[⟳ Refresh](#)

This competition has completed. This leaderboard reflects the final standings.

■ In the money   ■ Gold   ■ Silver   ■ Bronze

#	△pub	Team Name	Notebook	Team Members	Score ⓘ	Entries	Last
1	▲ 438	Tom	</> GAN dogs starter B...		55.42142	25	2y
2	▲ 372	[kaggler-ja] yabea	</> [Sub] RaLS BigGAN...		56.47019	39	2y
3	▲ 407	Ili			70.41224	32	2y
4	▲ 362	bestfitting	</> i96_1e-4_6e_4_bs3...		77.80956	33	2y
5	▲ 367	[ods.ai] Dmitry Vorobiev	</> doggies BigGAN su...		82.30464	78	2y
6	▲ 431	Theo Viel	</> ProCGan 30		82.78395	15	2y
7	▲ 209	Doge			83.77130	59	2y
8	▲ 441	tkato	</> submit_ac_2		89.17595	38	2y
9	▲ 460	Mark Peng	</> small-stylegan-v6...		89.64117	43	2y
10	▲ 457	Johannl	</> SN-DCGAN		95.45683	43	2y



# Outline

- Purpose and task overview
- Aneo
- Data and task details
- Evaluation
- Virtual Teams
- What should be delivered?
- Grading

# Outline

- Purpose and task overview
- Aneo
- Data and task details
- Evaluation
- Virtual Teams
- What should be delivered?
- Grading

# Each group

1. Selects two predictions on kaggle before the deadline
  2. Uploads **Two short** Jupyter notebooks:
    - Only necessary steps to reproduce your selected predictions
    - Naming: «short\_notebook\_1.ipynb»
    - Put your group number!
  3. **A long** Jupyter notebook:

Contains all attempts in your group work (EDA, all models, algorithms, feature engineering, results interpretation)
- **Submissions to Kaggle, notebooks to Blackboard**

# What can I use?

- Any languages, tools, platforms, AutoML (offline), libraries, file formats during development.
- But you should use Jupyter Notebook for your delivery
- You **can not** use external data, other than we provided
- Writing massive data in code is not allowed

# Late submissions

- **Deadline: November 12, 2023 22:00 Central European Time (GMT+2)**
- Strict.
- Up to this time you should submit both submissions to Kaggle and notebooks to Blackboard
- Late submission deadline: 15/11/2023 22:00 CET **(-10 points)**
- Even later = the whole group fails the course

# Grading

- Project points = base points + possible deductions

Letter	Points
A	89-100
B	77-88
C	65-76
D	53-64
E	41-52
F	0-40

Base points are proportional to the number of VTs defeated:

- max 100 (all VTs)
- min 41 (defeat 1 VT)

We use the best (out of 2 submitted) to evaluate your solution

# Grading

## Possible deductions:

- Pass Individual assignment in the second chance (-5)
- Late submission (-10)
- No exploratory data analysis (EDA) (-3)
- Only one predictor is used (-3)
- No feature engineering (-3)
- No model interpretation (-3)

All deductions are binary (full or no deduction)

# Computing resources

NTNU computing resources (TAs: Erlend Hordvei and Erlend Heir)

Google Cloud Credits (TAs: Jonas Nordsrøm and Erik Svendsmark)

The detailed instructions are going to be published on Blackboard soon...



NTNU



Norwegian University of  
Science and Technology

# Project Presentation

TDT4173 Machine Learning  
Fall 2023

Ruslan Khalitov  
PhD Candidate, Department of  
Computer Science