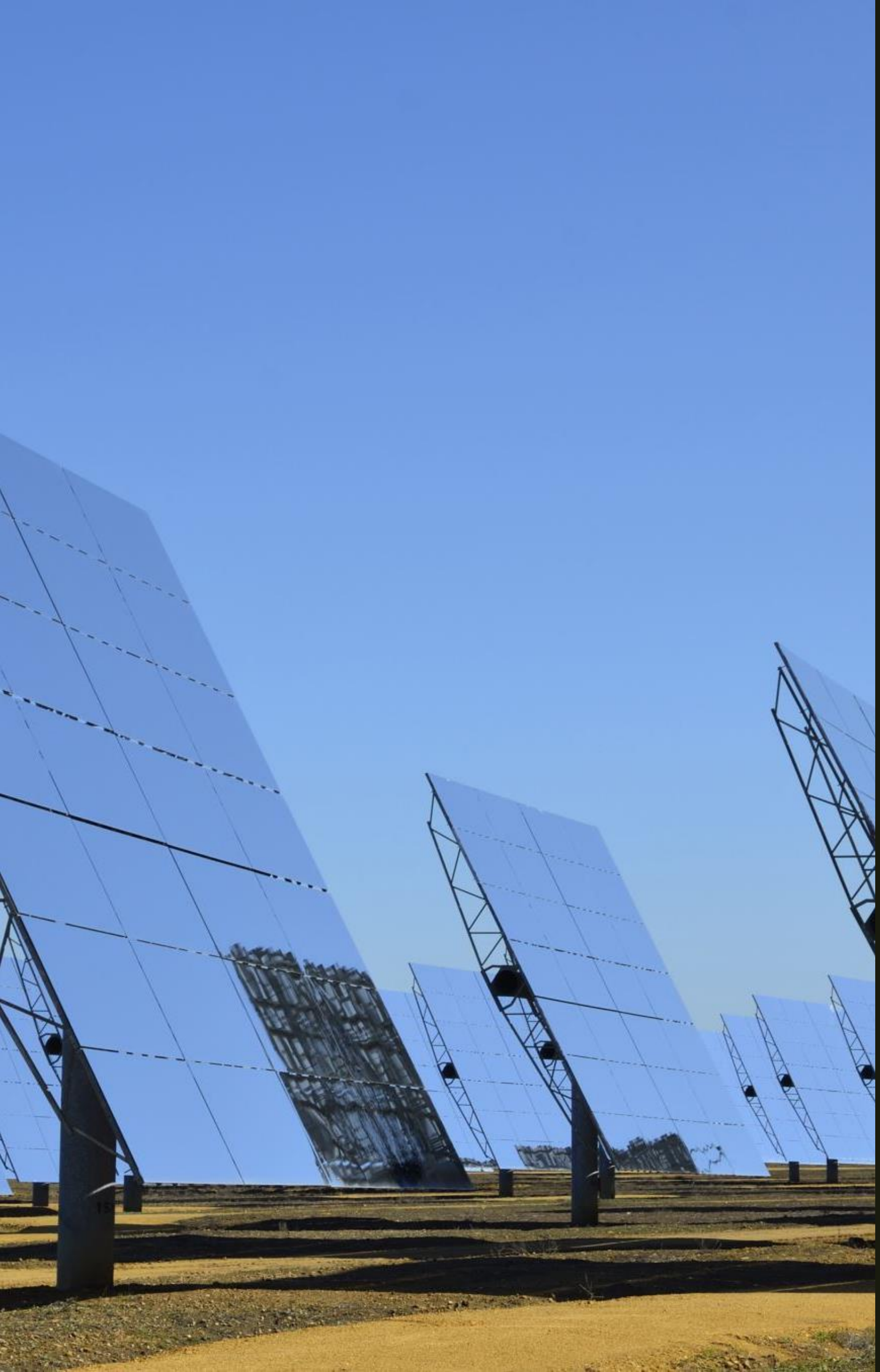
EXPLORING CLIMATE DATA’S RELEVANCE TO PREDICT ENERGY CONSUMPTION

Conducted in Colaboration with Energinet



Task

- Can climate data be used in order to produce precise forecasting for electricity consumption?

Why climate data?

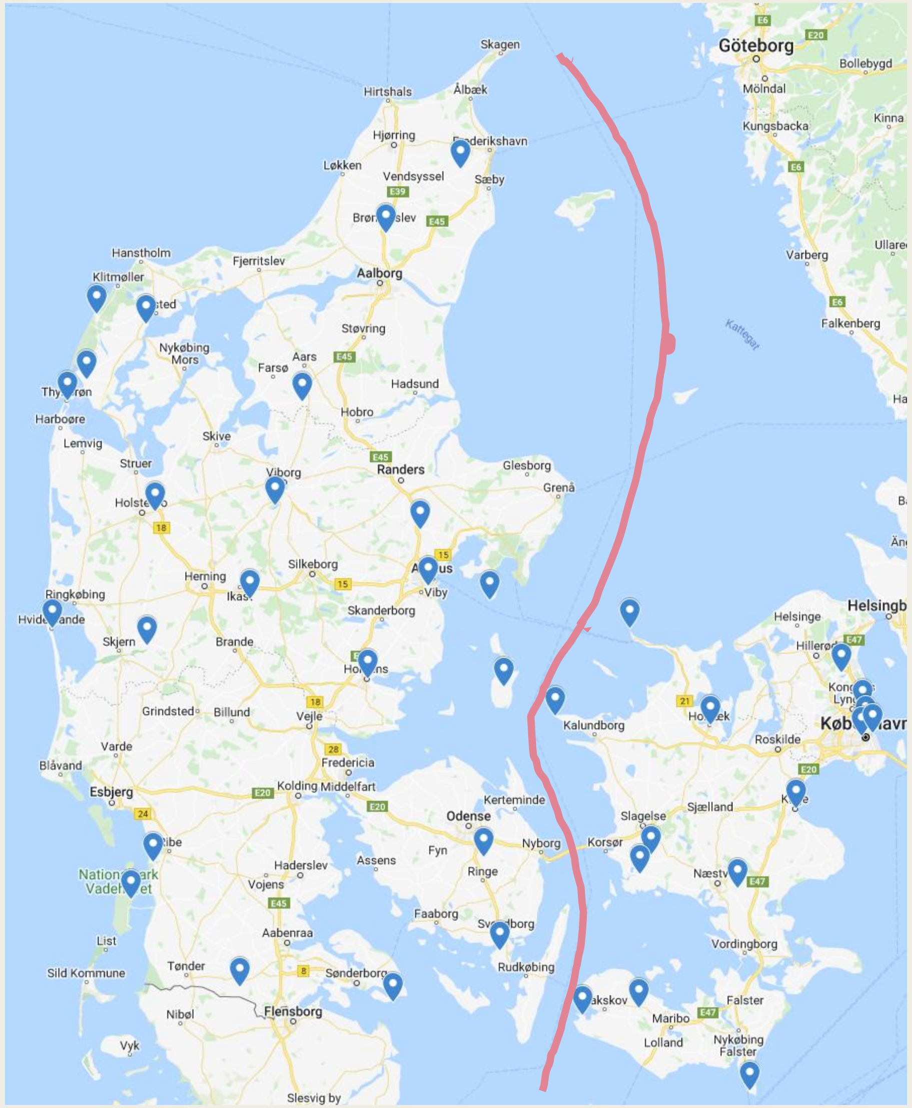
Increasingly many sectors are dependent on electricity

Stable electricity grid

80% of energy consumption goes to residential heating

The danish electrical grid and weather stations

- Two zones, DK1 & DK2
- 40 weather stations
- 25 for DK1
- 15 for DK2



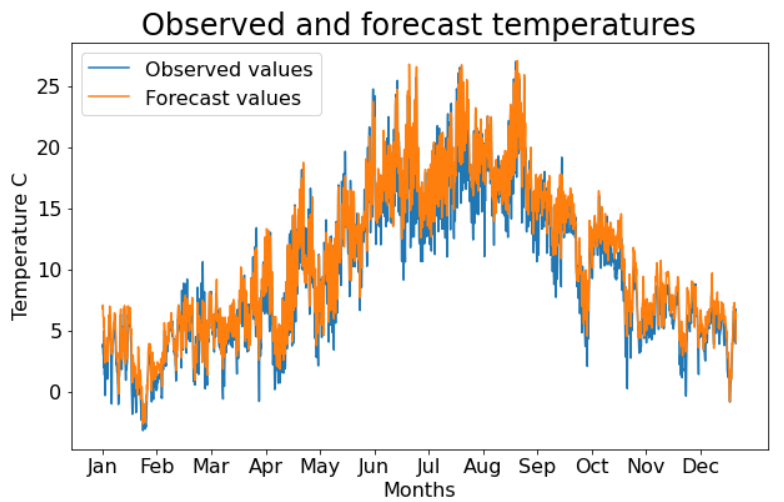
3 types of data

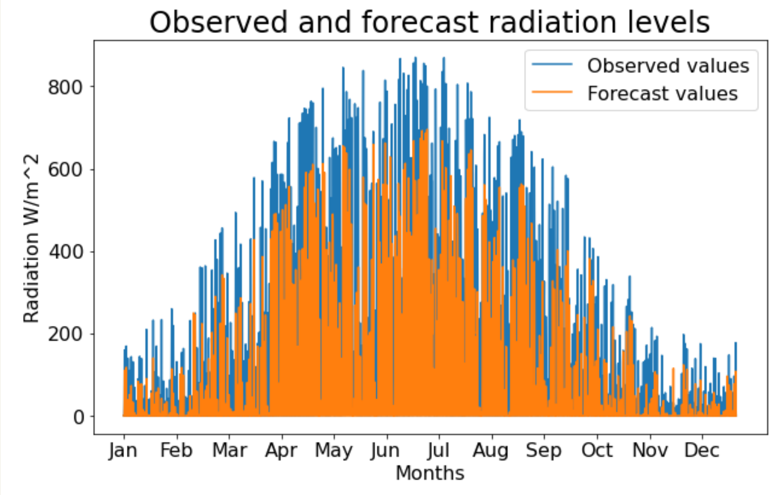
- Data for electricity consumption
- Weather data for observed values
  - Will be used to train the model
- Weather data from forecasts
  - Will be used as test set

Accuracy of DMI's forecasts?

- Temperature surprisingly accurate
- Radiation not so much
- Easier to predict temperature

R^2 values for forecast data		
	1 hour ahead	48 hours ahead
temperature	94.2	93.8
Radiation	70.5	67.4

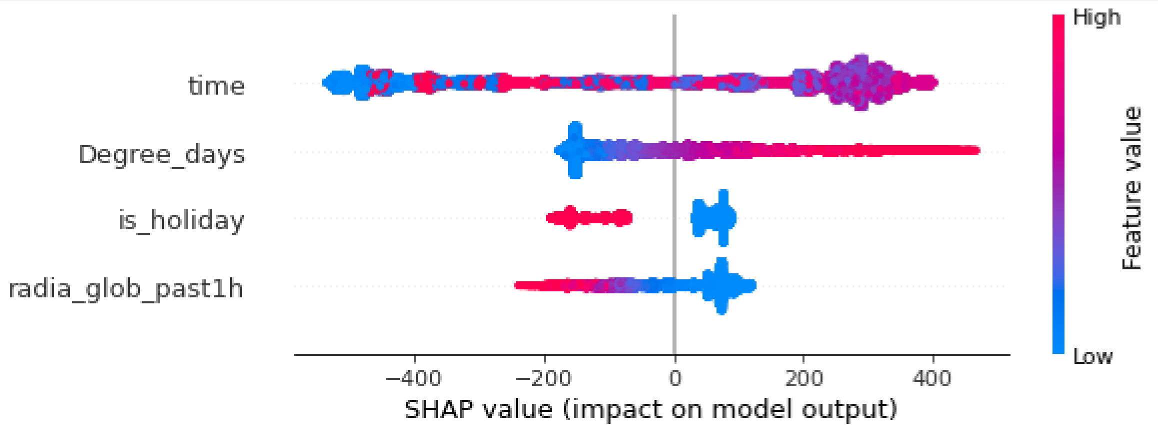




Data exploration

Training Dataset				
hour	Degree-days	Radiation	holiday	Consumption

- Searching for correlations in the aquired data
- SHAP-values
- Time is difficult to decipher

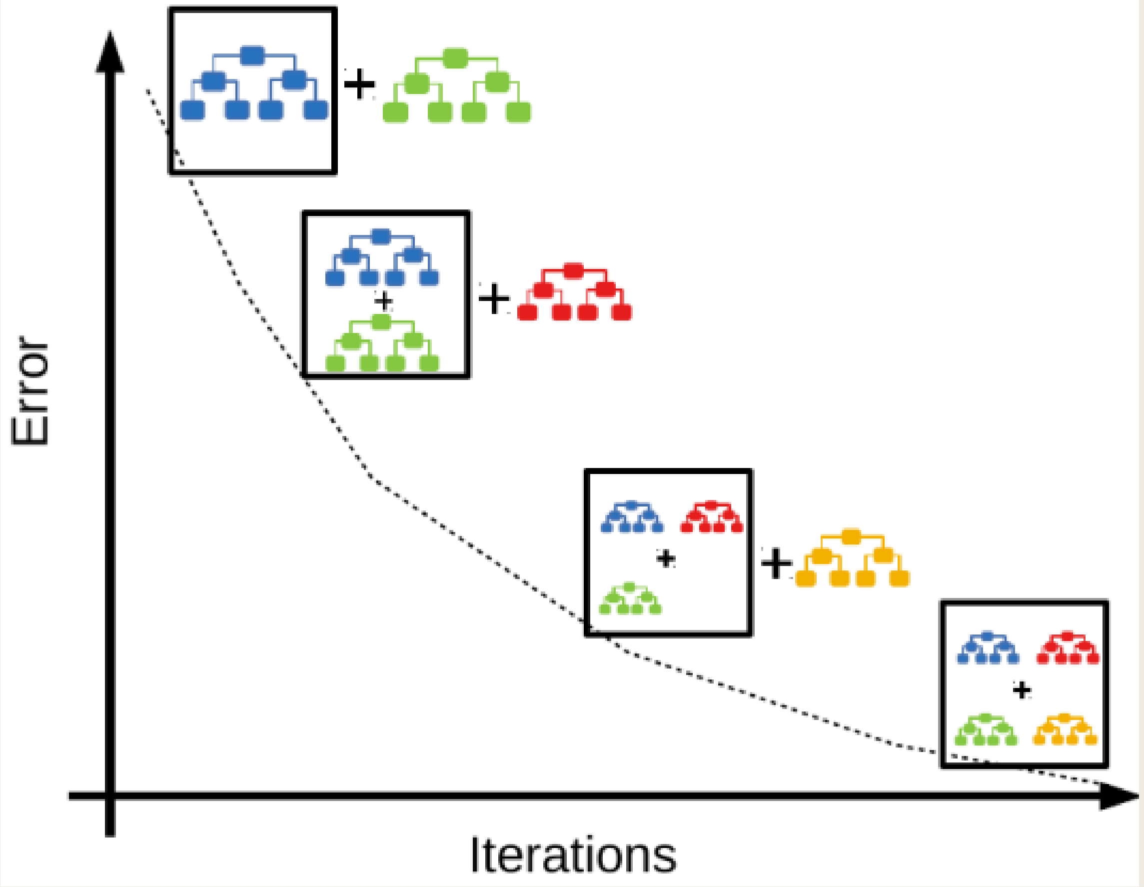


Two models

- XGBoost
  - Extreme Gradient Boosting
- LSTM-NN
  - Long-Short-Term Memory Neural Network

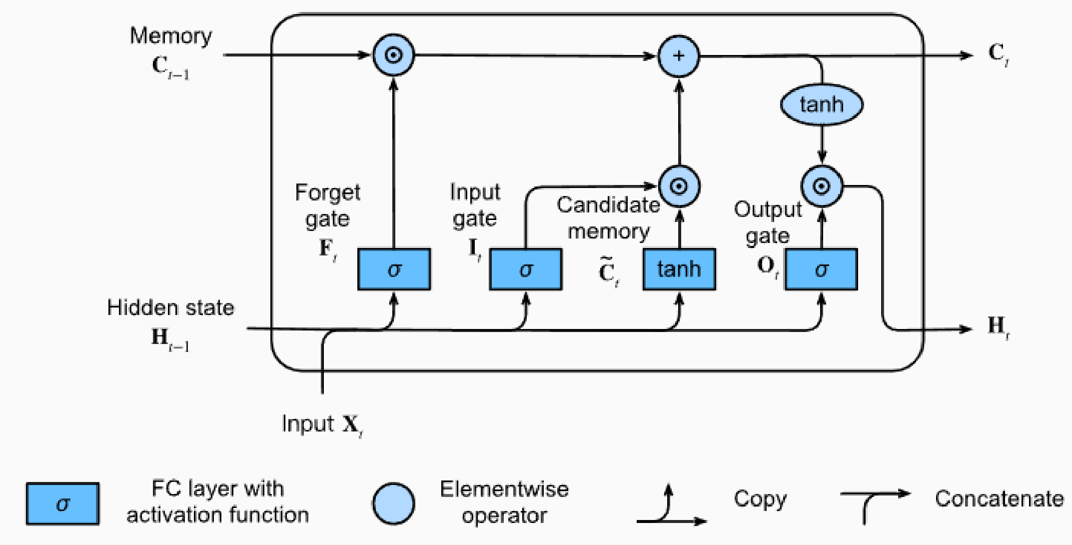
XGBoost

- Ensemble of weak learners
- Decision trees
- Gradient boosting
- eXtreme regularization
- Percieved as being very accurate



LSTM-NN

- Subcategory of a recurrent neural network
- Subcategory of a neural network
- Can store information between iterations
- Hidden state



Implementation

- Train data from 2010-17
- Validation set from 2018-19
- Two error metrics:
  - Mean Squared Error (MSE)
  -R^2 (R squared)
- A naive model

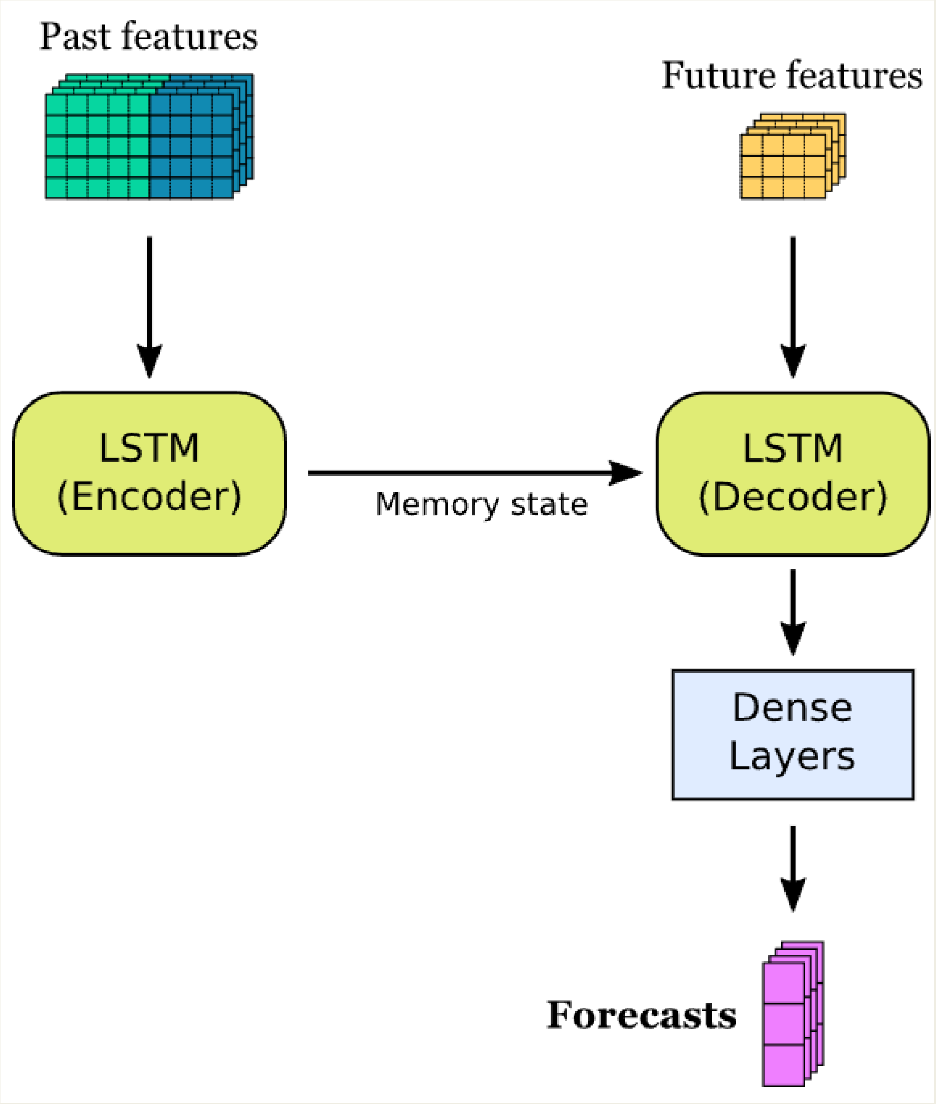
Optimal XGBoost

- Found by brute-forcing
- Proving the speed of the XGBoost
- 341 models trained

Zone	Max Tree Depth	Learning Rate	γ	Min Child Depth	Iterations
DK1	10	0.01	4	8	500
DK2	10	0.03	6	8	100

Encoder/Decoder LSTM

- For training everything is observed values
- Different for actual forecasts



Optimal LSTM-NN

- Same setup
- Different epochs
- Adam optimizer overfitted

Zone	Past input	Dropout	Activation function	Optimizer	Momentum	Learning rate	Loss function	Epochs
DK1	48	0.2	Relu	SGD	0.1	0.01	MSE	700
DK2	48	0.2	Relu	SGD	0.1	0.01	MSE	1000

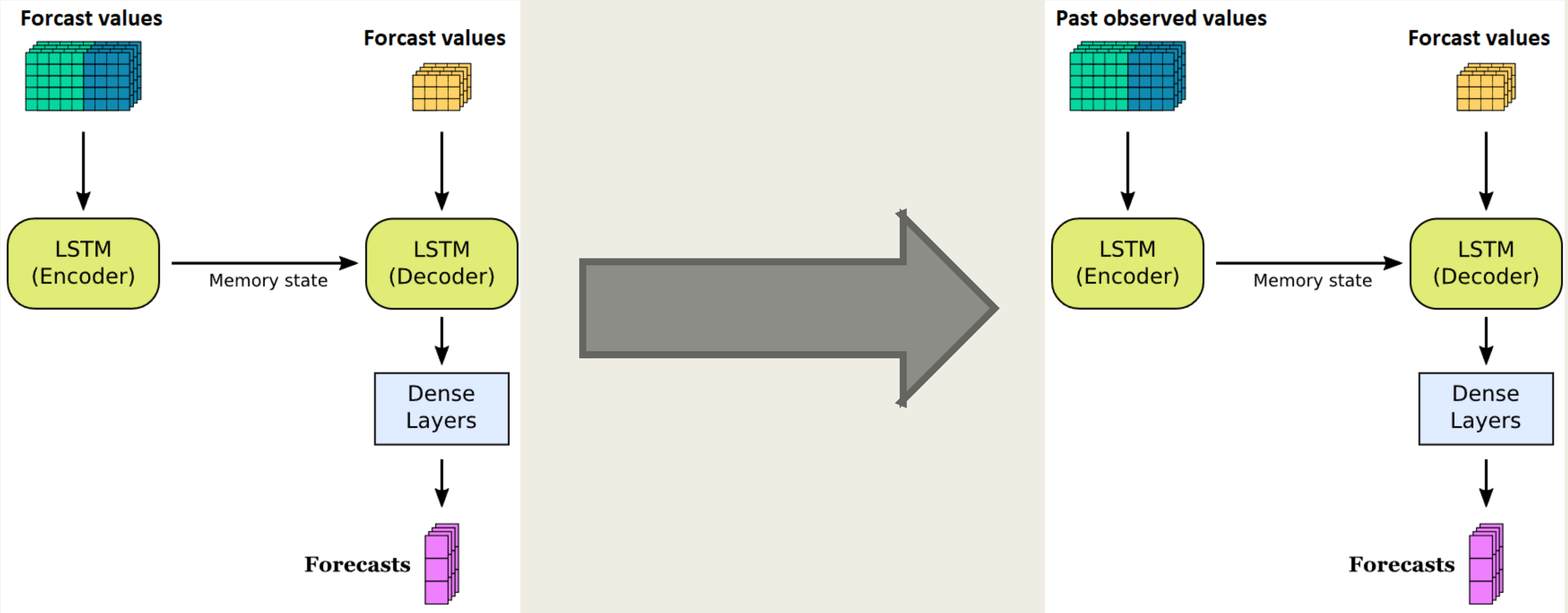
Training results

- Overall great performance
- XGBoost best for DK1
- LSTM-NN barely outperforms XGBoost for DK2
- Both outperforms the naive model

Results of predictions on the validation set			
Model	zone	MSE	R^2
XGBoost	DK1	18645.2	90.5
LSTM	DK1	22104	88.7
Naive model	DK1	67384.9	64.7
XGBoost	DK2	7187.5	92.2
LSTM	DK2	7095.9	92.3
Naive model	DK2	13714.9	84

Now for something very interesting

- New results
- After the project was turned in a problem in the code was found
- No observed values in the forecasts



Why did it happen?

- The results appeared to be true
- I did not illustrate the results enough
- Rookie mistake
- Rushed

Results of predictions on the test set			
Model	zone	MSE	R^2
XGBoost	DK1	33216.7	82.7
LSTM	DK1	33100.4	82.7
Naive	DK1	67384.9	64.7
XGBoost	DK2	9279.9	89.2
LSTM	DK2	17038.9	80.2
Naive	DK2	13714.9	84

New results

- Not better
- Atleast they are correct
- LSTM for DK2 is overfitting
  - It was run with lower epochs and the accuracy went up

Results of predictions on the test set			
Model	zone	MSE	R^2
XGBoost	DK1	33216.7	82.7
LSTM	DK1	33254.4	82.6
Naive	DK1	67384.9	64.7
XGBoost	DK2	9279.9	89.2
LSTM	DK2	19784	70.5
Naive	DK2	13714.9	84

Conclusion

- Climate data can be used to predict elecricity consumption
- In terms of this project, LSTM is not worth it
- XGBoost performs well, for less recources

Additional discussion points

- The LSTM should be trained differently
- Using months to try and capture summer months
- The naive model is tough competition
- Illustrate your predictions in different manners