Algorithm Reliability in the Development of Offshore Wind Turbine Digital Twins

Presented by:

Tolga YALÇIN

In collaboration with:

Hector del Pozo Gonzalez, Pol Paradell-Solà and Jose Luis Dominguez-Garcia

Deepwind 2024 18th January 2024 Trondheim, Norway

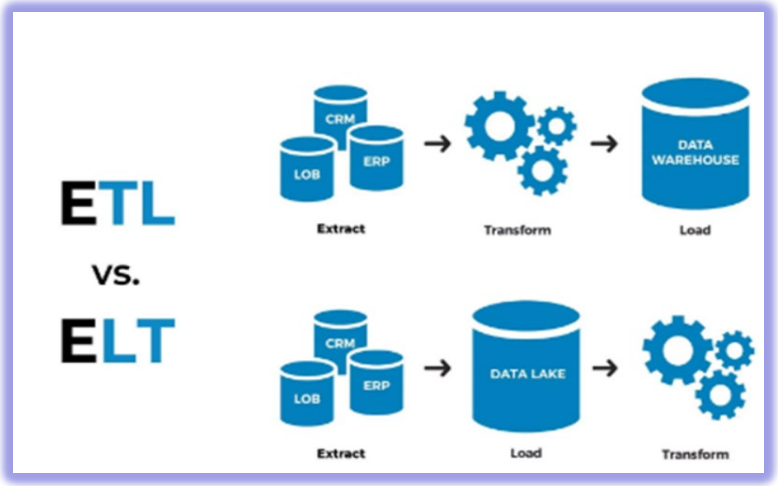
Presentation Outline

1. **Project Data**
   - The SCADA dataset is introduced.
2. **Feature Selection**
   - Introduce algorithms for decreasing the complexity while keeping information level.
3. **Machine Learning (ML) Models**
   - Different regression algorithms are tested to find the best performance.
4. **Digital Twin Concept (DT)**
   - A DT concept is presented to enable real-time monitoring of the offshore wind turbine through a virtual mode.
5. **Results:**
   - A brief comparison of DT's concept is introduced.

Shaping Energy for a Sustainable Future

Project Data (I): Description

- ✓ The study leverages SCADA data from the 7MW Levenmouth Offshore Wind Turbine.
- ✓ The database comprises 574 features per second and spans over a period of more than 2 years.
- ✓ The database comprises 517 float, with 56 of them being of integer and object types.

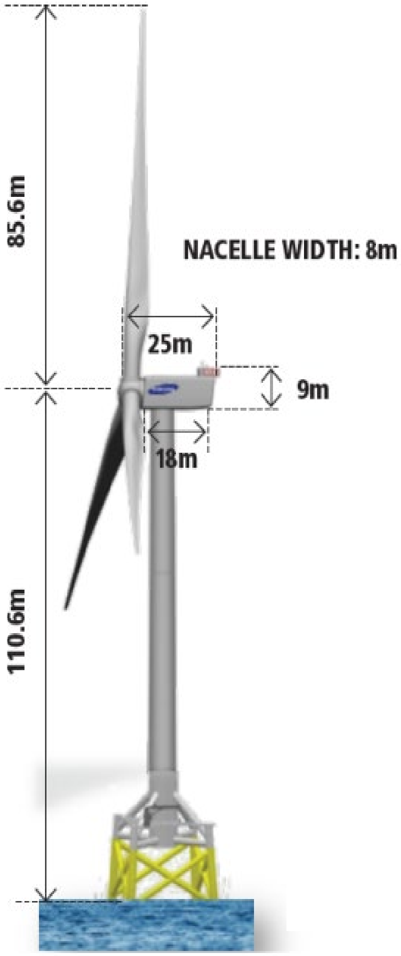


- Which one is better for this project?
- ETL is giving opportunity to work with big data.

Shaping Energy for a Sustainable Future

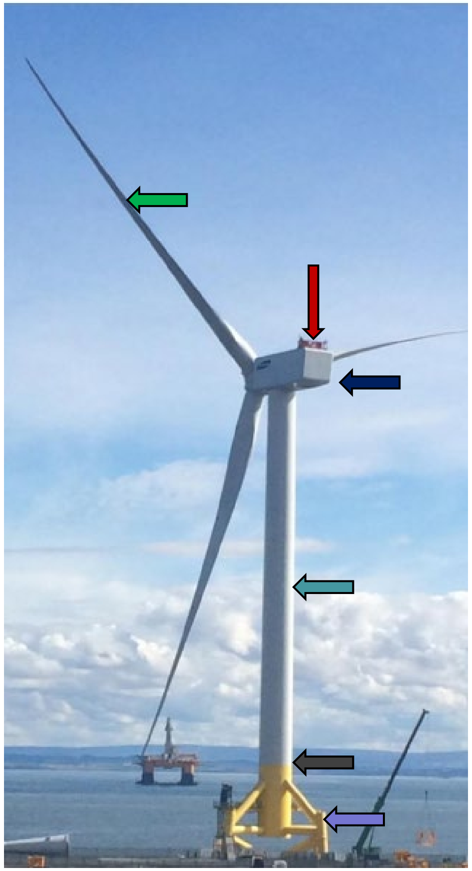
Project Data (II): Time Series

High Frequency Raw Data Signals → Structuring Data → Finding Missing values → Denoising Features → Removing Outliers



Available Measurements

1. Controller Signals
2. Acceleration sensors on the tower and blades
3. Multiple strain gauges on the tower and blades
4. Wind measurements
5. Atmospheric measurements
6. Sea-State measurements
7. Other critical structural measurements
8. Other electrical signals



- Lidar
- Power Train
- Pitch System
- Strain
- Accelerometer
- Tower
- Transition Piece
- Jacket

Figures taken from: https://www.sintef.no/globalassets/project/eera-deepwind-2020/presentations/g1_mckeever.pdf

Feature Selection

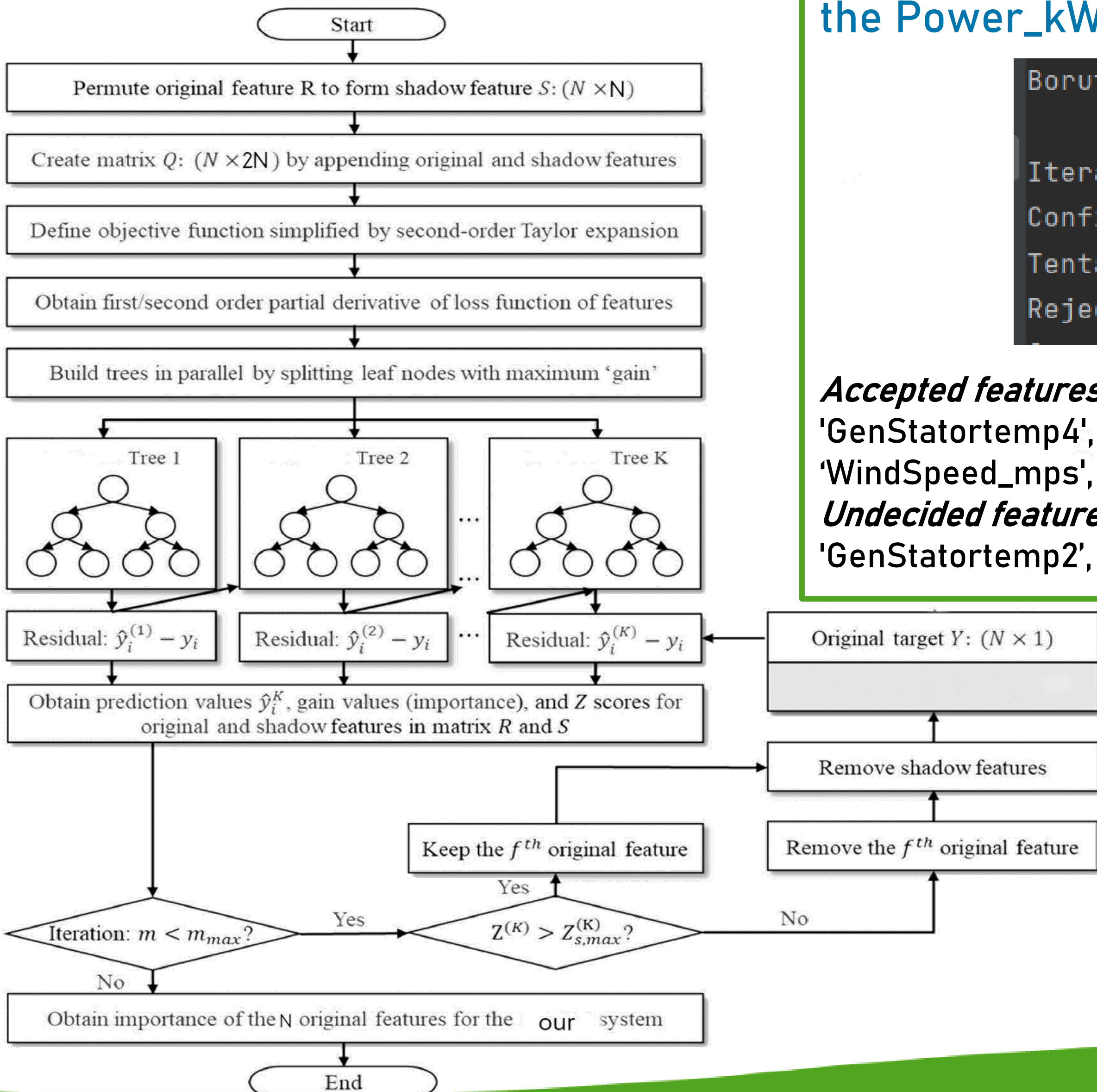
- ## Correlation Matrix:

Is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables.

	SubPcsPrivlgctTemp	GBoxCoolingLinetemp	GearBoxTemperature_DegC	GBoxOpShaftBearingTemp1	GBoxExtnlHeatertemp	GBoxDisttemp1	GBoxTanktemp2	GenStatortemp6	GenStatortemp4	GenStatortemp2	Hputemp2	NacOutsidetemp1	AmbTemp_DegC	AmbTemp	GenInternalAirMidtemp2	Pitch_Deg	RotorSpeed_rpm	NacelleOrientation_Deg	Power_kW
SubPcsPrivlgctTemp	1.00	-0.30	0.81	0.86	0.81	0.80	0.81	0.86	0.86	0.86	-0.01	0.07	0.07	0.07	0.86	0.15	0.89	0.27	0.94
GBoxCoolingLinetemp	-0.30	1.00	-0.28	-0.31	-0.28	-0.27	-0.28	-0.35	-0.35	-0.35	0.39	0.50	0.50	0.50	-0.17	-0.10	-0.28	-0.47	-0.29
GearBoxTemperature_DegC	0.81	-0.28	1.00	0.98	1.00	0.99	1.00	0.93	0.93	0.93	0.08	0.09	0.09	0.09	0.92	0.22	0.73	0.34	0.68
GBoxOpShaftBearingTemp1	0.86	-0.31	0.98	1.00	0.98	0.98	0.98	0.94	0.94	0.94	0.06	0.07	0.07	0.07	0.94	0.21	0.78	0.34	0.73
GBoxExtnlHeatertemp	0.81	-0.28	1.00	0.98	1.00	0.99	1.00	0.94	0.93	0.94	0.08	0.09	0.09	0.09	0.92	0.22	0.73	0.33	0.68
GBoxDisttemp1	0.80	-0.27	0.99	0.98	0.99	1.00	0.99	0.92	0.92	0.92	0.08	0.10	0.10	0.10	0.91	0.24	0.72	0.34	0.67
GBoxTanktemp2	0.81	-0.28	1.00	0.98	1.00	0.99	1.00	0.93	0.93	0.93	0.08	0.09	0.09	0.09	0.92	0.22	0.73	0.34	0.68
GenStatortemp6	0.86	-0.35	0.93	0.94	0.94	0.92	0.93	1.00	1.00	1.00	0.03	0.06	0.06	0.06	0.96	0.20	0.75	0.33	0.79
GenStatortemp4	0.86	-0.35	0.93	0.94	0.93	0.92	0.93	1.00	1.00	1.00	0.03	0.06	0.06	0.06	0.96	0.19	0.75	0.33	0.79
GenStatortemp2	0.86	-0.35	0.93	0.94	0.94	0.92	0.93	1.00	1.00	1.00	0.03	0.06	0.06	0.06	0.96	0.20	0.75	0.33	0.79
Hputemp2	-0.01	0.39	0.08	0.06	0.08	0.08	0.08	0.03	0.03	0.03	1.00	0.20	0.20	0.20	0.13	0.05	0.00	-0.30	-0.01
NacOutsidetemp1	0.07	0.50	0.09	0.07	0.09	0.10	0.09	0.06	0.06	0.06	0.20	1.00	1.00	1.00	0.20	0.04	0.02	-0.15	-0.00
AmbTemp_DegC	0.07	0.50	0.09	0.07	0.09	0.10	0.09	0.06	0.06	0.06	0.20	1.00	1.00	1.00	0.20	0.04	0.02	-0.15	-0.00
AmbTemp	0.07	0.50	0.09	0.07	0.09	0.10	0.09	0.06	0.06	0.06	0.20	1.00	1.00	1.00	0.20	0.04	0.02	-0.15	-0.00
GenInternalAirMidtemp2	0.86	-0.17	0.92	0.94	0.92	0.91	0.92	0.96	0.96	0.96	0.13	0.20	0.20	0.20	1.00	0.23	0.75	0.24	0.78
Pitch_Deg	0.15	-0.10	0.22	0.21	0.22	0.24	0.22	0.20	0.19	0.20	0.05	0.04	0.04	0.04	0.23	1.00	0.12	0.23	0.11
RotorSpeed_rpm	0.89	-0.28	0.73	0.78	0.73	0.72	0.73	0.75	0.75	0.75	0.00	0.02	0.02	0.02	0.75	0.12	1.00	0.25	0.84
NacelleOrientation_Deg	0.27	-0.47	0.34	0.34	0.33	0.34	0.34	0.33	0.33	0.33	-0.30	-0.15	-0.15	-0.15	0.24	0.23	0.25	1.00	0.23
Power_kW	0.94	-0.29	0.68	0.73	0.68	0.67	0.68	0.79	0.79	0.79	-0.01	-0.00	-0.00	-0.00	0.78	0.11	0.84	0.23	1.00

Shaping Energy for a Sustainable Future

• Boruta



An example of the results running Boruta for the Power_kWh variable:

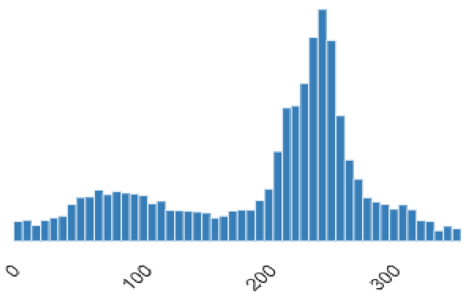
```
BorutaPy finished running.

Iteration:  100 / 100
Confirmed:  8
Tentative:  4
Rejected:   61
```

Accepted features: ['SubPcsPrivIgctTemp', 'GenStatortemp5', 'GenStatortemp4', 'Pitch_Deg', 'RotorSpeed_rpm', 'WindSpeed_mps', 'GenStatortemp6', 'WindSpeed1']
Undecided features: ['HtngWindSensor2', 'GenStatortemp3', 'GenStatortemp2', 'GenInternalAirMidtemp2']

Feature Analysis

NacelleOrientation_Deg
Real number (ℝ)

Distinct	159800	Minimum	0.0007103424
Distinct (%)	24.0%	Maximum	359.9998
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	201.27198	Memory size	10.1 MiB



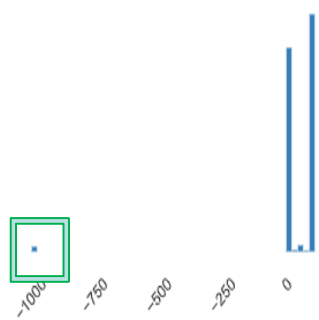
More details

Using statistic and histogram diagram to define the outliners and any unusual data.

Pitch_Deg
Real number (ℝ)

HIGH CORRELATION

Distinct	158645	Minimum	-1000.004
Distinct (%)	23.9%	Maximum	129.4825
Missing	0	Zeros	255
Missing (%)	0.0%	Zeros (%)	< 0.1%
Infinite	0	Negative	217747
Infinite (%)	0.0%	Negative (%)	32.7%
Mean	35.36796	Memory size	10.1 MiB



More details

IREC
Shaping Energy for a Sustainable Future

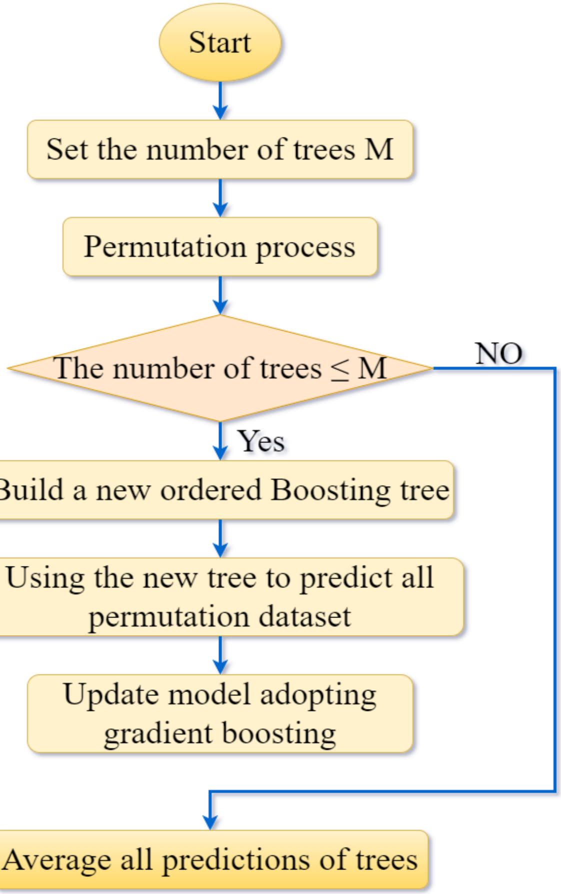
Machine Learning (ML) Models

- Inherent to their design, wind turbines showcase non-linear behavior, demanding swift adaptations in their systems.
- Today, with the surge in extreme natural events, ML algorithms must align with this reality.
- Given their role in control and monitoring, speed and stability are paramount.
- Consequently, we opted to forego overly complex or linear-based models, opting instead for the efficiency of the following four algorithms:
    1. CatBoost Regression
    2. Random Forest (RF) Regression
    3. Deep Neural Network Regression
    4. XGBoost Regression

IREC
Shaping Energy for a Sustainable Future

Machine Learning Models (I): CatBoost

CatBoost is a gradient-boosting library that uses decision trees as the base model. It is specifically designed to handle categorical variables and includes some other features such as handling missing values and built-in cross-validation.



MAE:50.74

- MSE:13196.93

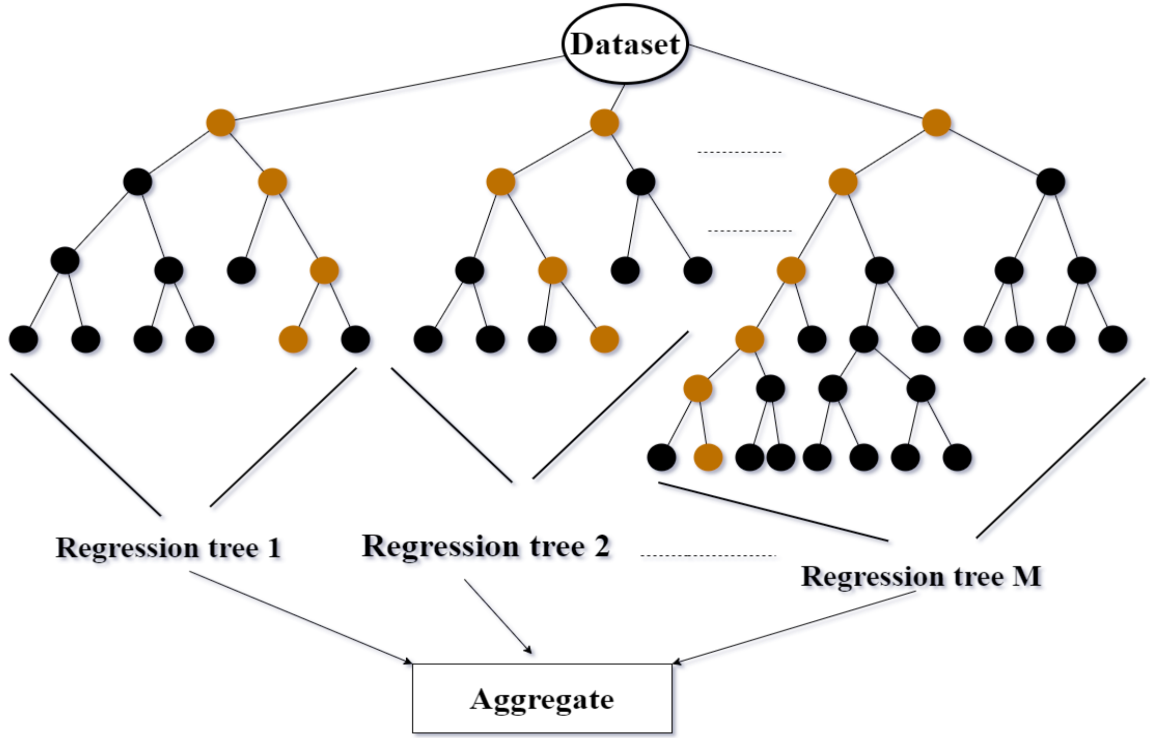
RMSE:114.88

- R-Square= 0.9973356088169721

Shaping Energy for a Sustainable Future

Machine Learning Models (II): Random Forest

Random forest is a Supervised Machine Learning technique that uses the group learning approach to do classification and regression.



MAE:40.04

- MSE:11417.72

RMSE:106.85

- R-Square= 0.9976948223386025

IREC
Shaping Energy for a Sustainable Future

Machine Learning Models (III): Artificial Neural Networks

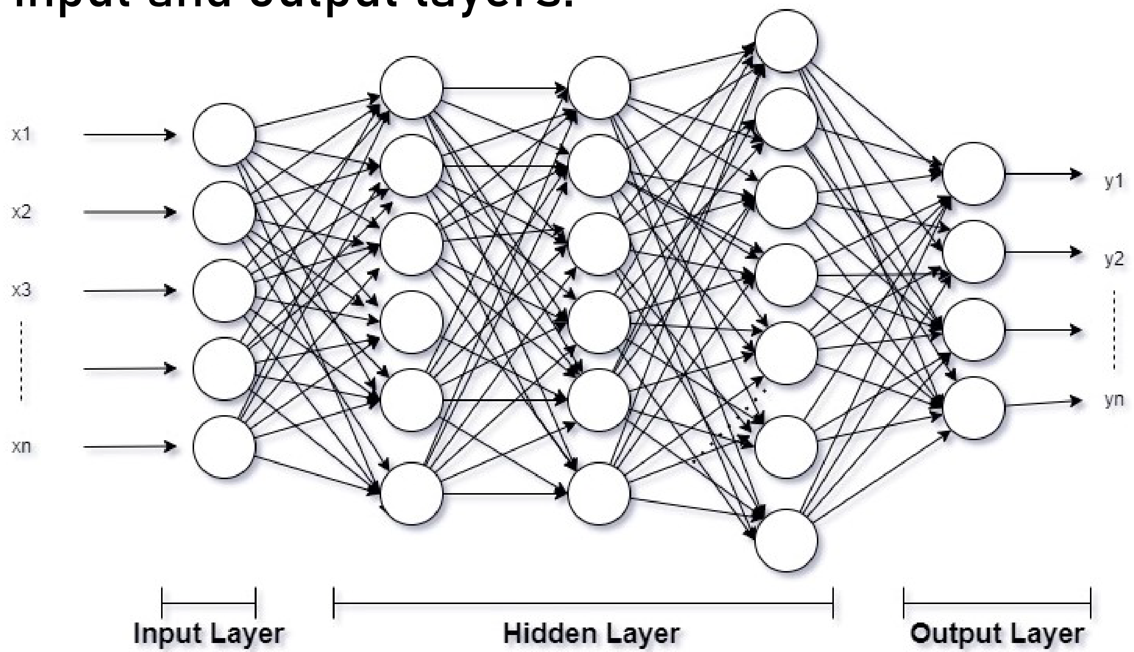
Deep neural networks (DNN) is a class of machine learning algorithms similar to the artificial neural network and aims to mimic the information processing of the brain. DNN share more than one hidden layer situated between the input and output layers.

MAE:46.94

- MSE:13703.76

RMSE:117.06

- R-Square= 0.9972332826361944



IREC
Shaping Energy for a Sustainable Future

Machine Learning Models (IV): XGBoost

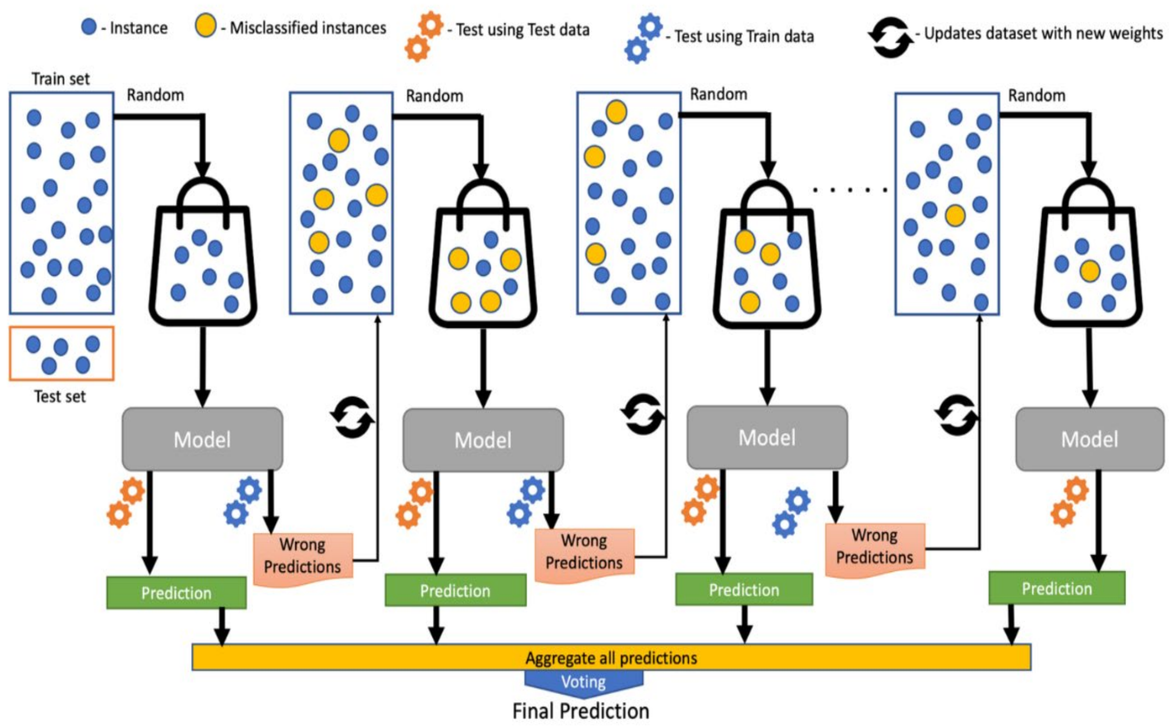
XGBoost is an ensemble learning and a gradient boosting algorithm for decision trees that uses a second-order approximation of the scoring function. This approximation allows XGBoost to calculate the optimal "if" condition and its impact on performance.

MAE:50.05

- MSE:13875.25

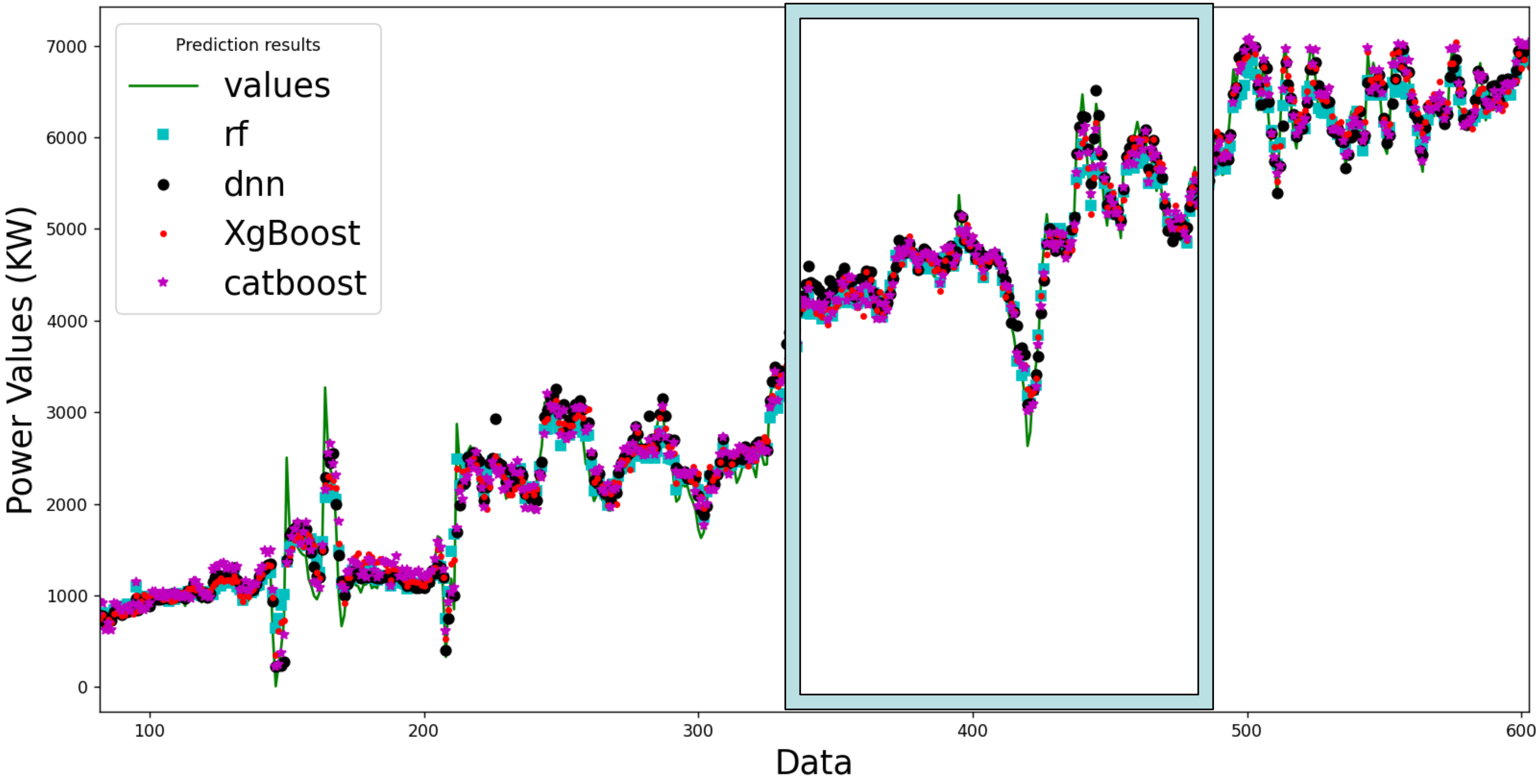
RMSE:117.79

- R-Square= 0.997198659343749



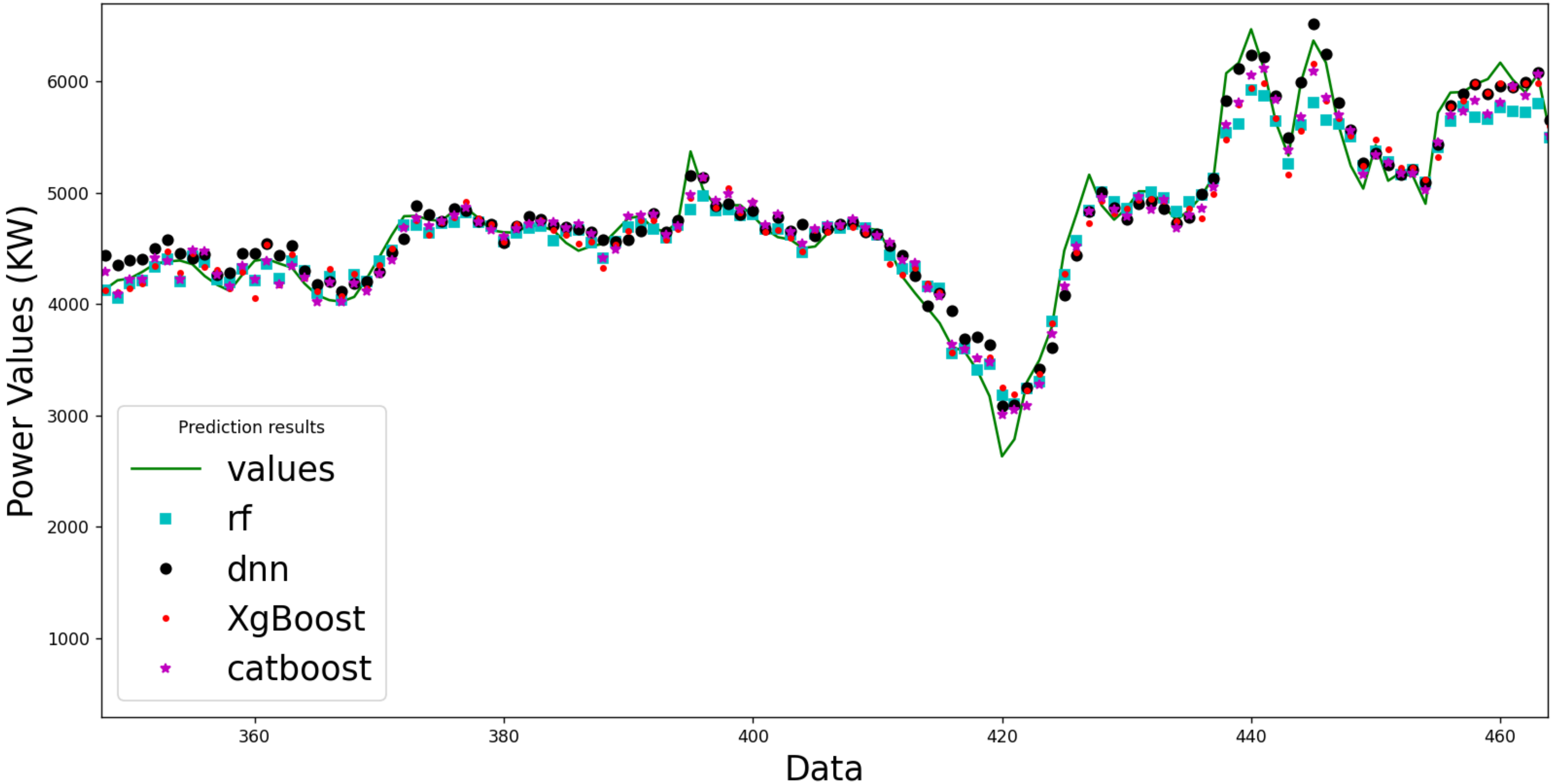
Shaping Energy for a Sustainable Future

Machine Learning Models: Results



Shaping Energy for a Sustainable Future

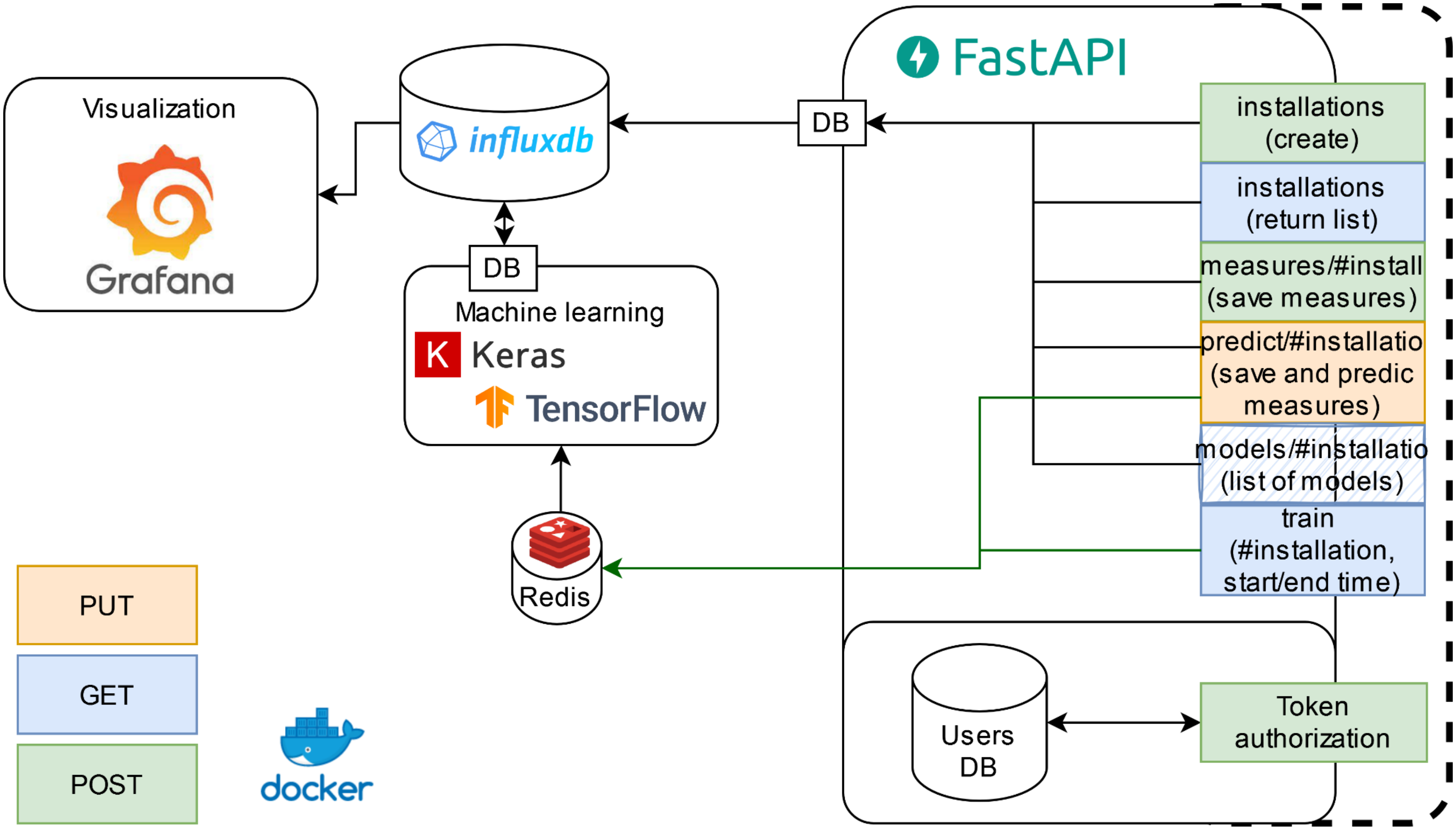
Machine Learning Models: Results



Comparative evaluations spotlight that while Random Forest excels in term of Error results, Deep Neural Networks (DNN) demonstrate superior fitting under extreme conditions and exhibit commendable speed.

IREC
Shaping Energy for a Sustainable Future

Complete Digital Twin Concept



Digital Twin Dashboard

IREC
Shaping Energy for a Sustainable Future

Conclusion

- Leveraging feature engineering has enhanced our comprehension of feature significance and informed the design of our digital twin concept.
- Comparative analysis of machine learning models not only gauges their accuracy but also elucidates behavioral nuances.
- The introduced digital twin concept boasts speed, agility, and reliability, positioning it as a dependable resource.
- Nevertheless, there is a necessity for updates, particularly in visualizations and the integration of enhanced control mechanisms, for further elevate the usability of the digital twin concept.

IREC
Shaping Energy for a Sustainable Future

Thanks for your attention

e-mail: tyalcin@irec.cat

Amb el finançament de:

European Union
European Regional Development Fund