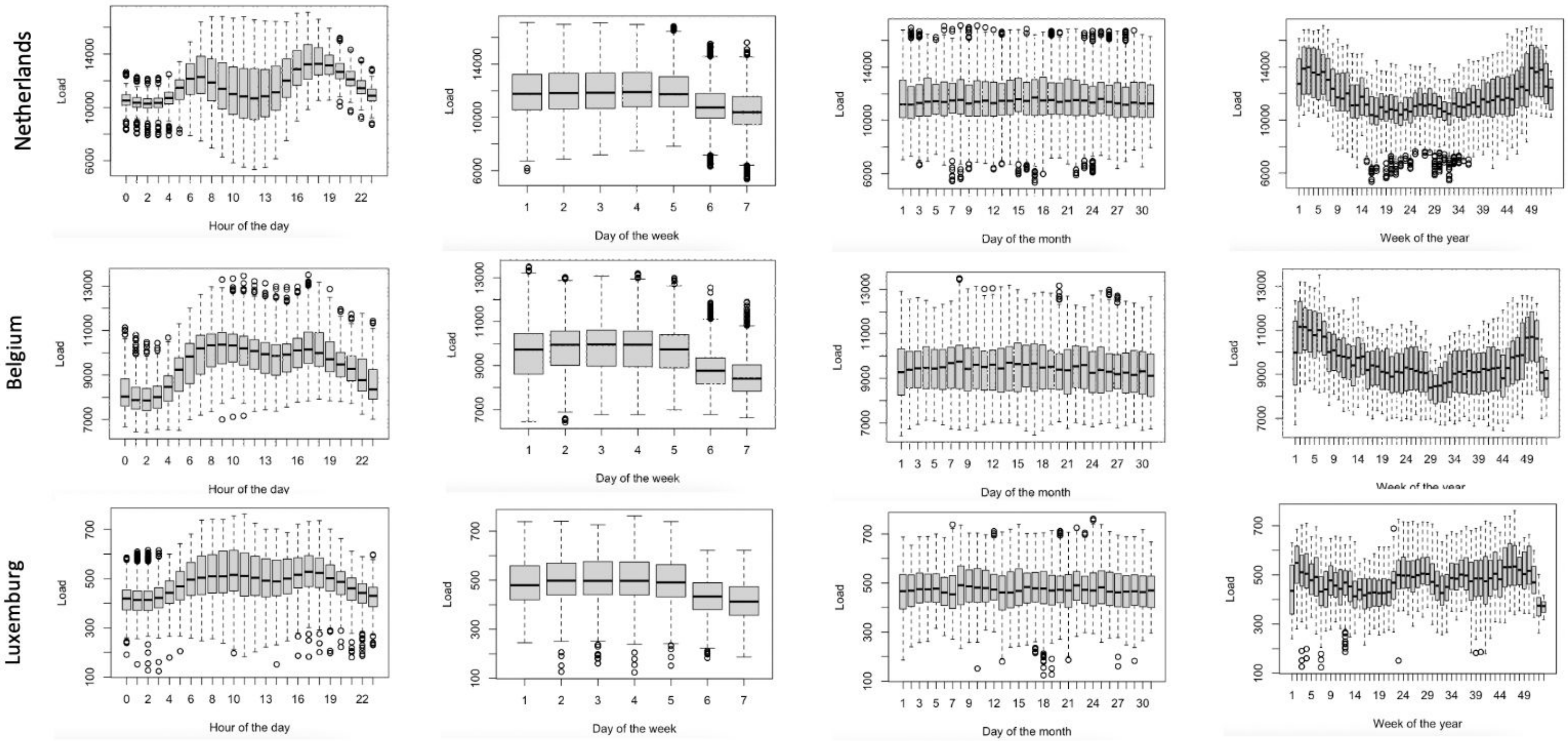
Forecasting Study of the Energy Demand in the Netherlands, Belgium, and Luxembourg

Alex Amaguaya - 242833
Sebastian Reyes - 242966
Anarghya Murthy - 243089

Dataset

- Countries: Belgium, Netherlands, Luxemburg.
- Tables:
    - MOSMIX: Temperature, Wind Speed, Global Irradiance, Wind Direction, Effective Cloud Cover
    - ENTSO-E: 240 hours-ahead load forecasts that are used as a benchmark for our own model
- The initial date for the data sets used from all countries was January 12, 2021, and the end date was November 28, 2022.
- Number of observations 163.413

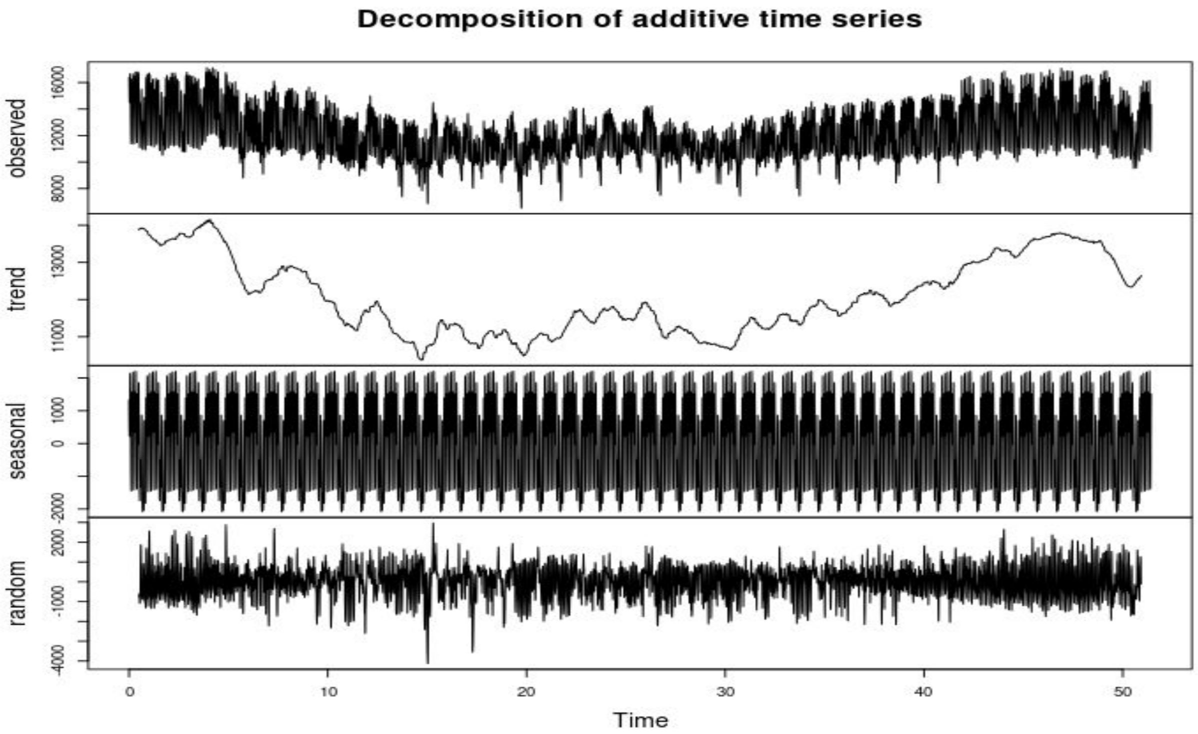
Exploratory Data Analysis



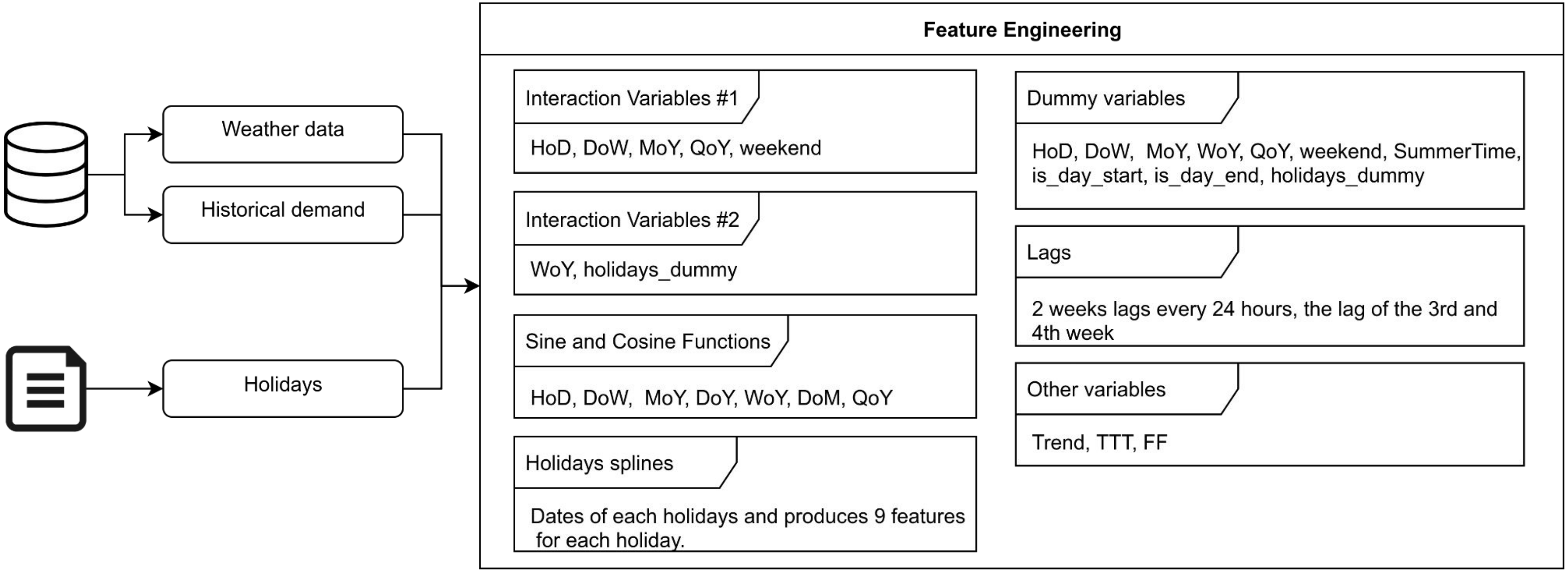
Data Preparation

- Data observations from 1st Jan., 2015, to 28th November, 2022
- Missing values were imputed using spline interpolation
- Spline interpolation is better suited than linear interpolation because this type of data is unlikely to have linear relationships between data points

Time Series Decomposition of Energy Load in Netherlands



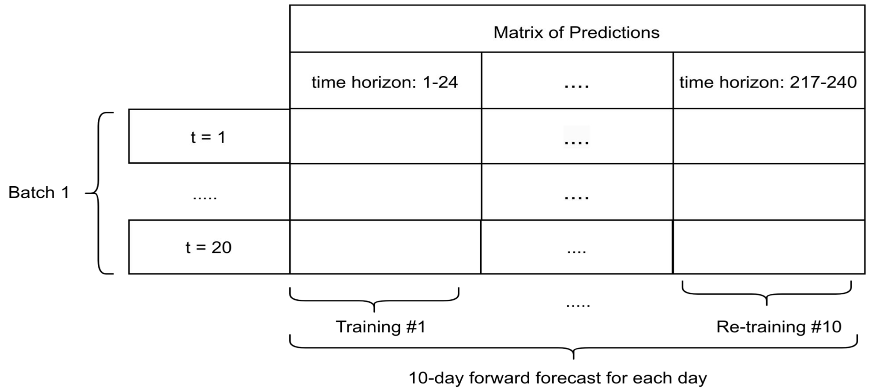
Feature Engineering

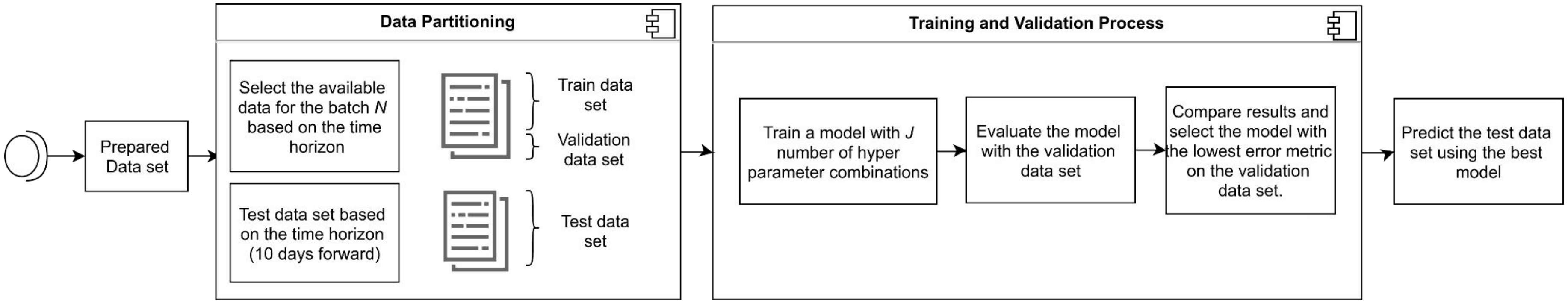


Methodology

- Elastic Net: it combines both the Ridge and Lasso regression techniques that seek to optimize the trade-off between variance and bias in order to minimize the error.
- Gradient Boosting: is a model that estimates a prediction through a combination of weak learners, usually decision trees, to create a strong predictive model through an iterative process.
- Random Forest: is an ensemble algorithm that builds decision trees on different samples and takes the majority vote for classification or regression.

Experiments Pipeline





Results - Error comparison

Netherlands	**Model**	Ensemble between Elastic Net, Random Forest and AR	Ensemble between Random Forest and AR	Ensemble between Elastic Net, and Random Forest	Ensemble between Elastic Net, Gradient Boosting, Random Forest and AR	Ensemble between Elastic Net and AR	Random Forest	Ensemble between Elastic Net, Gradient Boosting and AR	AR	Elastic Net	Ensemble between Elastic Net, and Gradient Boosting	Ensemble between Random Forest, and Gradient Boosting	Ensemble between Gradient Boosting and AR	Bench	Gradient Boosting
	RMSE	924,20	931,31	943,27	962,94	977,56	982,02	1.002,69	1.033,28	1.063,77	1.063,80	1.068,89	1.070,22	1.235,93	1.297,25

Belgium	**Model**	Bench	Ensemble between Random Forest and AR	Random Forest	Ensemble between Elastic Net, Random Forest and AR	Ensemble between Elastic Net, and Random Forest	Ensemble between Elastic Net, Gradient Boosting, Random Forest and AR	Ensemble between Elastic Net and AR	AR	Ensemble between Random Forest, and Gradient Boosting	Ensemble between Elastic Net, Gradient Boosting and AR	Ensemble between Gradient Boosting and AR	Ensemble between Elastic Net, and Gradient Boosting	Elastic Net	Gradient Boosting
	RMSE	322,22	525,40	528,80	541,48	561,66	581,44	605,49	625,00	629,35	636,62	669,91	709,83	723,44	845,85

Luxembourg	**Model**	Ensemble between Random Forest and AR	Random Forest	Ensemble between Random Forest, and Gradient Boosting	Ensemble between Gradient Boosting and AR	AR	Gradient Boosting	Ensemble between Elastic Net, Gradient Boosting, Random Forest and AR	Ensemble between Elastic Net, Random Forest and AR	Ensemble between Elastic Net, Gradient Boosting and AR	Ensemble between Elastic Net, and Random Forest	Ensemble between Elastic Net and AR	Ensemble between Elastic Net, and Gradient Boosting	Elastic Net	Bench
	RMSE	66,68	68,44	68,96	69,93	71,81	74,53	140,88	181,40	182,01	266,89	267,17	267,41	531,27	5.351,84

Results - Speed comparison

Algorithm	Gradient Boosting	Random Forest	Elastic Net	AR
Hyper parameters	**max_depth:** 8-9, **num_leaves:** 23-24, **min_sum_hessian_in_leaf:** 34-35, **num_iterations:** 29-31, **lambda_l1:** 0.12, **lambda_l2:** 0.08, **learning_rate:** 0.38-0.39	**max_depth:** 5-12, **num_leaves:** 20-27, **num_iterations:** 40-50, **lambda_l1 and lambda_l2:** 0.01-0.1, **learning_rate:** 0.18-0.3	**lambda1:** 20-23, **lambda2:** 0.98-1	**order.max:** 672
Duration time	**NL:** 7,11 min. **BE:** 7,20 min. **LU:** 7,54 min.	**NL:** 10,03 min. **BE:** 10,4 min. **LU:** 10,51 min.	**NL:** 12,43 min. **BE:** 12,75 min. **LU:** 12,99 min.	**NL:** 13 sec. **BE:** 14,09 sec. **LU:** 29,44 sec.
Numb. of Features	**NL:** 955 **BE:** 957 **LU:** 966	**NL:** 955 **BE:** 957 **LU:** 966	**NL:** 955 **BE:** 957 **LU:** 966	**NL, BE, LU:** < 15 lags

Conclusions

- Trade-off between low error and high speed
- The detection of structural breaks or trend change points can help improve performance, but generally leads to longer training times.
- The ensemble estimations gain accuracy and provide the best result in comparison with the individual estimations.
- The ensemble estimator of Random Forest and AR are the best estimators for Belgium and Luxembourg. For the Netherlands it includes Elastic Net.
- Using an external deep learning model takes a lot of training time in comparison to a tree model.