On Short-Term Load Forecasting Using Machine Learning Techniques

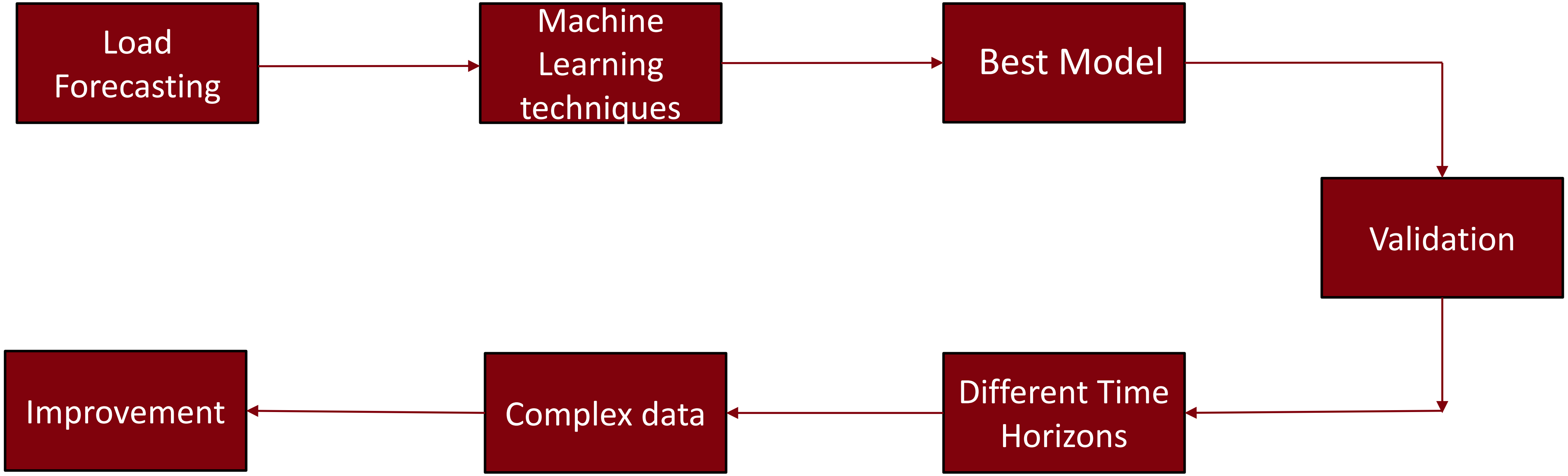
By: Behnam Farsi

Supervisors: Dr. Bouguila, Dr. Eicker, Dr. Amayri

December, 2020

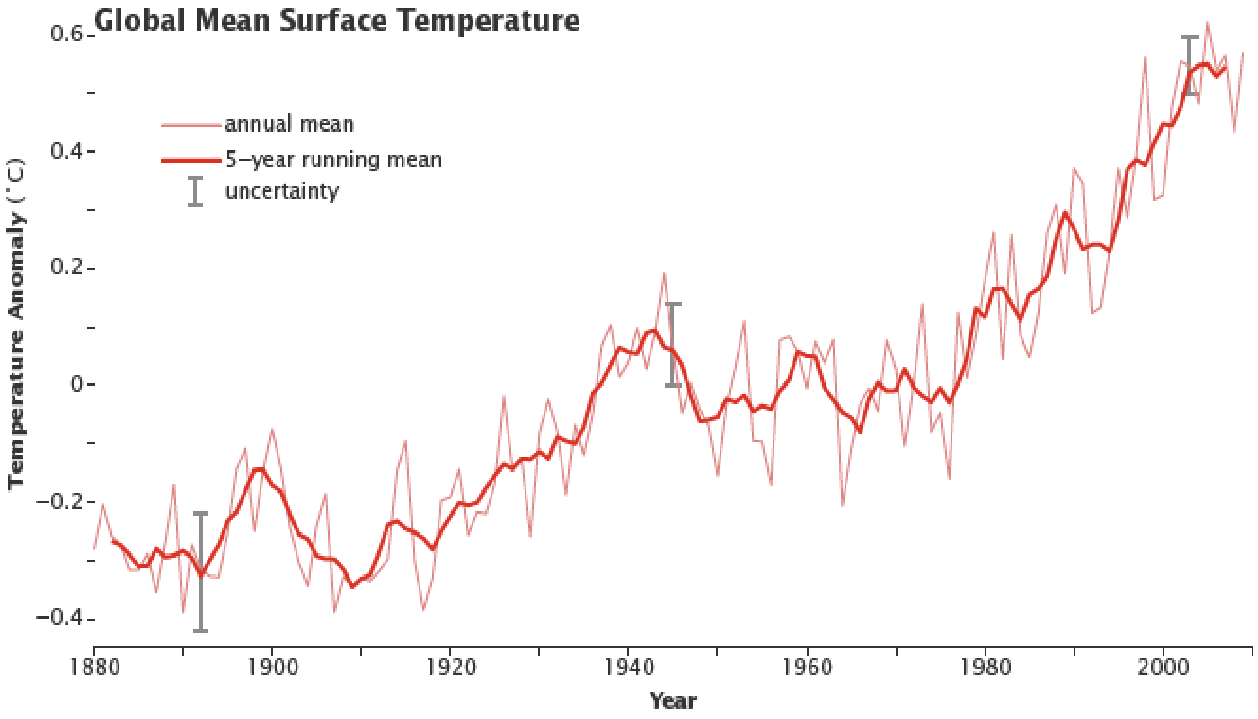
Contents

Workflow


Load Forecasting
Machine Learning techniques
Best Model
Validation
Different Time Horizons
Complex data
Improvement

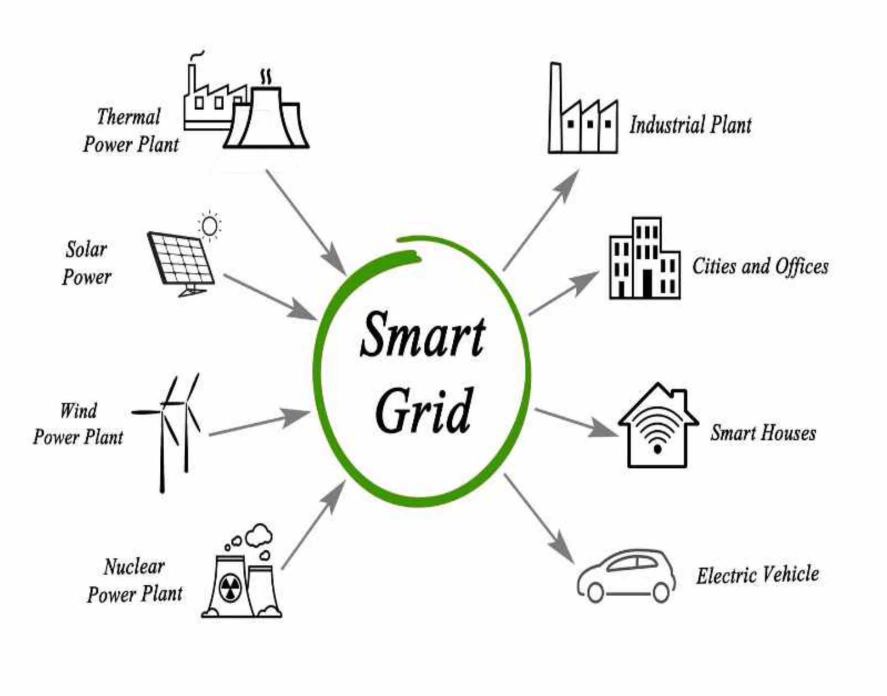
Global Warming

https://earthobservatory.nasa.gov



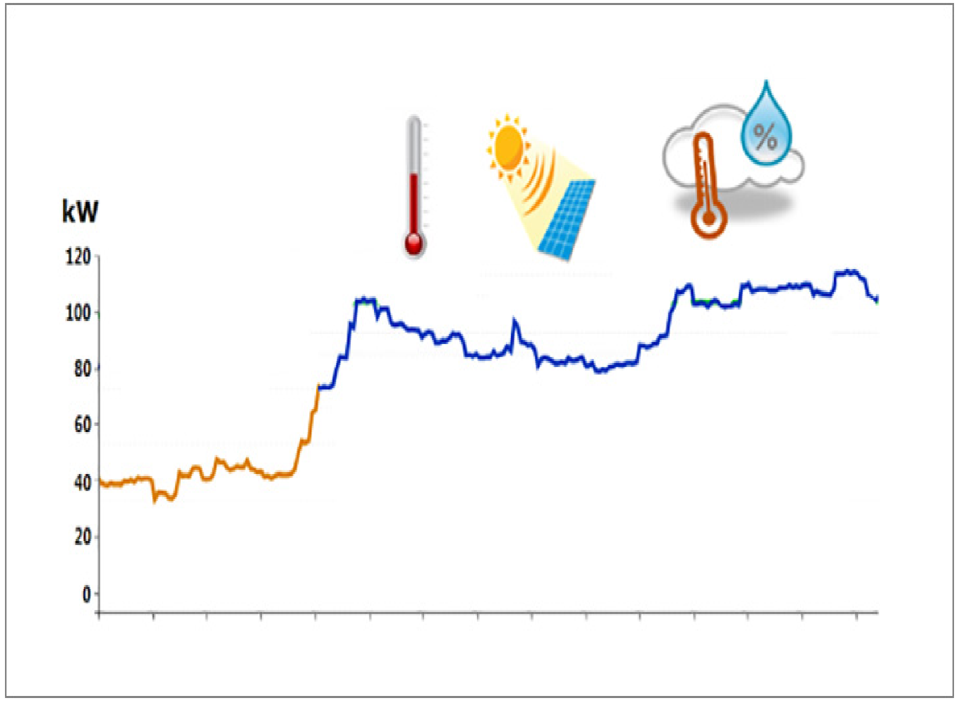
Global Warming

- Lifestyle and behavior
- Dietary change (low-carbon diet)
- Low-carbon energy (solar, wind, etc.)
- Carbon dioxide removal
- **Smart grid and load management**



https://www.smart-energy.com/

Load Forecasting



What is Load Forecasting?

- Electricity consumption
- Historical Data
- Electrical instruments
- Human Behavior
- Other sciences

Types of Load Forecasting

Very Short-Term

Short-Term

Medium-Term

Long-Term

Factor Affecting Short-Term

- Time
- Economy
- Weather
- Humidity
- Holidays and weekends

Factor Affecting Medium and Long Term

- Description of appliances
- Technology changes
- Population
- Economic factors
- Age of equipment

Time Series

Time Series data depend on the time. During time, they have different behavior.

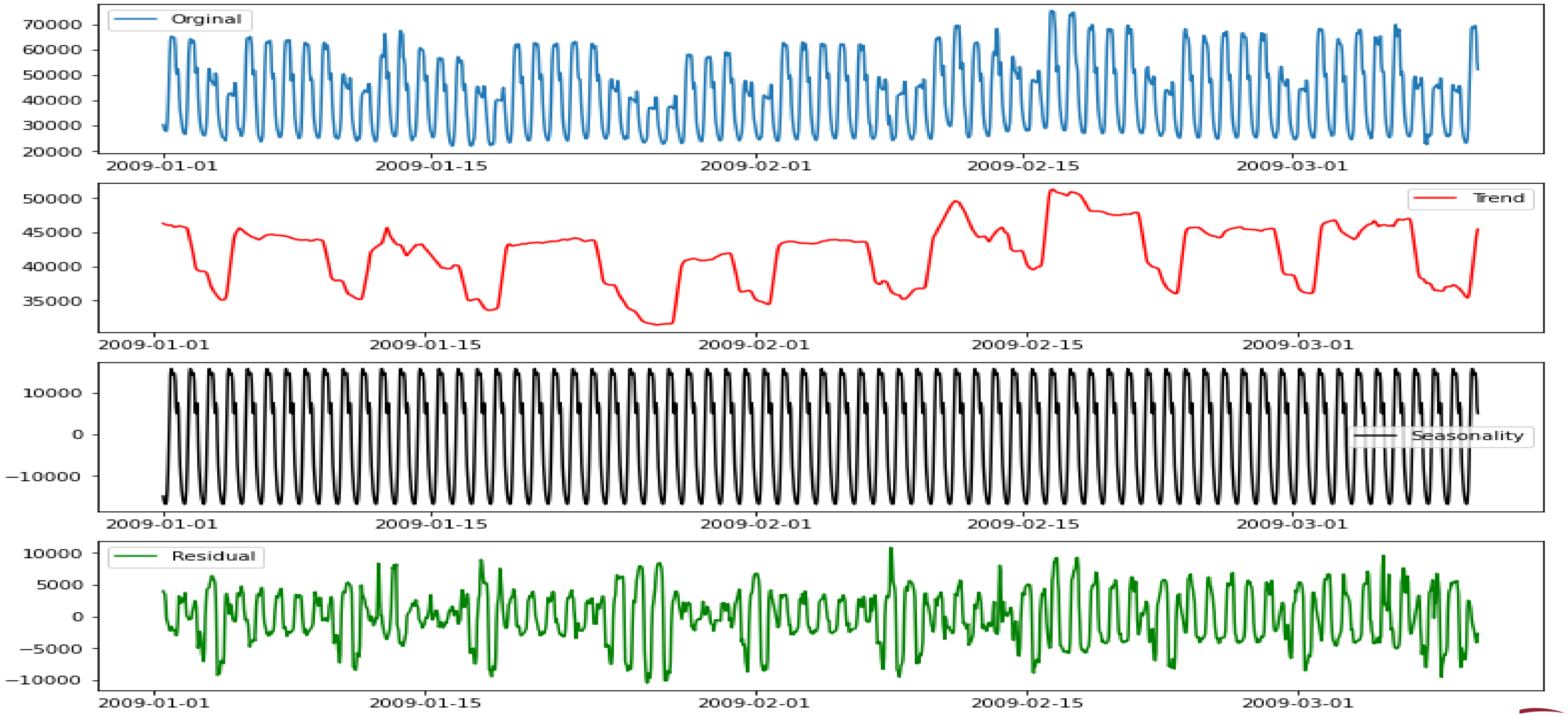
3 Specific attributes:

1. **Trend**: There is an increase or decrease during the whole time.

2. **Seasonality**: Many of time dependent data in a certain time have a seasonal algorithm.

3. **Noise** (residuals): By subtracting both trend and seasonality from original data, residuals (Noise) will remain.

Time Series



Orginal
70000
60000
50000
40000
30000
20000
2009-01-01
2009-01-15
2009-02-01
2009-02-15
2009-03-01
Trend
50000
45000
40000
35000
Seasonality
10000
0
−10000
Residual
10000
5000
0
−5000
−10000

Time Series

Stationary data: a stationary time series does not have pattern to predict the future by looking to it. Moving Average is almost constant.

Non-stationary

Moving Average: The average of specific time like every 10 minutes or one hour.

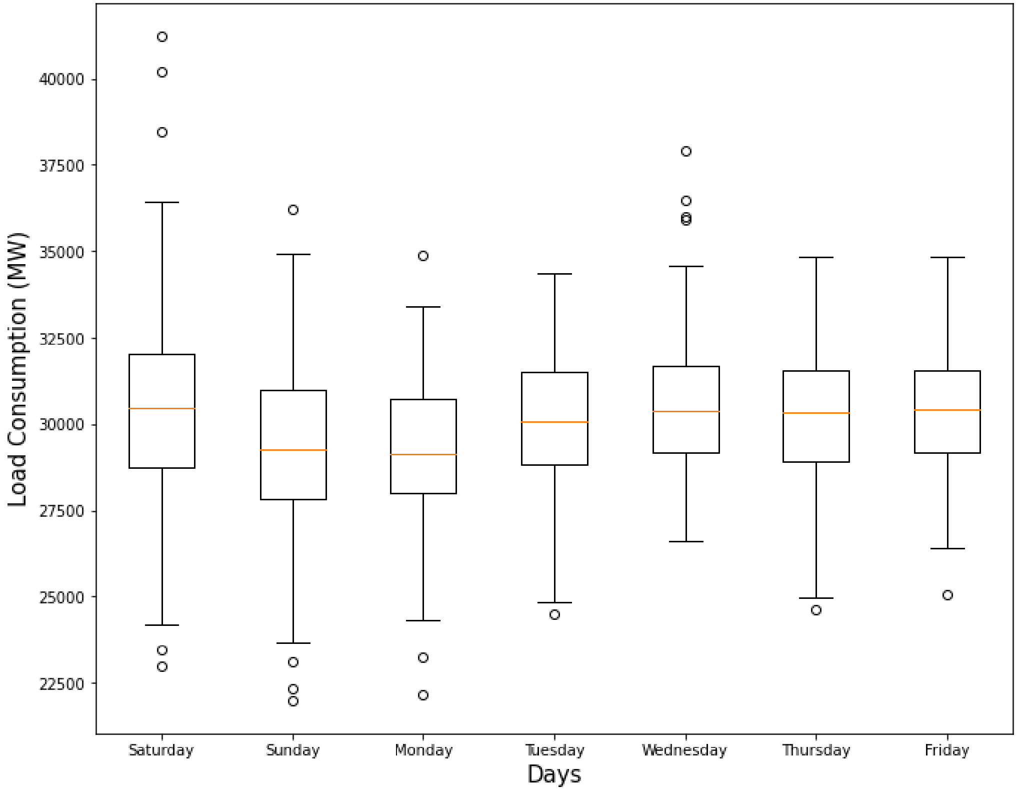
2 Different tests:

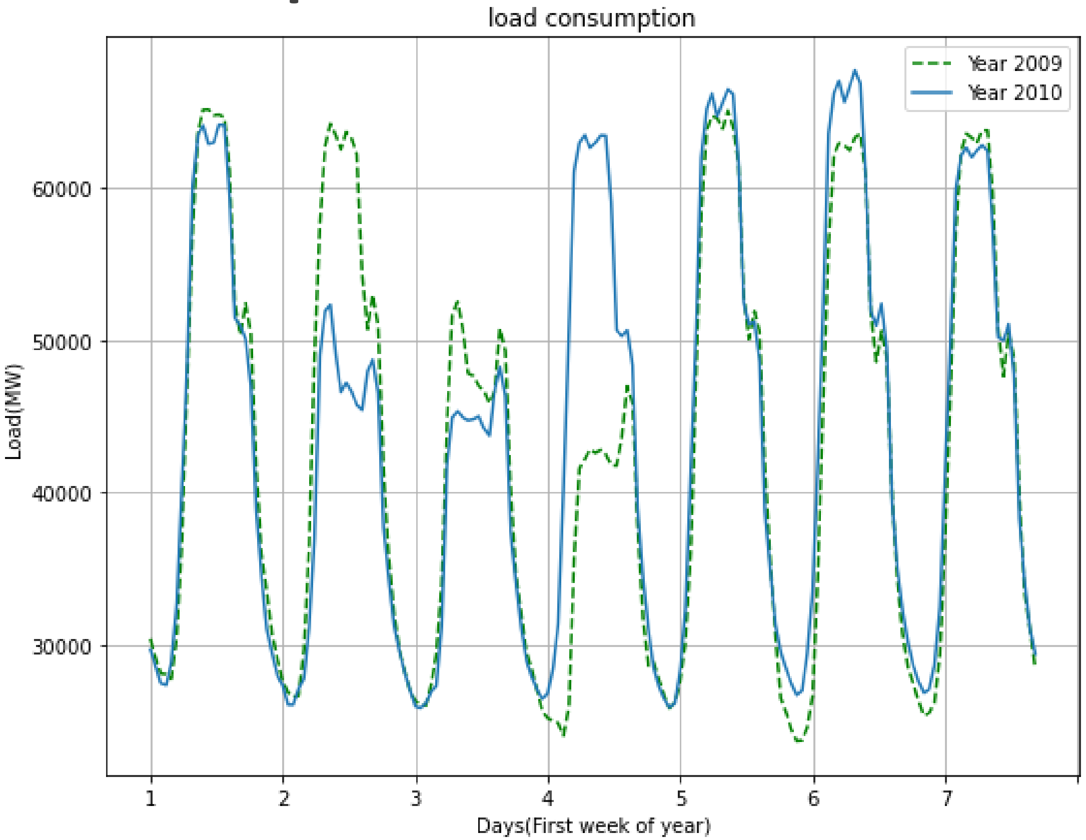
- Dickey Fuller Test (P-Value)

- Rolling Test (Plotting)

Data Sets

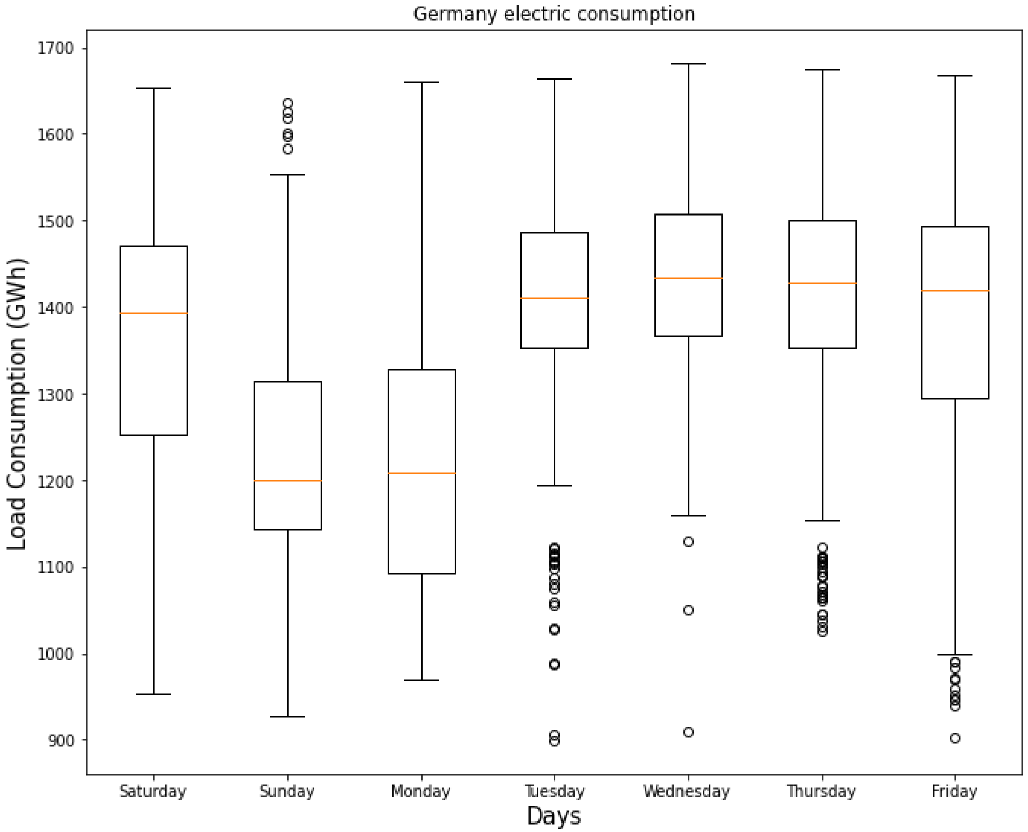
Malaysian Dataset → 17518 recorded sample.

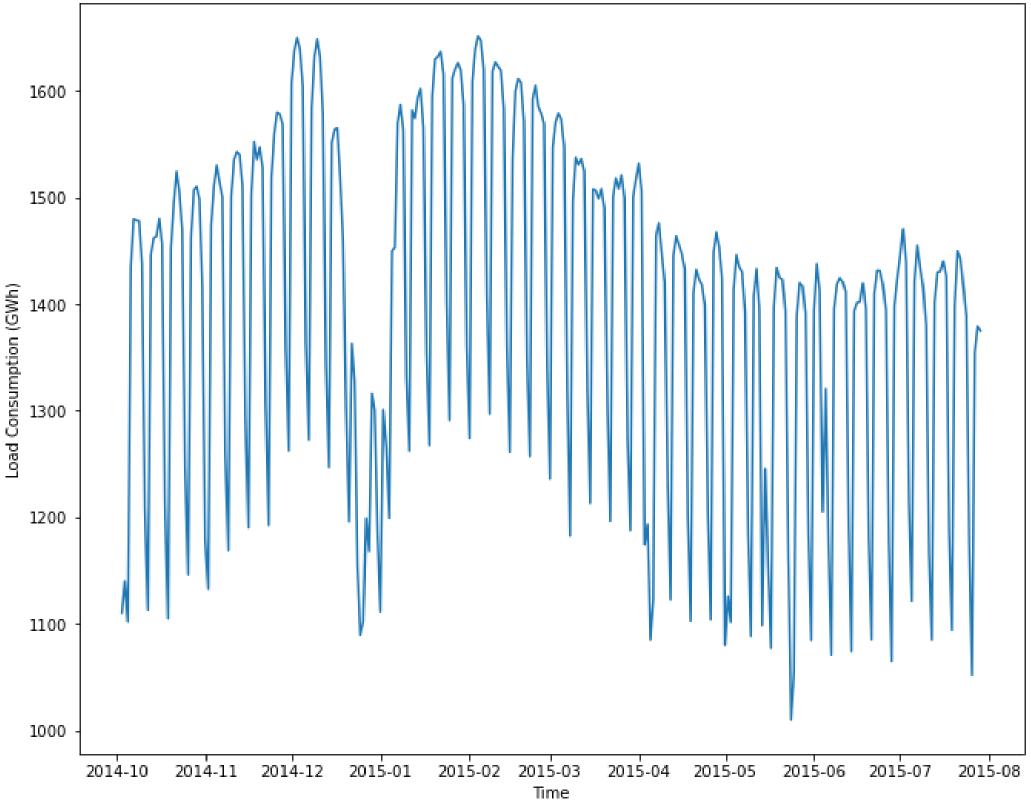




Data Sets

German Dataset → 2186 recorded sample.





Machine Learning Techniques

Statistical Models

- *ARIMA*
- *ETS*

Regression based Models

- *Linear Regression*
- *SVR*

Deep Learning Models

- *Fully Connected*
- *LSTM*
- *CNN*

ARIMA

ARIMA stands for Autoregressive(AR) Integrated (I) Moving Average (MA).

ARIMA (p, d, q) / SARIMA (p, d, q)(P, D, Q,[m])

p: Lag order

d: Degree of differencing

q: Order of Moving Average

AR

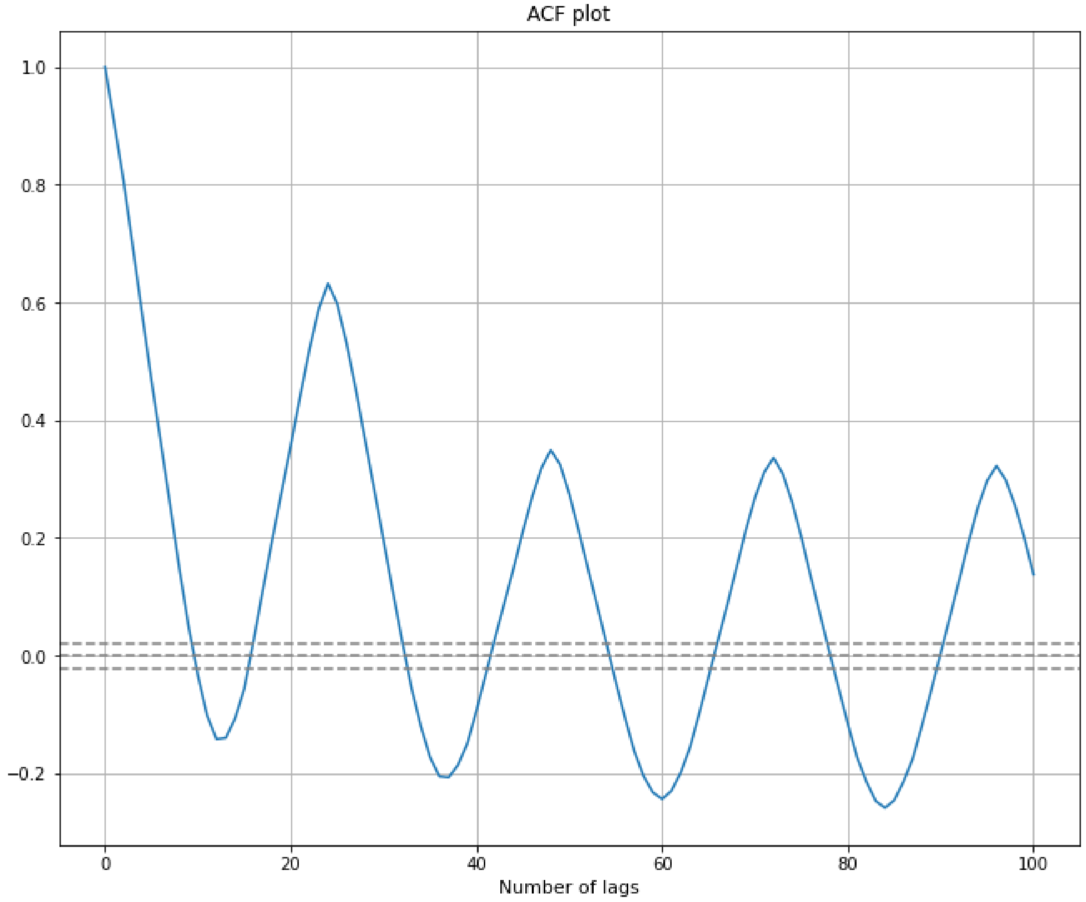
- $y_{\mathrm{t}} = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \mu$

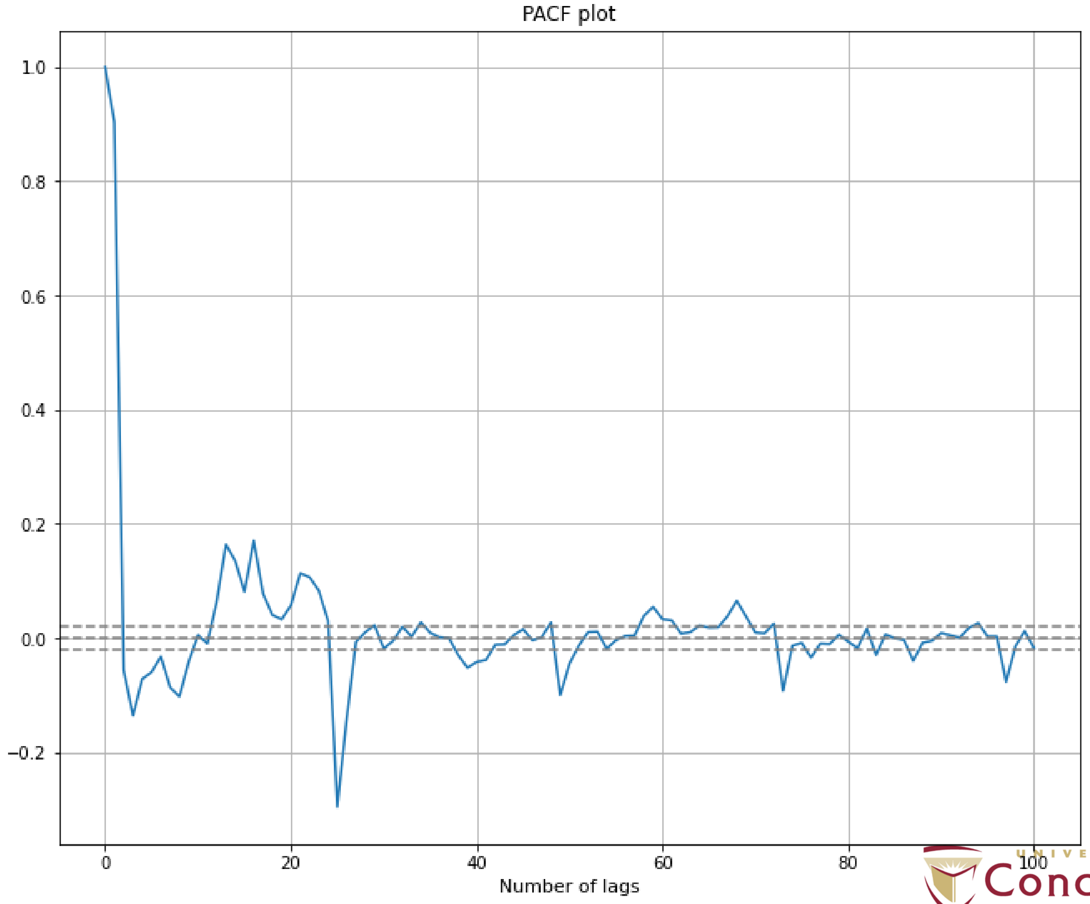
MA

- $X_{\mathrm{t}} = \theta_0 + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_p \epsilon_{t-q} + \epsilon$

ARIMA

Malaysian: SARIMA (1,0,1)(2,0,0,[24]) , German: SARIMA (5,1,0)(5,0,5,[7])





ETS

Based on previous observations.

An alternative for ARIMA.

$$F_{\mathrm{t+1}} = \alpha \mathrm{A}_t + (1 - \alpha)\, \mathrm{F}_t$$

$$0 < \alpha < 1$$

Linear Regression

Simple Linear Regression : $Y_i = \beta_0 + \beta_1 X_1 + \mu_i$

where **Y** is dependent variable, $\boldsymbol{\beta_0}$ interceptor, $\boldsymbol{\beta_1}$ is the slope, **X** is the independent variable and $\boldsymbol{\mu_i}$ is residual of the model.

If there were more independent variables, we would have **multiple linear regression (MLR).**

To work with this model, best number for coefficients should be found. In order to achieve that Least-squared error is used:

$$\text{LSE} = \Sigma\,(Y_i - X_i)^2$$

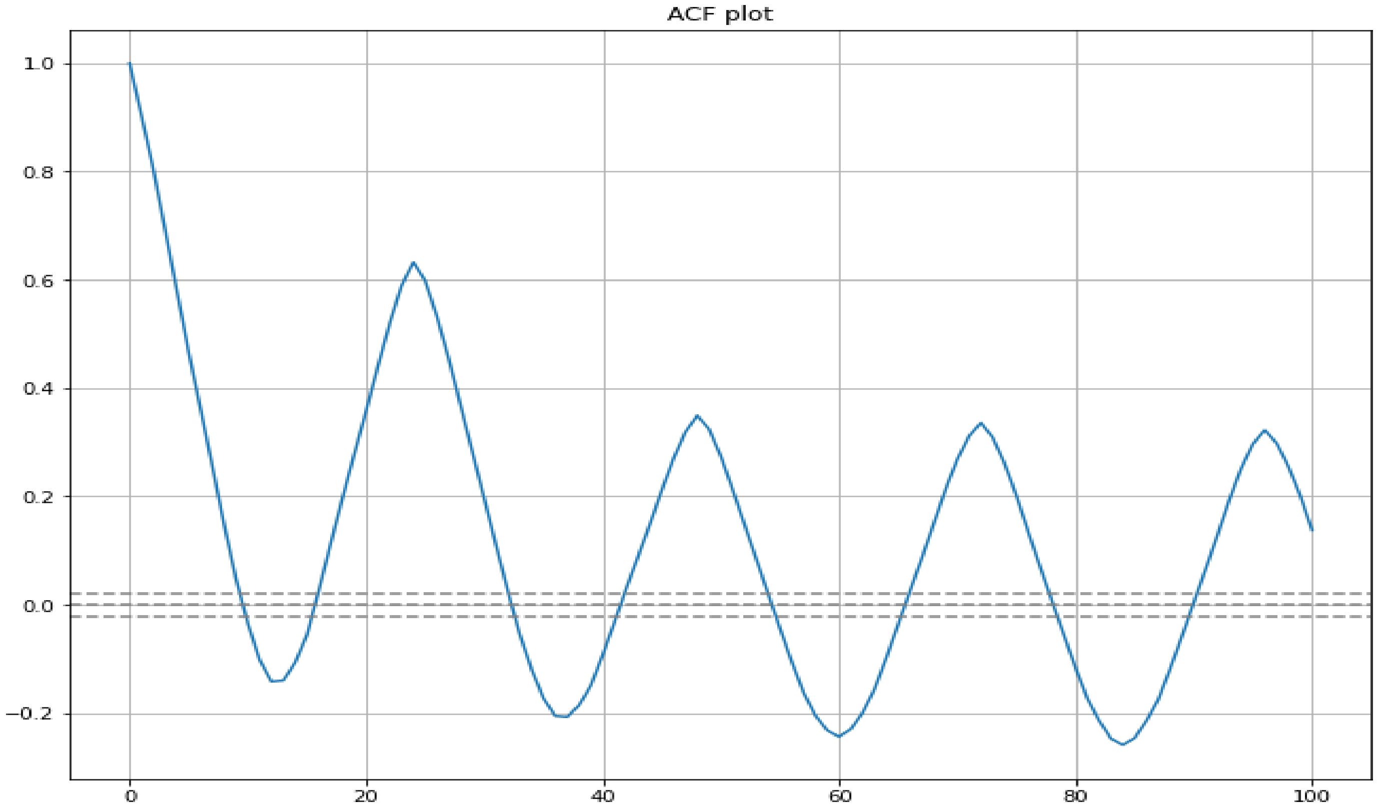
Where X is predicted value, Y is actual value and i is indicator of i-th variable.

Linear Regression

Using ACF plot.

Lags as independent variables.

Selecting threshold.



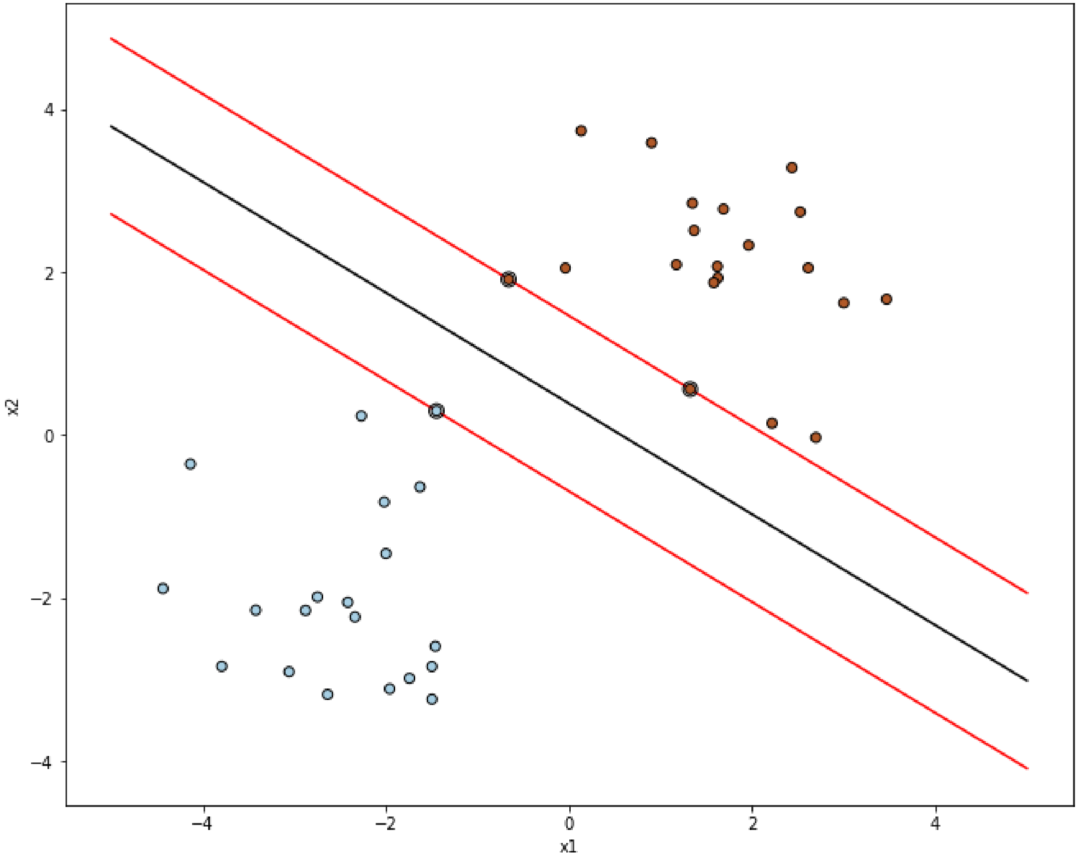
SVR

$$y_{\mathrm{i}} = wX_{\mathrm{i}} + b$$

Solution: min ½ $||w||^2$

Constraints : y_i - wX_{i} - b < ϵ

wX_{i} + b - y_i < ϵ



Deep Learning

Fully Connected Neural Networks:

Input layer, hidden layer, outputlayer.

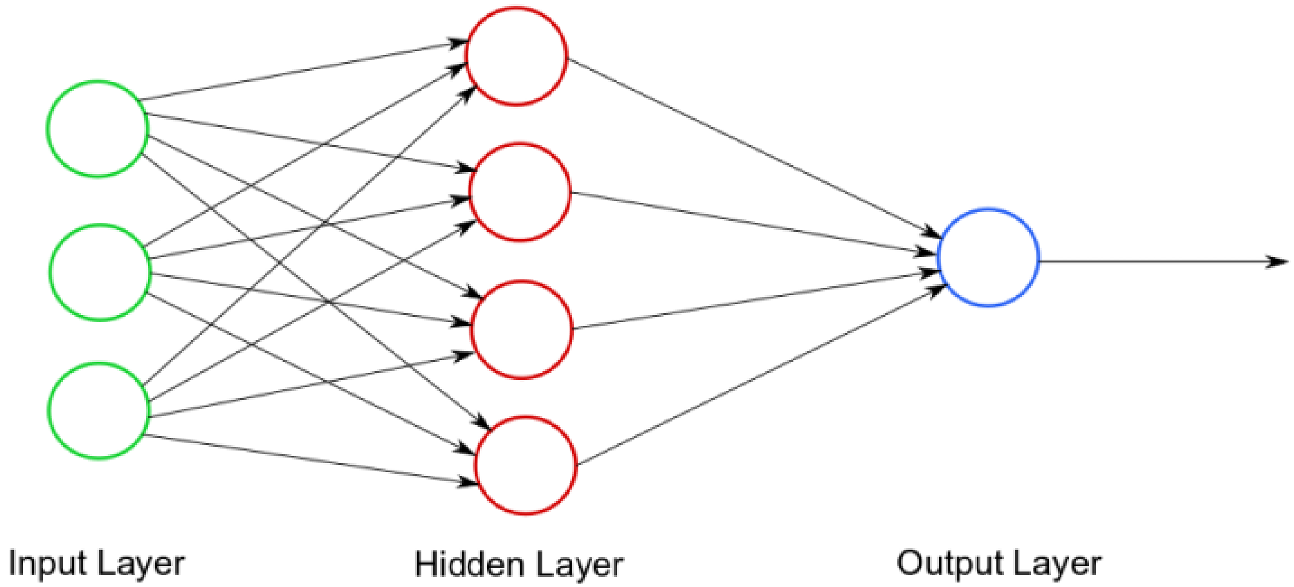
The complexity depends on number of hidden layer.

$a_l = W_l h_{l-1} + b_l$

$h_l = f(a_l)$ - Relu, Softmax, Sigmoid

Find the Weights with back-paropagation

$L = 1/N \Sigma (y-p(\alpha))^2$



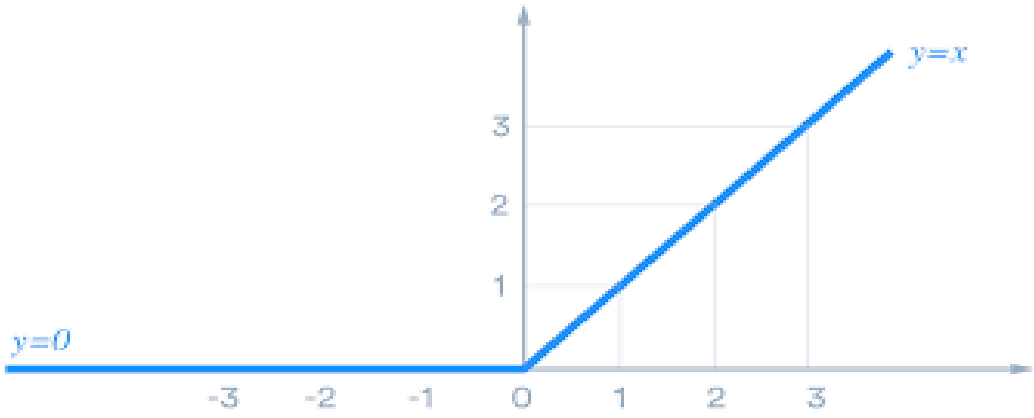
Deep Learning

Fully Connected Neural Networks:

ReLU:

ReLU stands for Rectified Linear Unit and it works like linear function with a difference which is output for negative inputs is zero. The mathematical formula is given as :

$$F(x) = \max(0, X)$$



Deep Learning model

Overfitting: The model is extremely trained.

Working based on ACF plot (Same as regression-based models).

Activation: ReLU

Deep Learning

CNN:

Convolutional Neural Networks

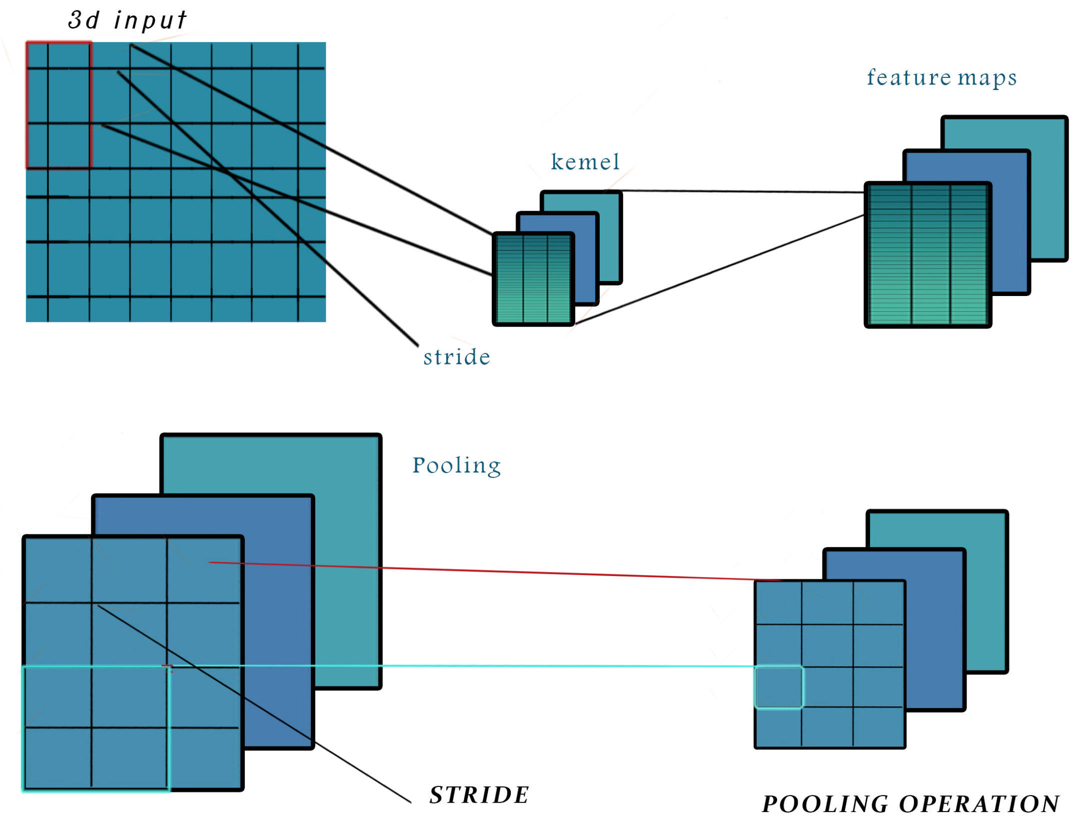
Based on Convolution opt.

Usable in 1-D, 2-D, 3-D

Common in NLP, image processing

Load forecasting and etc.

Feature extraction.



Deep Learning

LSTM:

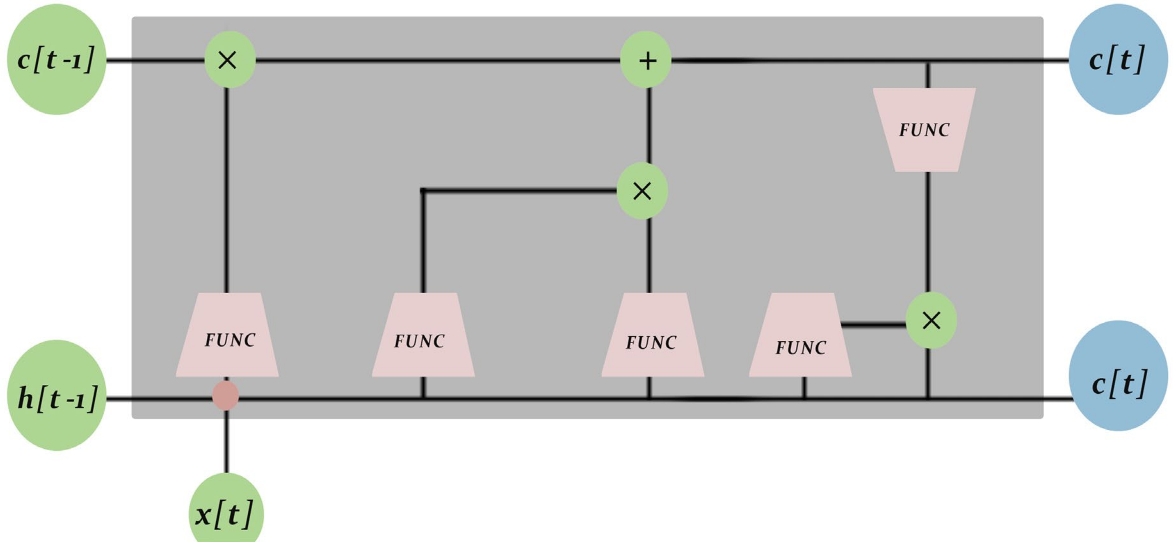
Long Short-Term Memories

Based on Control theory.

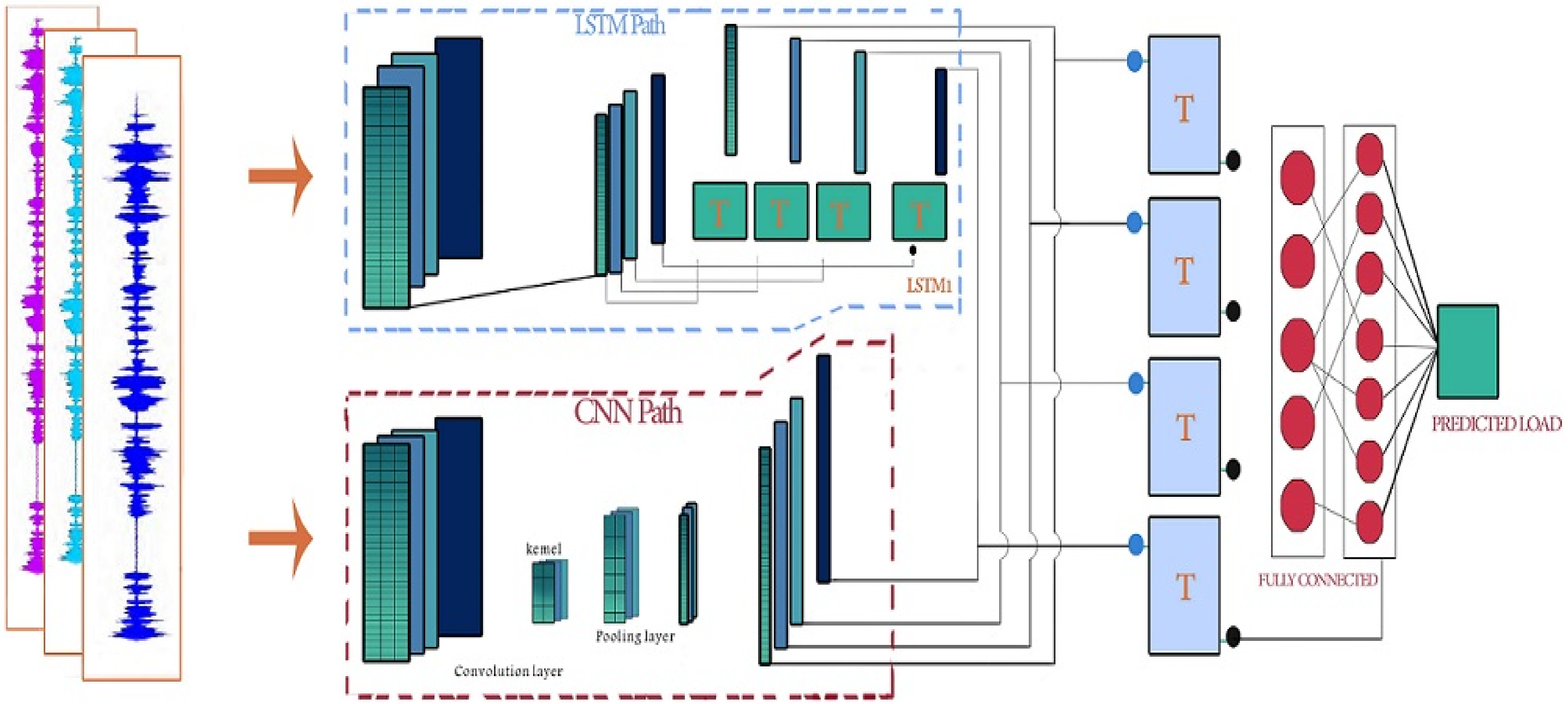
Including 3 different gates.

Looking back to previous data.

Useful to find the dependency.



Proposed model



Evaluation

RMSE : is the standard deviation of the residuals (prediction errors)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

MAPE: is the mean or average of the absolute percentage errors of forecasts

$$\text{MAPE} = \frac{\sum \frac{|\text{A-F}|}{\text{A}} \times 100}{\text{N}}$$

R-Squared: is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$

Results of Malaysian data

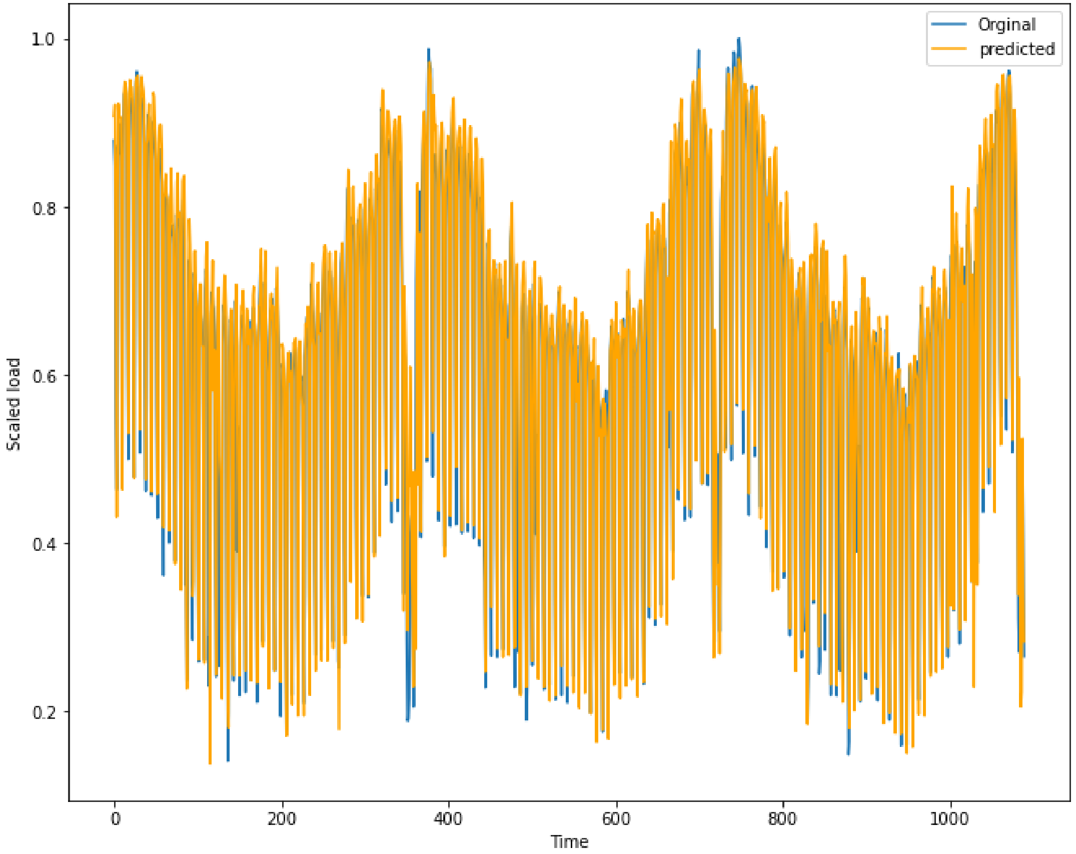
Models	RMSE	MAPE	R-squared score	Runtime (s)
ARIMA	0.102	3.56	94.19%	451.12
ETS	0.36	8.81	90.06%	380.35
Linear regression	0.092	2.33	95.50%	12.41
SVR	0.272	7.63	90.40%	10.23
DNN	0.128	3.62	95.38%	199.12
Vanilla LSTM	0.097	3.11	96.63%	902.56
CNN-LSTM	0.053	2.43	97.49%	487.33
Proposed model	0.031	2.08	98.23%	92.47

Results of German data

Models	RMSE	MAPE	R-squared score	Runtime (s)
ARIMA	0.201	18.4	80.04%	179.89
ETS	0.316	33.63	70.1%	167.03
Linear regression	0.214	19.12	79.86%	4.32
SVR	0.247	22.41	74.39%	3.11
DNN	0.25	26.47	73.47%	199.12
Vanilla LSTM	0.197	13.20	83.17%	431.11
CNN-LSTM	0.207	15.02	79.75%	180.22
Proposed model	0.061	5.12	91.18%	65.34

Results

Results of German data



Results of Malaysian data

Validation of the proposed model

Validation

The proposed model

1) Different Time horizons

2) Single building data

3) Exogenous variable

1) Different Time Horizons

Malaysian data. RMSE:

Model	1 Hour	24 Hours	48 Hours	10 Days
Vanilla LSTM	0.097	0.121	0.189	0.197
CNN-LSTM	0.053	0.069	0.0782	0.082
Proposed model	0.033	0.0379	0.0401	0.0575

R-Squared score:

Model	1 Hour	24 Hours	48 Hours	10 Days
Vanilla LSTM	96.63%	95.21%	92.65%	92.03%
CNN-LSTM	97.49%	96.62%	94.31%	92.88%
Proposed model	98.14%	97.55%	96.64%	94.16%

Université Concordia University

1) Different Time Horizons

German data. RMSE:

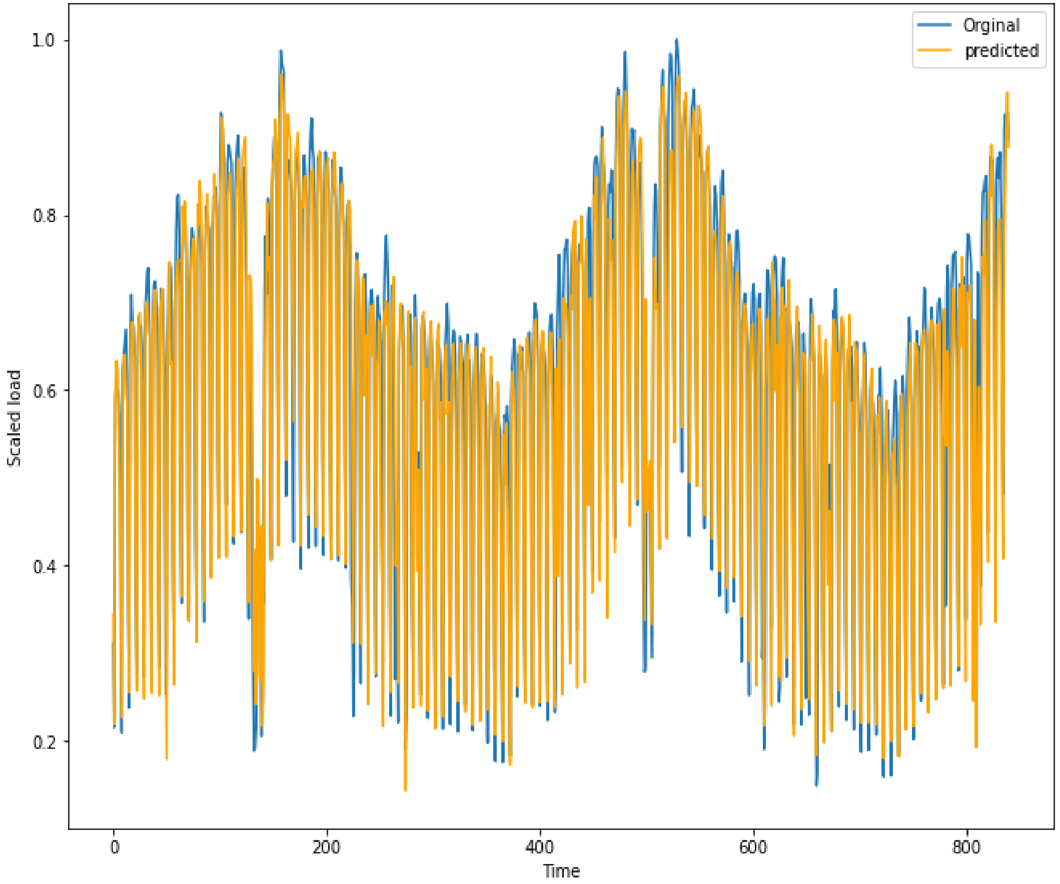
Model	1 Day	7 Days	10 Days	30 Days
Vanilla LSTM	0.207	0.215	0.231	0.312
CNN-LSTM	0.197	0.201	0.209	0.279
Proposed model	0.0063	0.0752	0.0761	0.117

R-Squared score:

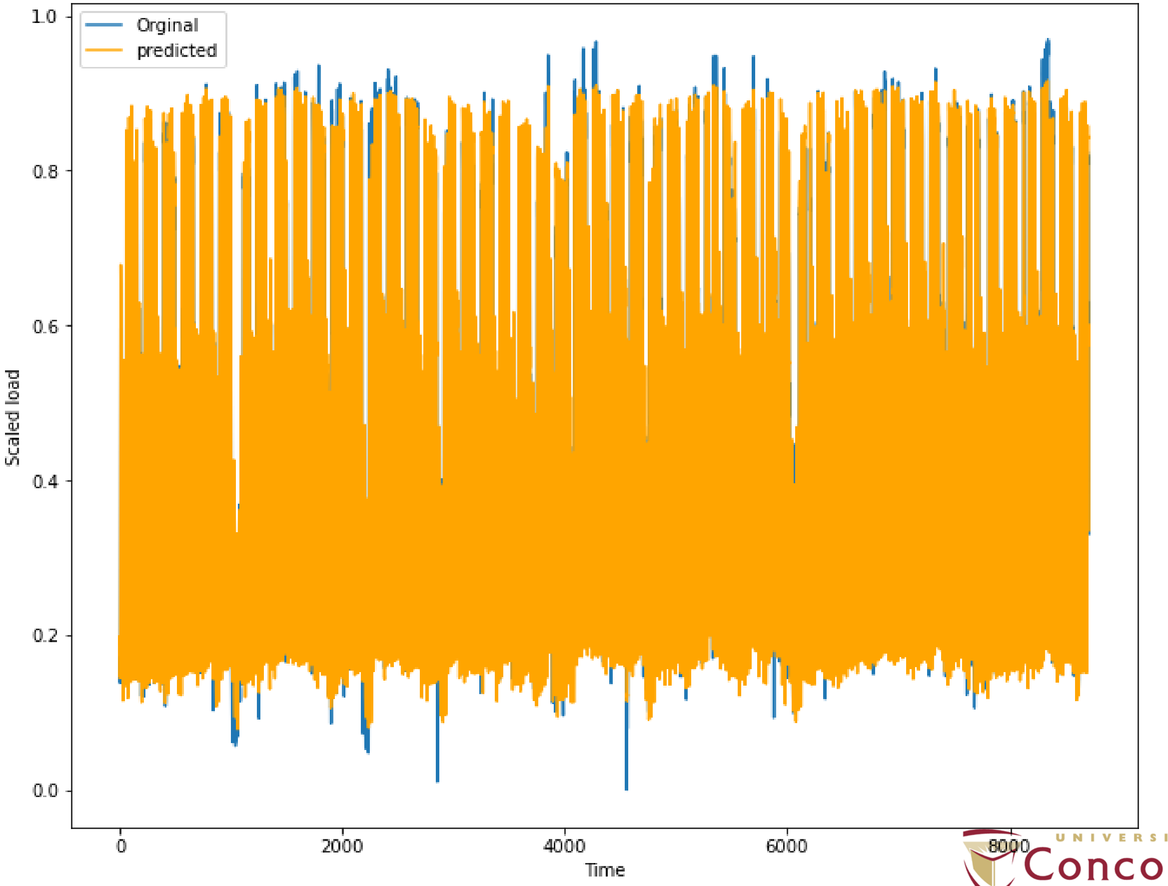
Model	1 Day	7 Days	10 Days	30 Days
Vanilla LSTM	79.75%	78.88%	78.02%	74.14%
CNN-LSTM	83.17%	80.87%	79.65%	76.88%
Proposed model	91.31%	89.53%	89.18%	82.49%

Results

Results of next 30 days German data

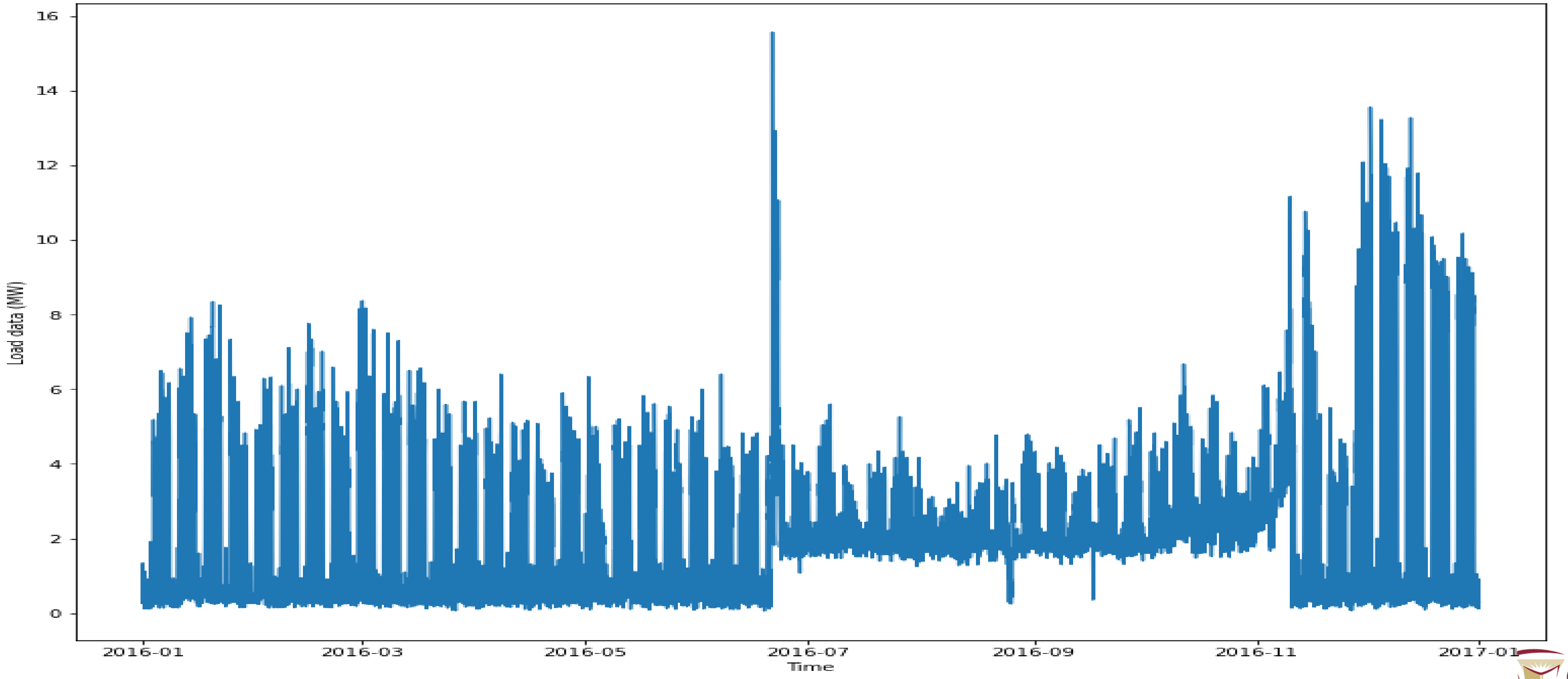


Results of next 10 days Malaysian data

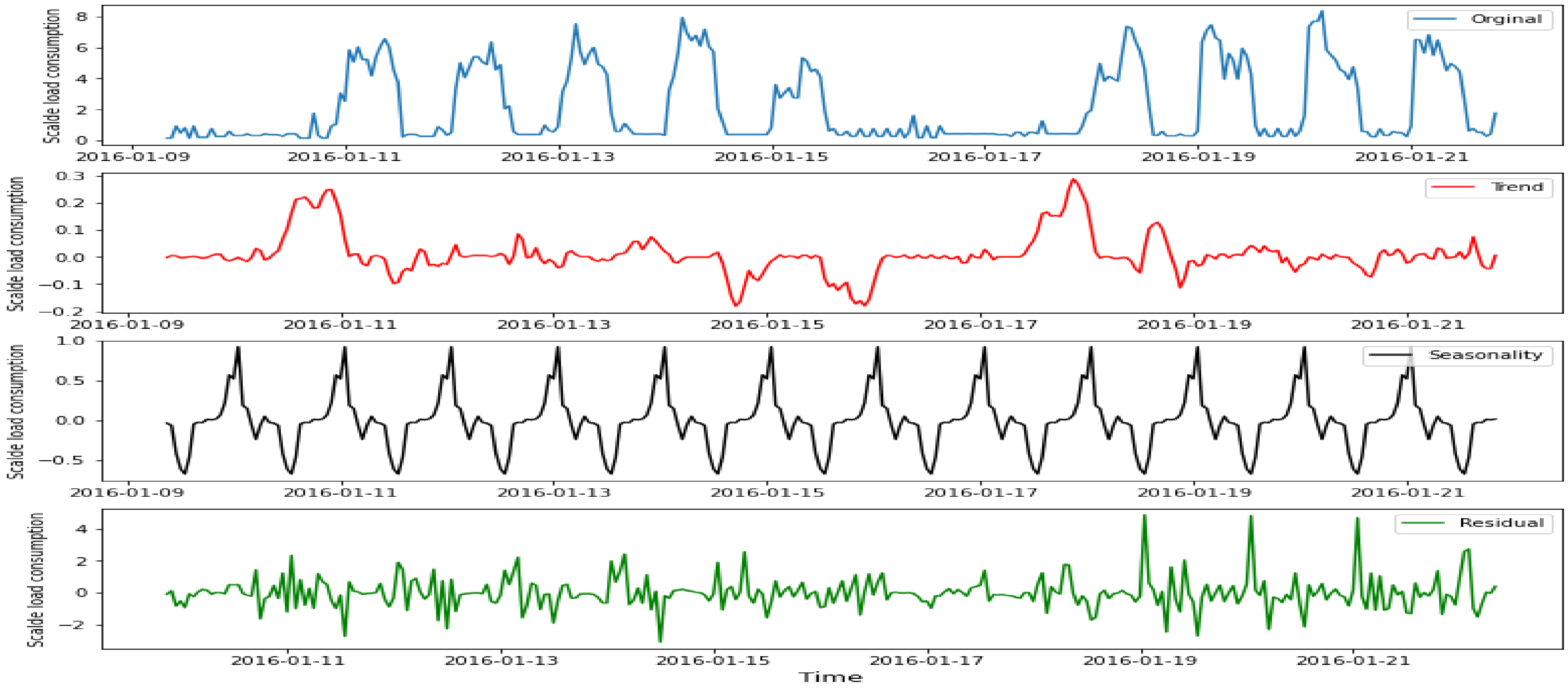


2) Single Building data

One-year hourly electricity and weather data→ Single building→ Grenoble



2) Single Building data



Results

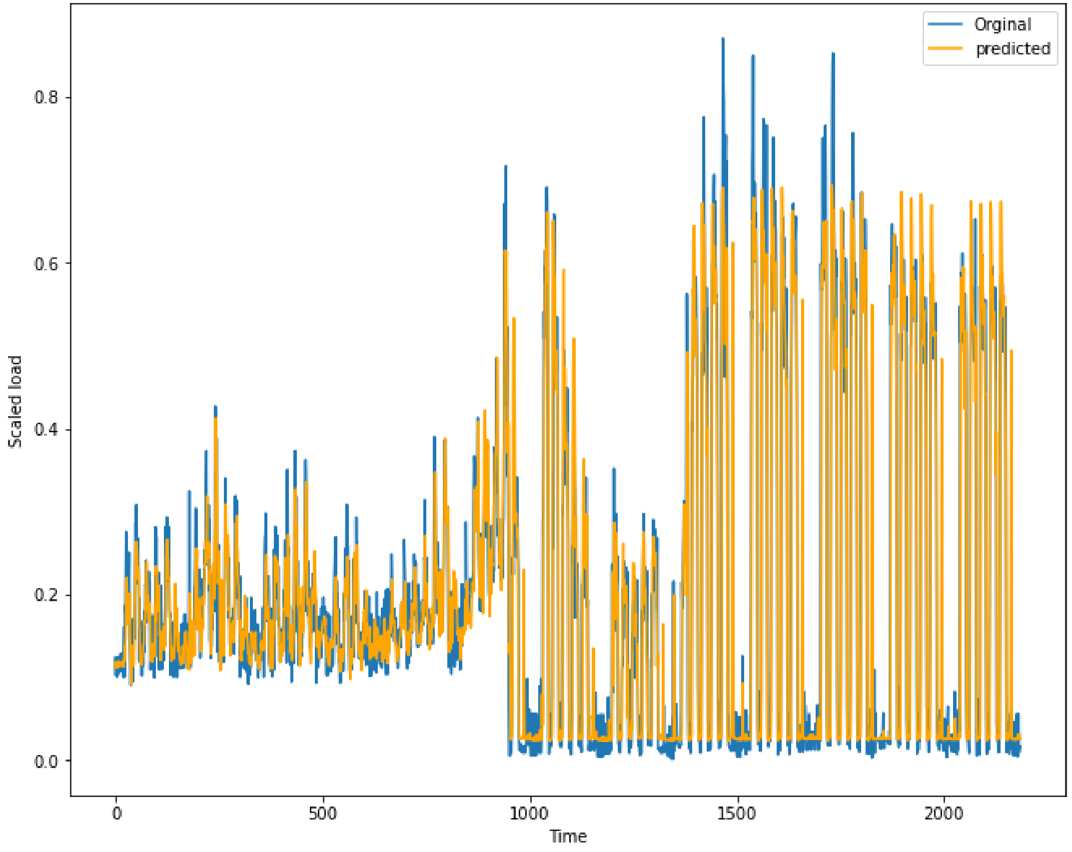
Training the model with 9 month and test on 3 month.

Next hour prediction.

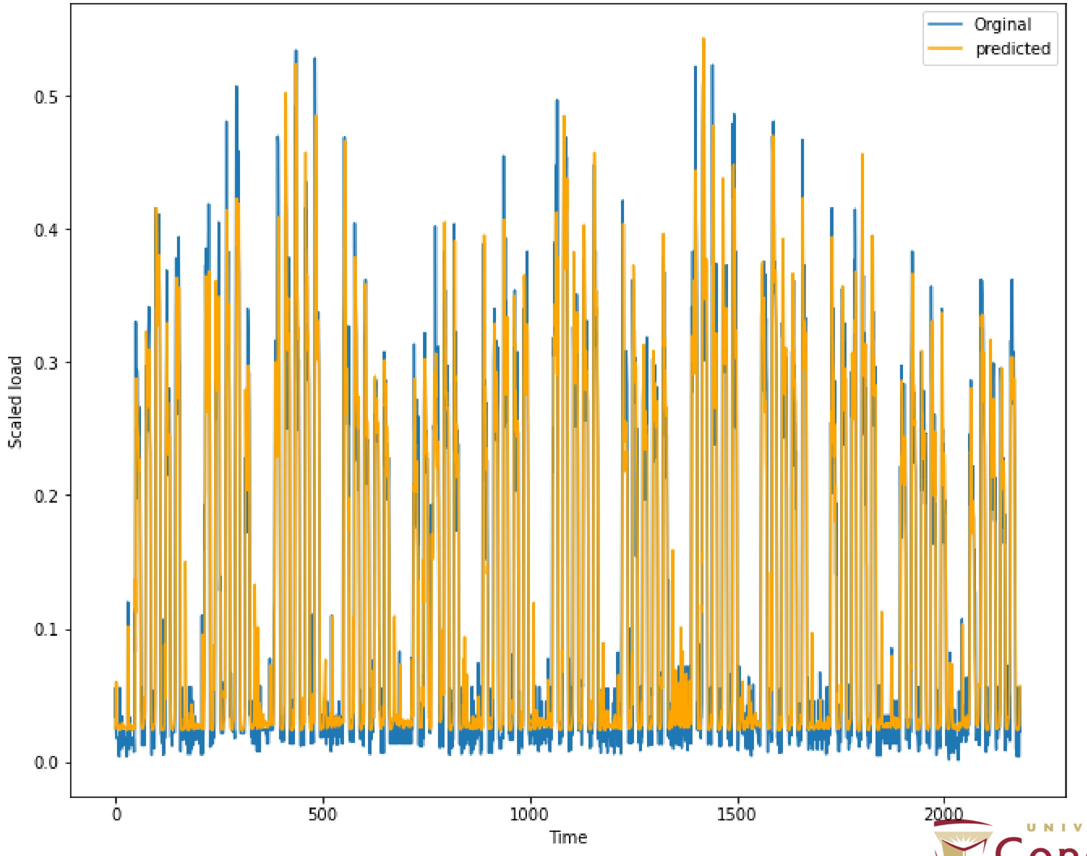
Time	RMSE	R-Squared score
Jan-Sept	0.0585	83.3%
Apr-Dec	0.0509	86.39%
Average	0.0547	84.84%

Results

First 9-month results

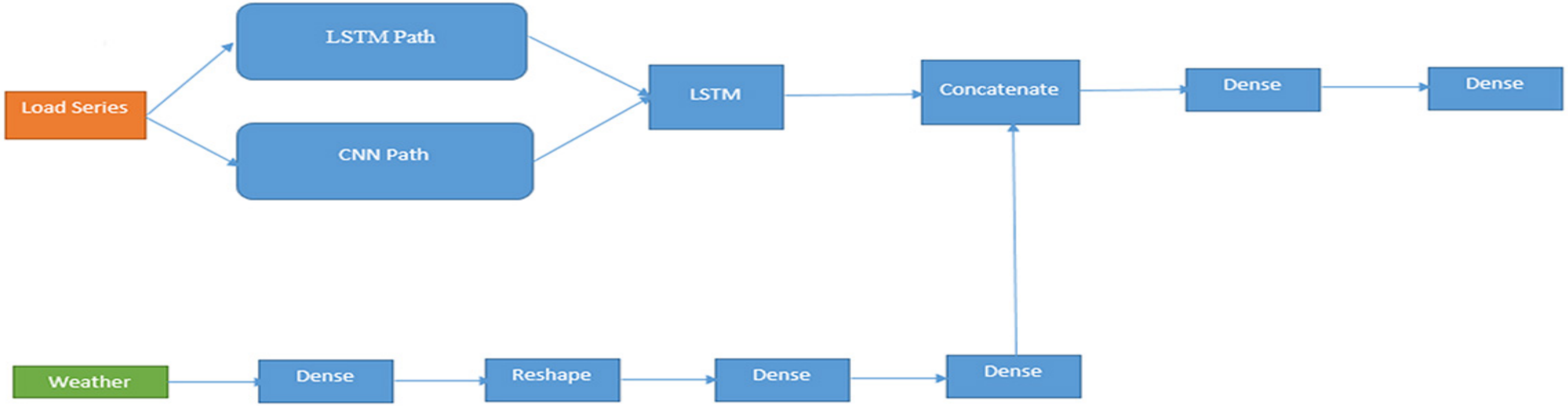


Second 9-month results



3) Exogenous variable

Weather data. Modification in the architecture.



3) Exogenous variable

Malaysian data: Correlation=0.56.

Time	RMSE	R-Squared	Runtime (s)
1 Hour	0.0289	99.15%	210.62
24 Hours	0.0302	98.2%	383.9
48 Hours	0.0385	97.26%	795.26
10 Days	0.492	96.05%	2821.08

French data: correlation= -0.08

Time	RMSE	R-Squared	Runtime (s)
Jan-Sept	0.0579	83.44%	260.18
Apr-Dec	0.0501	86.62%	165.73

Université Concordia University

Conclusion

- Even though regression-based models perform well, they are not trustful enough.
- Parallel CNN-LSTM models provide better performance than consecutive models.
- Deep Learning models are powerful models which can be used for STLF and LTLF.
- Weather can improve the accuracy of the models, but it leads to higher runtime.
- The effect of weather data on results depend on the correlation value.

Future Work

Modifications and improvements of the models.

Carrying out Long-term load forecasting using more variables.

Real-time forecasting (IoT).

Publication

Behnam Farsi, Manar Amayri, Nizar Bouguila and Ursula Eicker . On Short-Term Load Forecasting Using Machine Learning Techniques. IEEE Access Journal, 2020, submitted.

Thanks