PyDTS: A Toolkit for Deep Time-Series Prediction

DR. PASCAL SCHIRMER

AIMS & GOAL

The heart of machine learning and AI is signal processing and the language of its modelling is mathematics.

IEEE ICASSP 2023

- Structure of the talk
  - Presentation: 30 min
  - Questions: 20 min
  - Toolkit: 10 min
- This talk will be about
  - Providing a cooking receipt for time-series modeling -> Architecture
  - How to systematically improve the architecture -> Methodology
  - How to interpret the results
- This talk will not be about
  - The fanciest solution for time-series modeling -> Generalised solution
  - Generative models and transformers

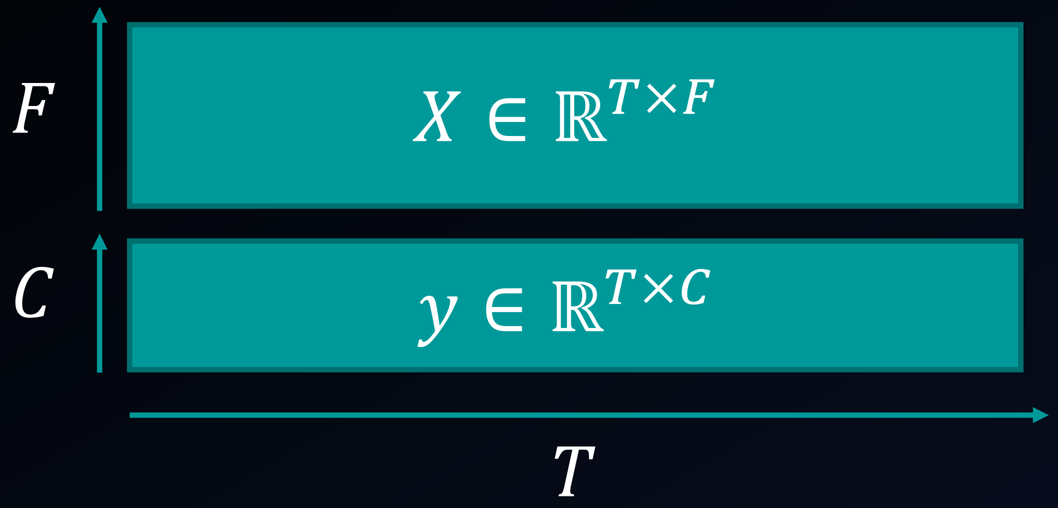
Content

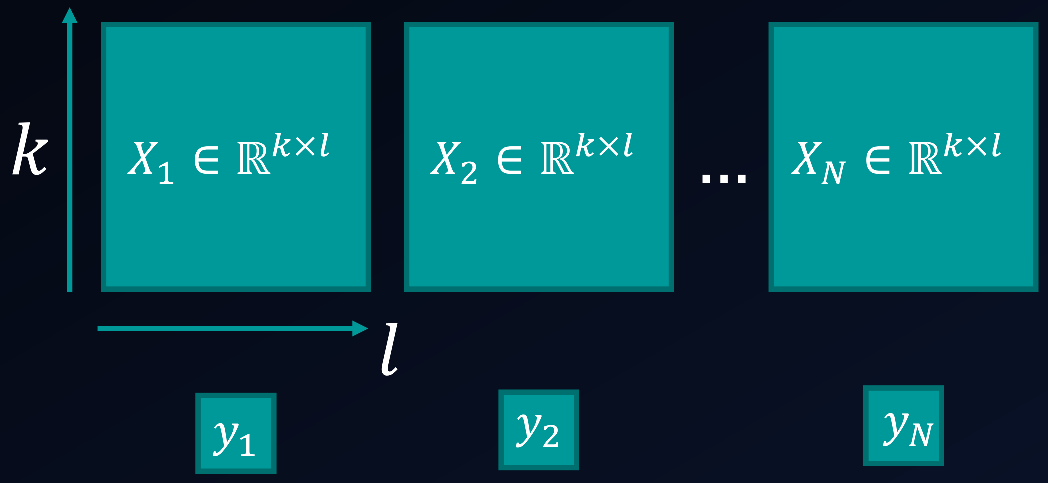
1) Short Introduction
2) PyDTS an Overview
3) Exemplary Workflow
4) Advanced Functionalities
5) Outlook and Conclusion

A

SHORT INTRODUCTION

Modeling Domain (Relation to the physical problem)





TEMPORAL DOMAIN

- Natural language processing
- Time series forecasting
- Non-linear regression

SPATIAL DOMAIN

- Image recognition
- Image to image learning
- Spatial transformers

Modeling Approaches (Relation to the solution)

Property	Statistics	Linear Algebra	Machine Learning
Runtime	+	o	-
Memory	+	o	-
Interpretation	o	+	-
Dimension	-	o	+
Transfer	-	o	+
Non-linear	-	o	+
Parameter	+	o	-
Data	o	+	-

[2]-[4]

DESCRIPTION

- Statistical Modeling
  - Autoregressions
  - Moving average processes
- Linear Algebra Modeling
  - State space systems and discrete PDE
  - Matrix and Tensor factorisations
- Machine/Deep Learning
  - Generic regression
  - Generic classification

Modeling Improvement (Relation to the performance)

ARCHITECTURAL LEVEL

- Feedback
- Pre-processing
- Post-processing

FEATURE LEVEL

- Time-Domain
- Frequency-Domain
- Statistical

LAYER OR SOLVER LEVEL

- Loss functions
- Learning rate updates
- Convergence criteria
- New layers

Model Applications (Areas where we should use ML)

ILL-DEFINED PROBLEMS

Physical model and boundary conditions are known, but system equations are ill-defined

- Inverse Problems
- Blind Sources
- Separations

MODEL OR BOUNDARY FREE PROBLEMS

Physical model or its boundary condition are unknown or cannot be predicted.

- Medical monitoring
- Anomaly detection
- Transfer learning

HUMAN LIKE PROBLEMS

Input and output is generated by humans, thus neither a model or boundary conditions exist.

- Text
- Audio
- Images

B

PYDTS AN OVERVIEW [5]

Considered Problems

PROBLEM DESCRIPTION

- Denoising
- Forecasting
- Non-linear Modeling
- Anomaly Detection
- Degradation Modeling

MATHEMATICAL FORMULATION

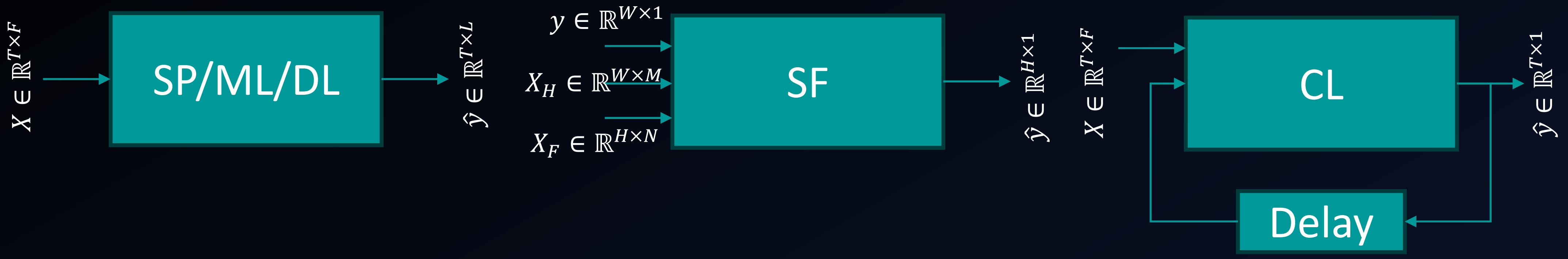
- $y(t) = x(t) + \epsilon$
- $y(t) = \alpha y(t-1) + \beta x(t) + \epsilon$
- $y(t) = f_\theta\big(x(t)\big)$
- $y(t) = \varphi\left(f_\theta\big(x(t)\big)\right)$
- $y(t) = y_0 + \beta x(t) + \epsilon$

Implemented high-level architecture

MATHEMATICAL FORMULATION

$$\left.\begin{array}{ll} \text{• True Function:} & y(t) = f_{\theta}\big(x(t)\big) \\ \text{• Approx. Function:} & \hat{y}(t) = g\big(x(t)\big) \text{ s.t. } \min\|y - \hat{y}\|_2^2 \end{array}\right\} g \approx f_{\theta}$$

Model Configurations



OPEN LOOP

- Input (X)
  - T samples
  - F features
- Output (y)
  - T samples
  - L channels

SHORT FORECAST

- Input (y, X_H, X_F)
  - W previous samples
  - W exogenous samples
  - H horizon samples
- Output (y)
  - H horizon samples

CLOSED LOOP

- Input (X, y)
  - T samples
  - F features
  - Delayed Output
- Output
  - T samples with 1 channel

Operating Modes

CONVENTIONAL REGRESSION

- Constant Split
- K-Fold Validation
- Transfer Learning

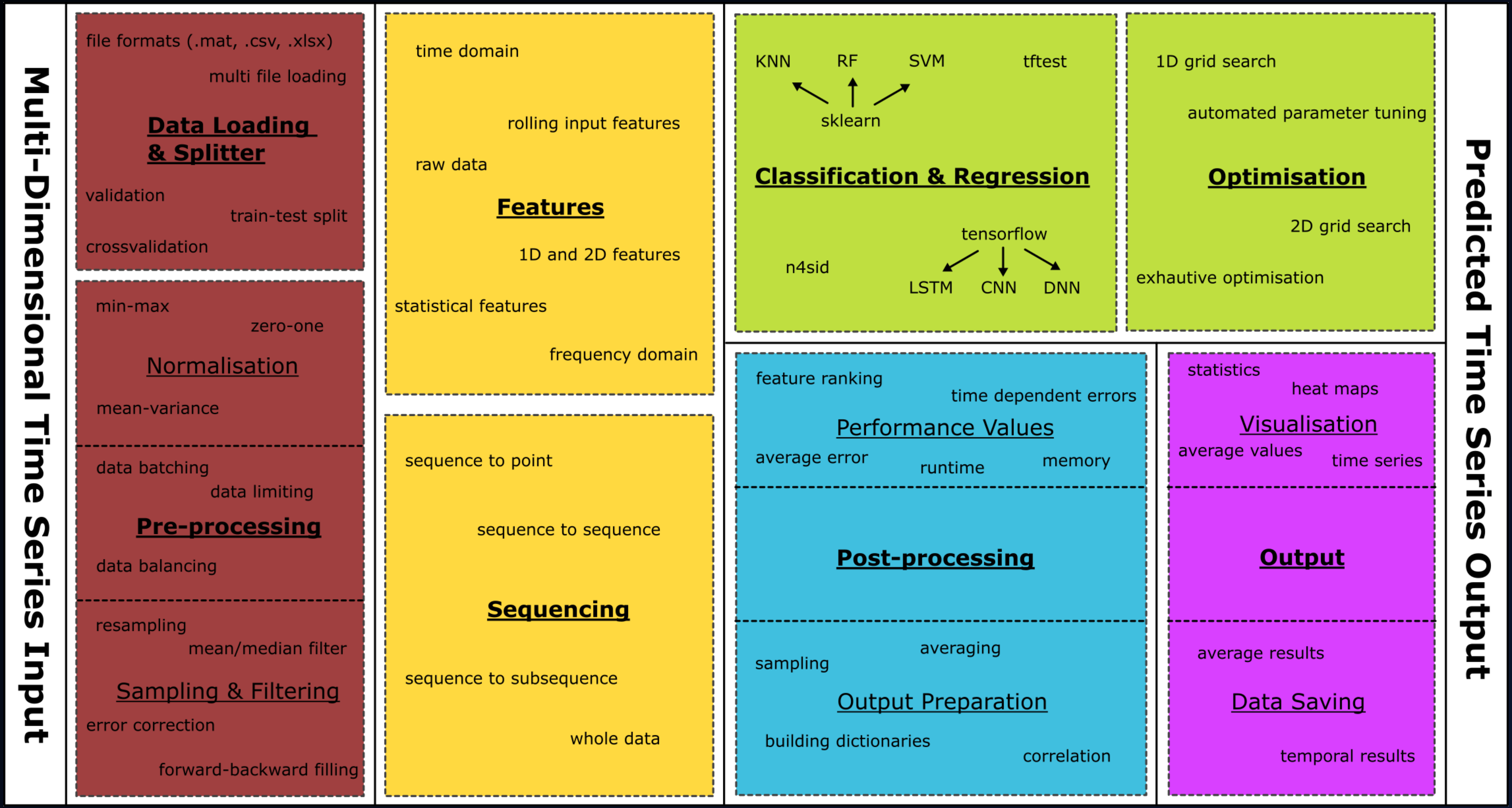
EXHAUSTIVE GRID SEARCH

- Window Length
- Sampling
- Filtering

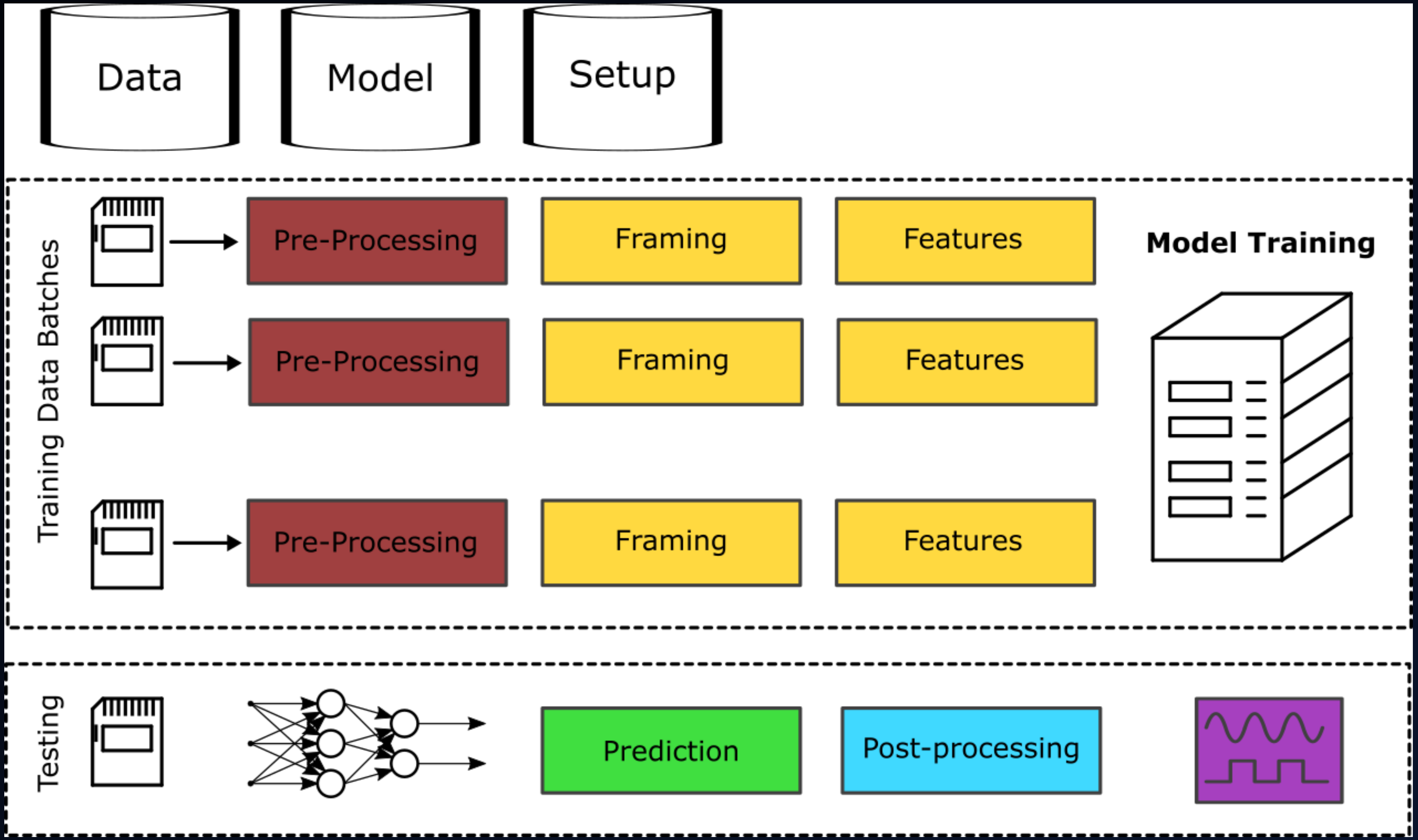
HYPERPARAMETER OPTIMISATION

- Number Layers
- Number Neurons
- Activation Fnc.

Included Modules



Data Pipeline



C

EXEMPLARY WORKFLOW

Problem Description

DESCRIPTION

Based on the observation of previous three years of regenerative energy generation the aim is to predict the generation of the following year.

- Inputs
  - Average Temperature (°C)
  - Average Solar Irradiance (W/m^2)
  - Average Wind speed (m/sec)
- Outputs
  - Regenerative solar generation (MWh)
  - Regenerative wind generation (MWh)

Initial Parameters

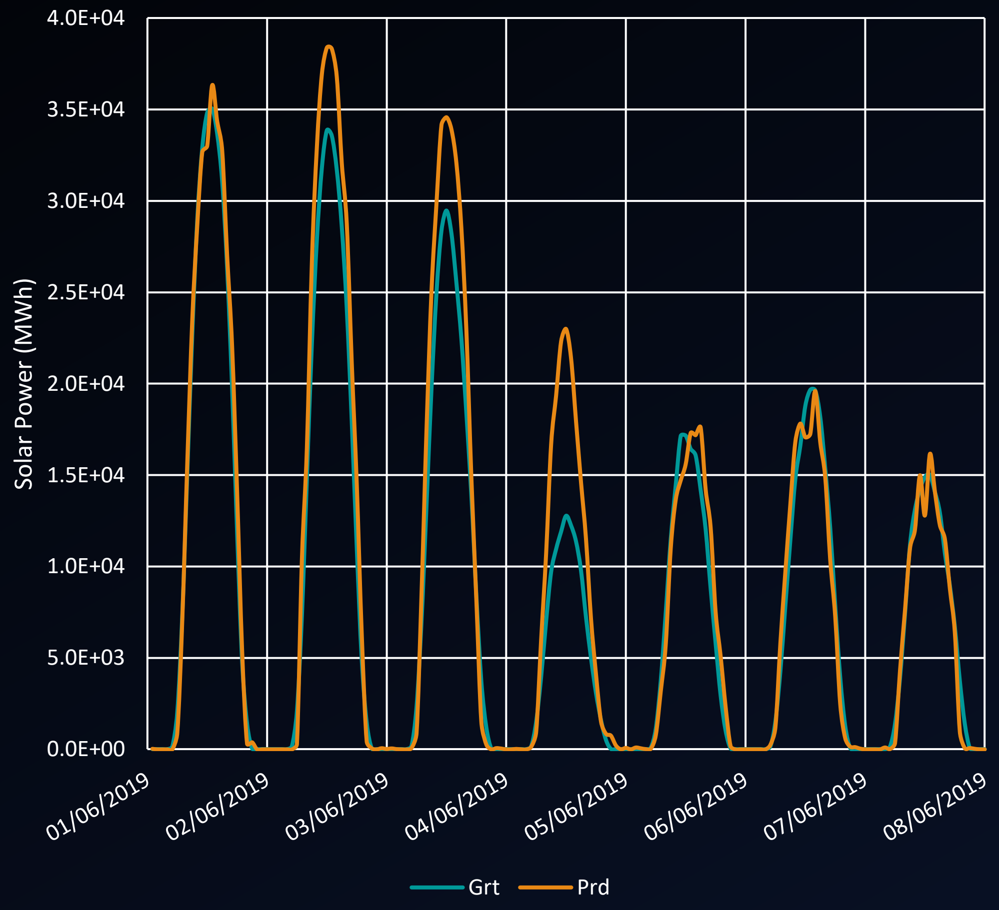
INPUT PARAMETERS AND CONFIGURATION

- Sampling rate: 60 min
- Window length: 24 samples (24 hrs)
- Overlap: 23 samples
- Method: Seq2Point
- Focus sample: Last (causal)
- Normalisation X: mean-std
- Normalisation y: 0-1

MODEL AND HYPERPARAMETERS

- Model: DNN (3-layers, 32-nodes)
- Solver: Adam
- Epochs: 100 (15 patience)
- Loss: MAE
- Batch: 1000
- Learning rate: 1e-3
- Beta: 0.9 / 0.999

Initial Results (Solar)



DESCRIPTION

- General good correlation between true and predicted values
- Two types of errors
  - Average error (mismatch in the predicted energy)
  - Shape error (mismatch in the shape of the predicted energy)

Optimisation by "Thinking"

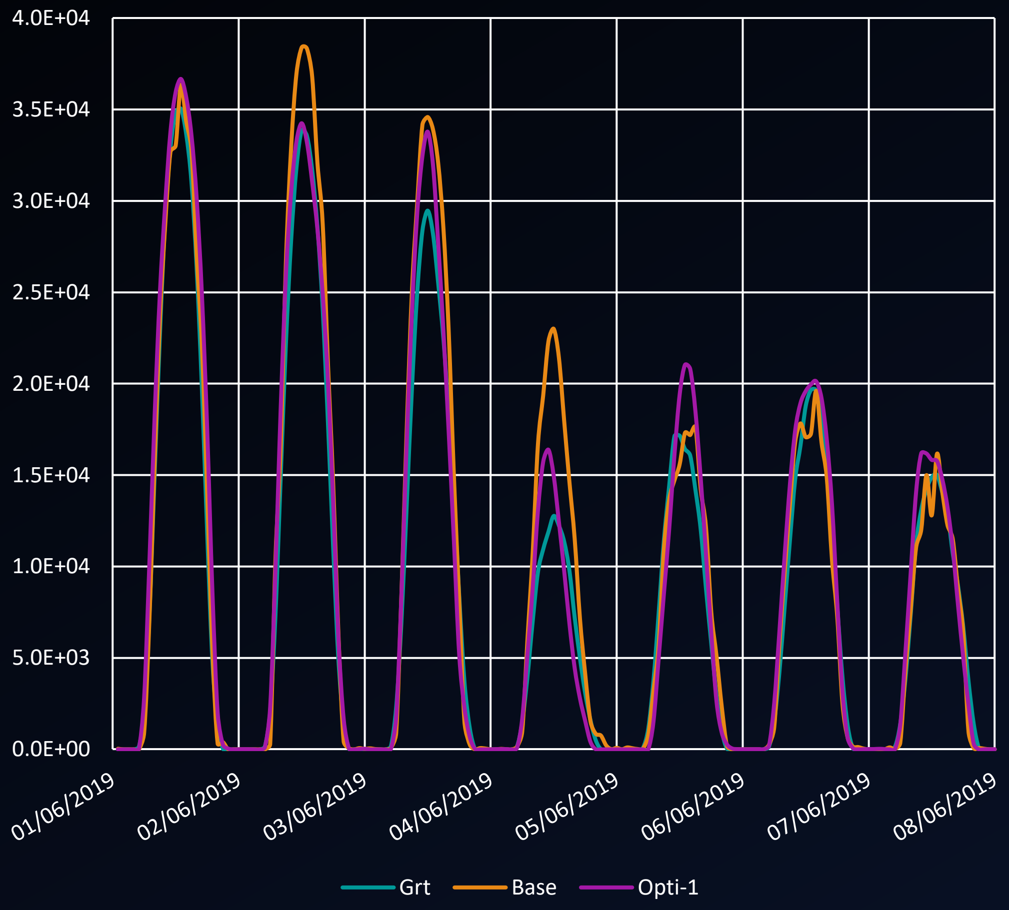
WINDOW LENGTH

- Observation: error shows daily and yearly temporal influence
- Reason: solar irradiance has a short term cycle (1 day) and a long term cycle (1 year)
- Action: input window length (12 hrs = 12 samples)

MODEL ARCHITECTURE

- Observation: error distribution is not mean centred
- Reason: additional PV capacity is added, which is cannot be captured by the DNN
- Action: adding two LSTM layers to capture long term changes

Optimisation by “Thinking” (Opti-1)



DESCRIPTION

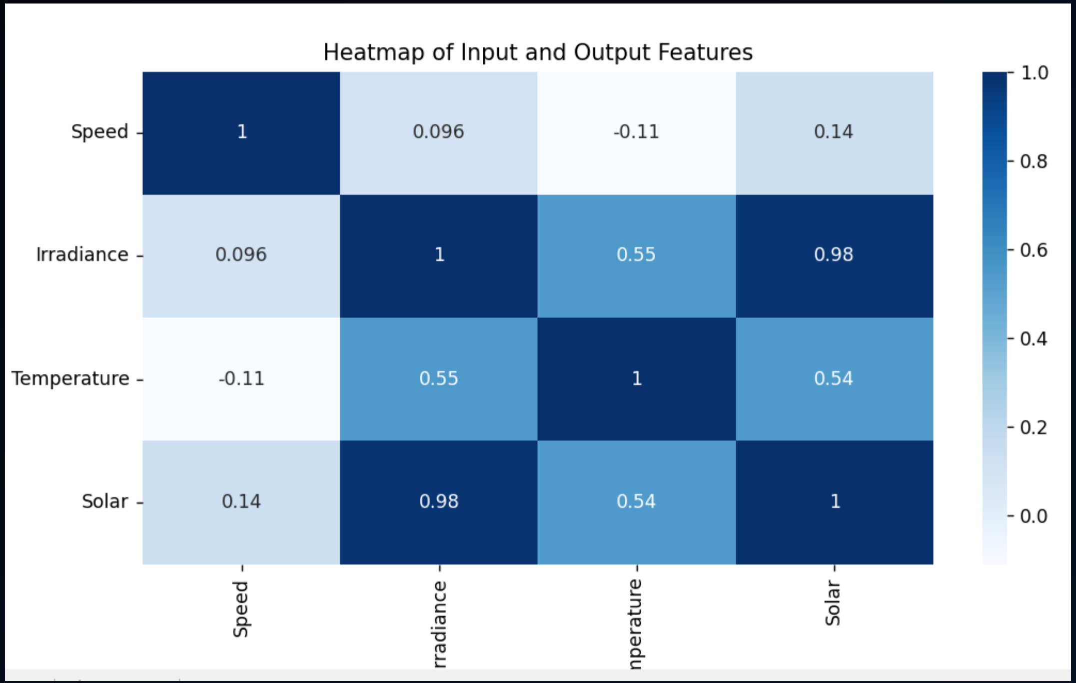
- Significant improvement over baseline system
  - Average error reduced
  - Shape is modelled better
- Only for the 5th of June the performance decreased

D

ADVANCED FUNCTIONALITIES

Choosing Input Features

FEATURE CORRELATION



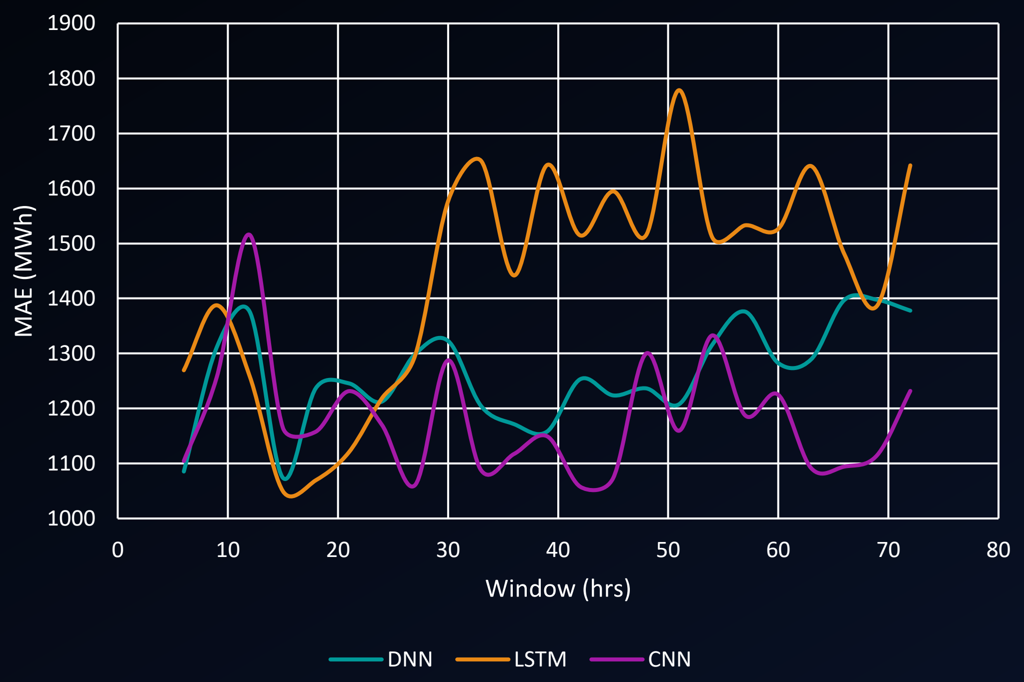
DESCRIPTION

Feature ranking methods rank the importance of a input with respect to the output

- Linear techniques: PCA, Cor
- Non-linear techniques: RF
- Leave-One-Out: ReliefF [1]

Optimising Input Parameters (Opti 2)

OPTIMISATION SPACE



DESCRIPTION

- Input window length is varied
  - Minimum: 3 hrs (3 samples)
  - Maximum: 72 hrs (72 samples)
  - Step: 3 hrs (3 samples)
- Best solution
  - 15 hrs (15 samples)
  - Shortest window capturing a full day of sunshine during summer

Optimising Model Architecture (Opti 3)

OPTIMISED ARCHITECTURE

LSTM(15, 48)

LSTM(15, 64)

LSTM(112)

DNN(96)

DNN(1)

DESCRIPTION

- Number of LSTM layers (1 – 3)
- Number of DNN layers (1 – 4)
- Number of Neurons (16 – 256)
- Dropout (0.0 – 0.5)
- Learning rate (1e-5 – 1e-2)
- 16-1024 Batch Size

Optimal Results (Solar): Average

IMPROVEMENTS

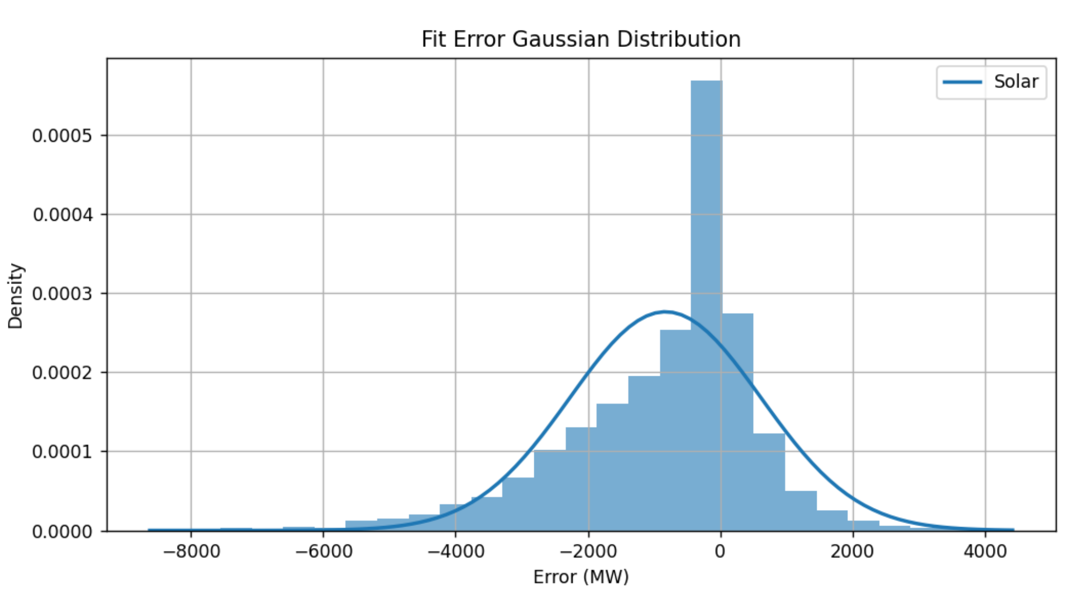
	Base	Opti 1	Opti 2	Opti 3
MAE	1212.7	912.4	718.1	640.9
RMSE	2142.5	1670.8	1339.3	1252.9
SAE	11.54%	8.22%	8.30%	8.50%
TECA	89.07%	91.78%	93.53%	94.22%
R^2	93.58%	96.09%	97.49%	97.80%

DESCRIPTION

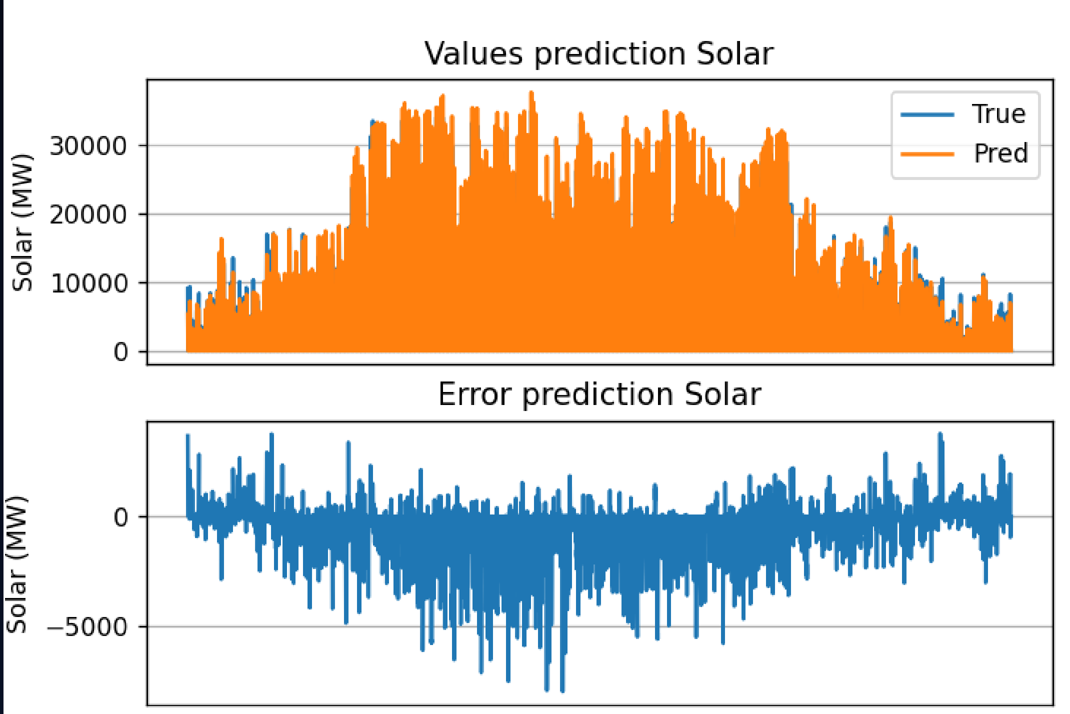
- Strategic optimisation leads to almost 50% reduction of error for MAE and RMSE
- SAE cannot be improved due to installation of new capacity
- Signal correlations for TECA and R^2 are increasing by 4-5%

Optimal Results (Solar): Time Domain

ERROR DISTRIBUTION



TIME DOMAIN ERROR



E

OUTLOOK AND CONCLUSION

Outlook and Conclusion

WHAT IT CAN DO

- Serve as a baseline system for time-series problems
- Help to understand correlations and select features
- Optimise existing architectures and parameters

WHAT IT CANNOT DO

- Does not produce cutting edge results
- Does not release you from thinking about your problem
- Does not support yet generative models and transformers

References

[1] Robnik-Šikonja, M.; Kononenko, I. Theoretical and empirical analysis of ReliefF and RReliefF. Mach. Learn. 2003, 53, 23–69.

[2] Zheng, H.S.; Liu, Y.Y.; Hsu, C.F.; Yeh, T.T. StreamNet: Memory-Efficient Streaming Tiny Deep Learning Inference on the Microcontroller. In Proceedings of the Thirty-Seventh Conference on Neural Information Processing Systems, 2023.

[3] Schirmer, P.A.; Mporas, I. Device and Time Invariant Features for Transferable Non-Intrusive Load Monitoring. IEEE Open Access J. Power Energy 2022, 9, 121–130.

[4] Chen, X.W.; Lin, X. Big Data Deep Learning: Challenges and Perspectives. IEEE Access 2014, 2, 514–525.

[5] Schirmer, P.A.; Mporas, I. PyDTS: A Python Toolkit for Deep Learning Time Series Modelling. Entropy 2024, 26, 311. https://doi.org/10.3390/e26040311