Investigating the Prediction of aFRR Activated Volume and Price Using Machine Learning

Claire Lambriex*, Felix Preuschoff, Denise Bangkeling, Albert Moser

IAEW der RWTH Aachen University, Schinkelstraße 6, 52062 Aachen, Tel.: 0241/80-92475, c.lambriex@iaew.rwth-aachen.de, http://www.iaew.rwth-aachen.de

Abstract: A separate balancing energy market was introduced in Germany in 2020 for automatic and manual frequency restoration reserves (aFRR/mFRR), representing a new trading opportunity for market participants. The introduction of the balancing energy market makes participation in the energy auction possible without having been successful in the capacity auction, enabling more short-term trading of balancing energy. The market, however, possibly changes the performance of prediction models used by market participants to maximize profits. To address this, this work investigates the prediction of aFRR activated volume and price on the new balancing energy market using different machine learning methods and input data combinations. The best performing method and optimal number of features in terms of R^2-values were identified. Despite various attempts with different methods and input data combinations, the models for activated volume generally performed poorly. This outcome, however, is logical since balancing energy would not be needed if its activation were predictable. For aFRR energy prices, R^2-values of more than 95% were reached, indicating that with the appropriate input data combination and method, predicting aFRR energy prices is possible.

Keywords: Balancing Energy, Electricity Balancing, Machine Learning, Prediction

1 Background and Motivation

The balanced outcome of the European and German electricity markets, where supply meets demand, does not always correspond to a physical balance between generation and consumption of electrical energy. Imbalances can result from unexpected weather conditions or consumer behavior and cause a deviation of the grid frequency from its nominal value. To compensate such short-term imbalances, transmission system operators (TSOs) procure and, if necessary, activate balancing reserves, which are composed of standard products in Europe. Automatic and manual Frequency Restoration Reserves (aFRR/ mFRR) are standard products that are used to restore the frequency to its nominal value. Different auctions for each of the standard products are organized by TSOs to secure a sufficient amount of reserves and a cost-efficient activation of these reserves. For aFRR and mFRR, remuneration in these auctions consists of a capacity price for the reservation of balancing capacity and an energy price for the actual provision of energy when activated.

Balancing energy markets were introduced in Germany in November 2020 for aFRR and mFRR as a result of the Electricity Balancing Guideline (EB GL) issued in 2017 [1]. Before the introduction of balancing energy markets, balancing capacity bids were awarded in a capacity auction based on a capacity price merit-order. Energy price bids from balancing service providers that were accepted in the capacity auction were considered for activation. With the new balancing energy markets, participation in the energy auction is also possible without having been successful in the capacity auction, enabling more short-term bidding of balancing energy products. Balancing service providers can now act at shorter notice and better align their bids in the balancing energy markets with known awards from prior electricity auctions. To couple aFRR balancing energy markets at European level and enable cross-border exchange of balancing energy, a platform called PICASSO was established in June 2022 [2]. With the introduction of PICASSO, some market parameters like the remuneration rule and validity period were changed [3].

The balancing energy market provides a new opportunity for market participants like power plant and storage operators to sell their capacity. Such market participants employ prediction tools to maximize their profits across several electricity markets. The new balancing energy market and the changed market parameters could affect the ability to predict activated aFRR volumes and prices. The question that arises is how activated aFRR volumes and prices can be predicted and how well prediction models perform.

In related studies, attempts were already made to predict the balancing energy prices and/ or the volume of activated balancing reserves using machine learning [4–6]. The achieved results in such studies were of low performance. These studies, however, often analyzed a time frame during which market design changes were implemented. Furthermore, none of these studies analyzed the prediction of the balancing energy market after the introduction balancing energy market in Germany. To expand the current state of research, this study therefore bundles the attempt to predict aFRR activated volumes and prices, while analyzing a time frame after the introduction of the balancing energy market in Germany.

2 Methodology

This work uses supervised learning techniques to investigate the prediction of four target variables: positive and negative aFRR volumes and prices. Four different tree-based machine learning methods are investigated: Gradient Boosting, Random Forest, XGBoost and LightGBM. Different input data combinations are investigated for all methods and compared as well. The first step of the methodology is the collection of input data. Then new features are engineered from the existing input data and some features are transformed to deal with inadequate scaling. Finally, the models are trained and their performance is evaluated across the different methods.

2.1 Collection of Input Data

The input data for the investigations in this work contains market data like forecasted electricity generation, actual electricity generation, forecasted and actual electricity consumption and day-ahead market prices. Forecasted generation is used for variable renewable energy sources like solar energy, wind onshore and wind offshore. Actual electricity generation is used for variable renewable energy sources as well, but also for other energy sources such as biomass, hydropower and conventional sources. The consumption data contains the total grid load and load from hydro pumped storages. [7]

Next to general market data, data on activated balancing energy and its price, but also results from the balancing capacity market are used. [8, 9]

All input data is collected in 15-minute resolution for the time period between May 2021 and April 2023.

2.2 Input Data Generation and Transformation

After the collection of input data, the quality of the data is inspected. Missing values are dropped and data from different sources is joined. Next, features are generated from the collected input data (feature engineering). Features engineered are generation ramps, consumption ramps and forecast errors of generation and consumption. Generation and consumption ramps represent how fast generation or consumption of electricity from a certain source changed within 15 minutes. As values, these features are represented by the difference of generation or consumption values at timesteps t and $t-1$. Forecast errors of generation and consumption are represented by the difference between forecasted and actual values.

The target values contain a certain recurring pattern. Particularly for activated volumes of aFRR, a peak in activation is seen for certain hours of the day, as shown in Figure 1 for positive aFRR.

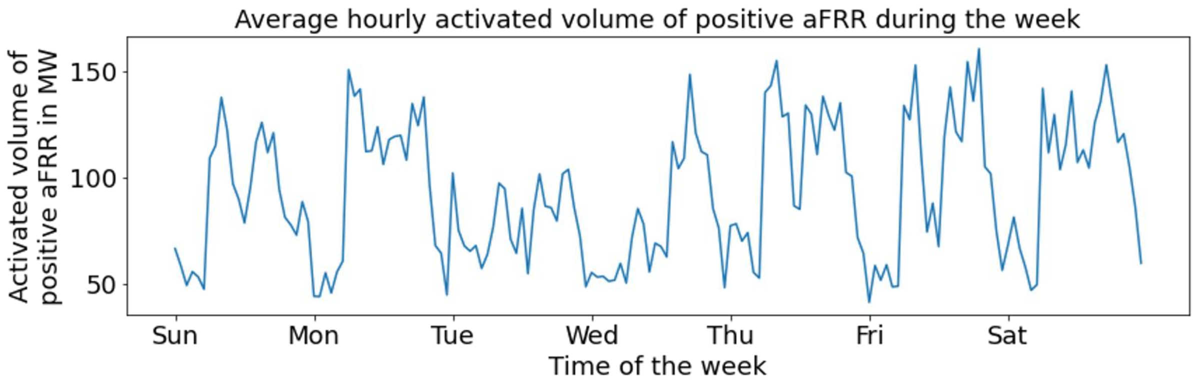

Figure 1: Seasonality in activated positive aFRR volume

For negative aFRR, a seasonality in activated volume can be seen as well, which is shown in Figure 2.

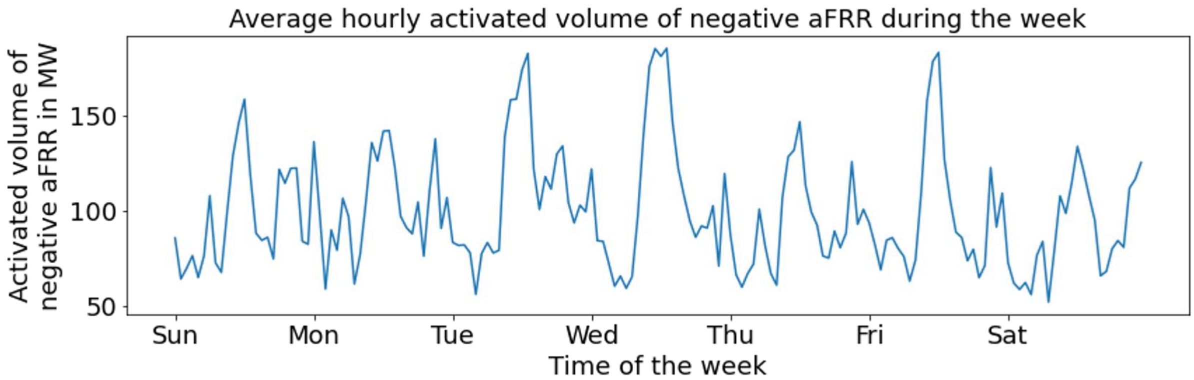

Figure 2: Seasonality in activated negative aFRR volume

To capture these relationships, categorical features for hour, time of day, day of week, month, and season are added. These features categorize which hour, time of day, day of week, month and season each activation occurred in. However, as categories are represented as numerical values in machine learning, these time-related features have to be encoded. The challenge that arises is to capture the cyclical nature of the time-related features in numerical values, i.e. the first day of the week follows the last day of the week. In the example of day of the week as a feature, the distance between Sunday and Monday is the same as between Monday and Tuesday, but if these features were to be represented as numbers 1 to 7, the distance between 7 and 1 is not the same as 1 and 2. To solve this issue, the numbers that represent the cyclical time features $feature_{value}$ are encoded into sine and cosine values. This way, the first number in the set of values always comes after the last. Formulas (1) and (2) are used for this.

$$feature_{sine} = \sin\left(\frac{2\pi \cdot feature_{value}}{feature_{max}}\right) \tag{1}$$

$$feature_{cosine} = cos\left(\frac{2\pi \cdot feature_{value}}{feature_{max}}\right) \tag{2}$$

Lastly, some feature values have to be transformed to support the algorithms' learning ability. For example, some source data are heavily skewed with most values close to zero and a few outliers. To even out this distribution, the values are transformed using a Yeo-Johnson transformation [10].

The different input data combinations that are compared vary in the use of actual and/ or forecasted generation and the handling of skewness. The different combinations are shown in Table 1.

Table 1 Input data combinations

Combination	***Generation and consumption data used***	***Max no. of features***	***Skewness threshold***
1	Only forecast values	19	1.5
2	Forecast and actual values	59	1.5
3	Forecast and actual values	59	1
4	Forecast and actual values	59	No transformation for skewed distributions

2.3 Machine Learning and Model Training

2.3.1 Machine Learning Methods

This study uses supervised learning, which means the input data is labeled. Each input has a specific "correct answer" or output. In this study, the output corresponds to the activated aFRR volume or its price. As the target prediction is a continuous value i.e. volume or price, it is a regression problem. [11]

The machine learning methods used in this study are Gradient Boosting (GB), Random Forest (RF), XGBoost (XG) and LightGBM (LG). These methods are based on combining two concepts: decision trees and ensemble learning. With a decision tree model, the model can be understood as a protocol that is followed by each observation of the input data, going through the nodes, and making decisions based on tests on its predictors [11]. Ensemble learning builds a stronger model by combining a series of weak models, in this case a series of decision tree models [12].

2.3.2 Hyperparameter Tuning

When building machine learning models, there are two different types of parameters that are to be determined. There are parameters that the model determines through training, these cannot be changed manually. Hyperparameters, on the other hand, are set manually before the training process. Hyperparameters can be optimized, a process also known as hyperparameter tuning. This improves model performance by finding the best hyperparameters for the specific case i.e. dataset and algorithm. Several methods for hyperparameter tuning exist, including Random Search. In Random Search, hyperparameters are selected at random until finally finding a near best set. [13, 14]

2.3.3 Model Training

The final input data is split into train and test sets randomly in a ratio of 80:20. The split is randomized to ensure an equal division of the input data, especially in regard to the categories i.e. to have equal observances in each season. The test set is used to evaluate model performance and is therefore not used for training. Hyperparameters of each model are tuned using a 5-fold cross validation on the training set using the randomized search algorithm.

For each method, the different input data combinations shown in Table 1 are investigated. In total, this results in 16 models with sets of best hyperparameters for each of the target variables. Each set of hyperparameters is used to initialize a model that is trained and evaluated. For each model, the features' importance is evaluated. The features are ranked in a list from most to least important. The model is consequently trained and evaluated with an increasing number of features starting from an initial model with only the most important feature and gradually with all features. This is done to find the optimal number of features, evaluating model performance at every increase of number of features.

2.4 Model Evaluation

Finally, model performance as well as feature importance is evaluated. Each machine learning method has an embedded feature importance evaluation that automatically evaluates feature importance during training. For every number of features in the input data, the model performance is evaluated. The evaluation metric chosen is the coefficient of determination R^2, which is a common assessment criterion for supervised learning methods.

3 Results

For each of the input data combinations shown in Table 1, the machine learning method that reached the best result in terms of R^2 value and the optimal number of features will be shown and discussed.

3.1 Volume

3.1.1 Positive aFRR

For the activated volume of positive aFRR, the best performing model for each combination of input data with the reached R^2 values and optimal number of features are shown in Figure 3.

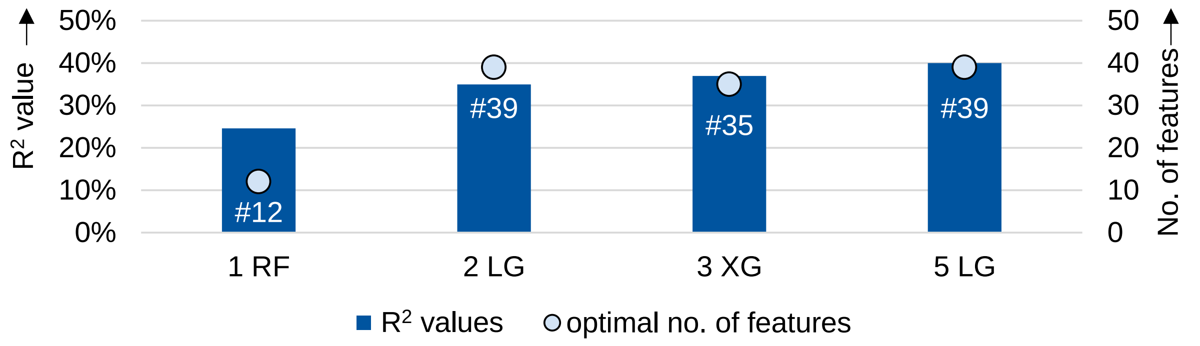

Figure 3: R^2 values and optimal number of features for each input data combination for positive aFRR volume

In the first combination only forecasted generation and consumption data were included. Using this input data, the best results were achieved by the Random Forest algorithm with an R^2 value of 24.59%. Starting from the second combination, both actual and forecasted values were included. Best performance was reached with the fourth input data combination using the LightGBM method. This combination reached an R^2 value of 40% with an optimal number of features of 39. Generally, it can be seen that model performances increased by adding the actual generation and consumption values. However, performances in general were not sufficient for this target variable.

Table 2 shows the most important features for each input data combination for the target variable positive aFRR volume.

Table 2 Best performing method and most important features for each input data combination for positive aFRR volume

Combination	***Best method***	***Most important features***
1	RF	Generation forecast others, generation forecast photovoltaics and wind, capacity price of positive aFRR
2	LG	Actual generation nuclear, actual generation other renewables, month
3	XG	Actual generation hydro pumped storage, actual generation hydro pumped storage ramp
4	LG	Actual generation hydro pumped storage, actual generation hydro pumped storage ramp, actual generation hydropower ramp

Table 2 shows that for the first input data combination, forecasted generation of variable renewable energy sources is an important feature. The absolute forecasted generation may be correlated with the magnitude of forecast errors, which are an important cause of aFRR activation. For the other combinations, hydropower-related features dominate, which might reflect the dominance of hydropower plants participating in the aFRR market in Germany [15]. The inclusion of actual generation and consumption data improves model performance, indicating that the prediction of activated positive aFRR is difficult if actual generation data is not available. The significance of actual generation from conventional sources aligns with the fact that positive aFRR is commonly activated from conventional power plants as these are easier to ramp up compared to e.g. variable renewable energy sources. However, in the application of volume and price prediction models by market participants, real-time actual generation data is not available, which means that model performance is likely to be weak for these applications.

3.1.2 Negative aFRR

The summarized results for the prediction of the activated negative aFRR volume is shown in Figure 4.

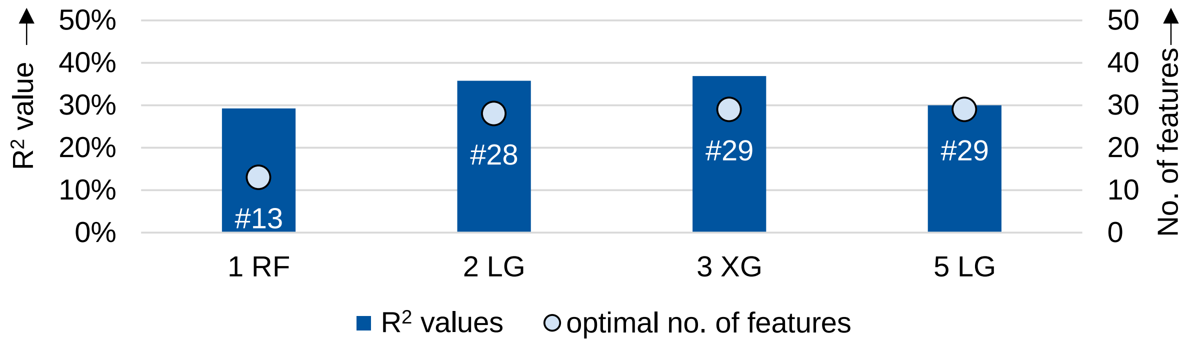

Figure 4: R^2 values and optimal number of features for each input data combination for negative aFRR volume

Generally, the results show that model performances for negative aFRR volume are worse than for positive aFRR volume, with none reaching an R^2 value of more than 36.9%. The third combination, in which skewed values above 1 or below -1 were transformed, achieved the best performance for this target variable. For this combination, the XGBoost model achieved an R^2 value of 36.9% with an optimal number of 29 features. Like the results for positive aFRR volume, the performance of the first combination with only forecasted values is worse than those with actual values.

The most important features for each input data combination for negative aFRR volume are shown in Table 3.

Table 3 Best performing method and most important features for each input data combination for negative aFRR volume

Combination	***Best method***	***Most important features***
1	RF	Generation forecast photovoltaics and wind, day ahead price DE/LU
2	LG	Residual load forecast error, actual consumption hydro pumped storage, capacity price of negative aFRR

3	XG	Actual consumption hydro pumped storage ramp, actual consumption hydro pumped storage, actual consumption residual load ramp
4	LG	Actual consumption hydro pumped storage ramp, residual load forecast error

For the target variable negative aFRR volume, forecasted generation of photovoltaics and wind is the most important feature for the first input data combination as well. As for positive aFRR volume, this indicates a correlation between the absolute forecasted generation from variable renewables and forecast errors which cause aFRR activation. For the other input data combinations, actual generation and consumption of hydro pumped storages still dominate as important features. Capacity price seems to influence this target variable as well.

3.2 Energy Price

3.2.1 Positive aFRR

The models for the prediction of positive aFRR energy prices generally achieved much better results than the ones for the activated volumes. The highest R^2 values and optimal number of features for each input data combination for the positive aFRR price are shown in Figure 5.

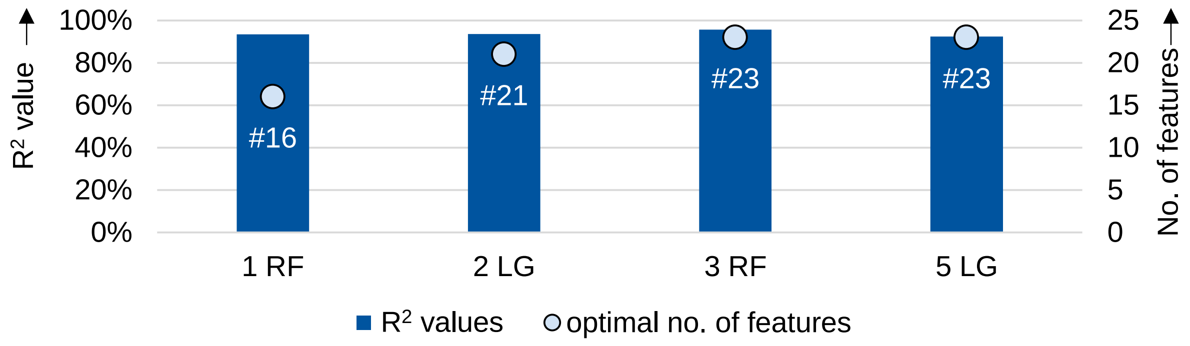

Figure 5: R^2 values and optimal number of features for each input data combination for positive aFRR price

For all four combinations, models achieved R^2 values of more than 90%. The best performance was reached by the third input data combination, with an R^2 value of 95.79% achieved by the Random Forest algorithm and an optimal number of 23 features.

Table 4 shows the most important features for each input data combination for positive aFRR prices.

Table 4 Best performing method and most important features for each input data combination for positive aFRR price

Combination	***Best method***	***Most important features***
1	RF	Capacity volume procured positive aFRR, month, capacity price of negative aFRR
2	LG	Residual load forecast error, actual consumption hydro pumped storage, capacity price of negative aFRR

3	RF	Actual generation other renewable, actual generation nuclear, month
4	LG	Actual generation nuclear, Capacity volume procured positive aFRR, capacity price of negative aFRR

To summarize, capacity prices of negative aFRR and the volume of positive aFRR capacity seem to dominate, as well as generation from conventional sources. The fact that negative aFRR capacity prices are an important feature seems counter-intuitive, but could mean that if the demand for negative aFRR capacity and thus its price is high, an excess supply in the grid is to be expected and therefore the need for positive aFRR is low. Out of all time-related features, only the month seems to have an effect on this target variable.

3.2.2 Negative aFRR

For negative aFRR prices, model performances were reasonably good. The summarized results are shown in Figure 6.

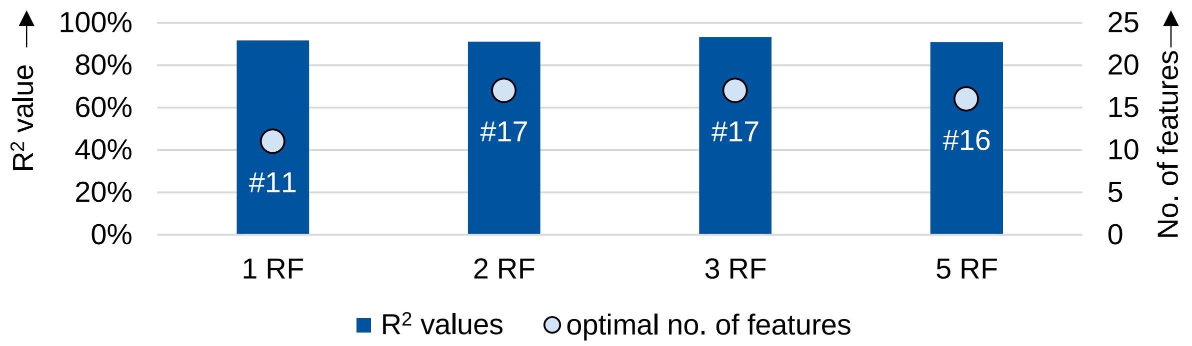

Figure 6: R^2 values and optimal number of features for each input data combination for negative aFRR price

Interestingly, the best performances were achieved by the Random Forest algorithm for all combinations. The third combination achieved the best model performance for this target variable with an R^2 value of 93.45% and an optimal number of features of 17.

The most important features for each input data combination for the prediction of the negative aFRR price are presented in Table 5.

Table 5 Best performing method and most important features for each input data combination for negative aFRR price

Combination	***Best method***	***Most important features***
1	RF	Capacity price of positive aFRR, generation forecast other, time of day
2	RF	Actual generation nuclear, actual generation other renewable, capacity price of positive aFRR, month
3	RF	Actual consumption hydro pumped storage ramp, residual load forecast error, actual generation hydro pumped storage

4	RF	Actual generation nuclear, actual generation other renewables, capacity price of positive aFRR

Although the general trend that hydropower-related features dominate persists, the most important features are quite different among the combinations. This is peculiar since all the best performances were achieved by the Random Forest algorithm. This points to the possibility that data transformations have an important function in determining which features are important. Furthermore, it is counter-intuitive that the capacity price of positive aFRR would influence the energy price of negative aFRR. Again, the reason could be that the model interprets the low capacity price of positive aFRR as a signal for the demand for negative aFRR that influences its energy price.

4 Conclusion

The goal of this paper was to update and expand the current state of research on the prediction of activated balancing energy volumes and prices. Four different machine learning methods were used with different input data combinations, varying the use of actual and/ or forecasted generation and consumption and the handling of skewness, to predict activated aFRR volumes and prices in Germany for a time period between May 2021 and April 2023.

For the activated volumes of positive and negative aFRR, model performances were generally not optimal. Although the entire analyzed time period was after the introduction of the balancing energy market in Germany, some market parameters were still changed during the analyzed period, which might complicate the prediction. Forecasted generation from variable renewable energy was the most important feature when no actual, real-time generation and consumption data was considered. When adding the actual generation and consumption data, model performances increased slightly, because activated aFRR is included in the actual generation and consumption data.

The prediction models performed better for balancing energy prices, reaching R^2 values up to 95,8%. Balancing capacity prices were important features to predict balancing energy prices. However, for the energy price of positive aFRR, the capacity price for negative aFRR was most important and vice versa. This could mean that the capacity price of the opposite direction reflects the need for the opposite product and thus decreases the price of the considered product. Future research should, however, investigate this in more detail.

No machine learning method performed best for all target variables and input data combinations. For each individual case, another method performed best. As for the handling of skewness, there is a certain threshold at which skewed values should be transformed, but it differs per method and input data combination.

Overall, the results show that a prediction of activated aFRR volume is difficult. This seems logical, because if the activated volume were predictable, the activation of aFRR could be prevented by trading on the intraday electricity market. The prediction of balancing energy prices reaches better results. However, if market participants were to use such a prediction to their advantage and in turn influence market outcome, it is not certain that predictions would stay reliable.

5 References

[1] European Commission, *Commission Regulation (EU) 2017/2195 of 23 November 2017 establishing a guideline on electricity balancing*, 2017.

[2] ENTSO-E, "Go-Live of PICASSO: the pan-European Platform for the International Coordination of Automatic Frequency Restoration Reserves and Stable System Operation," 2022.

[3] ACER, "Implementation framework for the European platform for the exchange of balancing energy from frequency restoration reserves with automatic activation," 2022.

[4] A. Lucas, K. Pegios, E. Kotsakis, and D. Clarke, "Price Forecasting for the Balancing Energy Market Using Machine-Learning Regression," *Energies*, 2020.

[5] J. Kruse, B. Schäfer, and D. Witthaut, "Secondary control activation analysed and predicted with explainable AI," 2022.

[6] S. Backe, S. Riemer-Sørensen, D. A. Bordvik, and S. Tiwari, "Predictions of prices and volumes in the Nordic balancing markets for electricity," 2023.

[7] Bundesnetzagentur. "Marktdaten." Accessed: Dec. 28, 2023. [Online]. Available: https://www.smard.de/home/downloadcenter/download-marktdaten/

[8] regelleistung.net. "Datencenter." Accessed: Dec. 28, 2023. [Online]. Available: https://www.regelleistung.net/apps/datacenter/tenders/?productTypes=SRL&markets=BALANCING_ENERGY&date=2023-12-28&tenderTab=SRL$ENERGY$0

[9] netztransparenz.de. "Aktivierte Regelleistung." Accessed: Dec. 28, 2023. [Online]. Available: https://www.netztransparenz.de/de-de/Regelenergie/Daten-Regelreserve/Aktivierte-Regelleistung

[10] S. Weisberg, "Yeo-Johnson Power Transformations," 2001.

[11] P. Gupta and N. K. Sehgal, *Introduction to Machine Learning in the Cloud with Python*, 2021.

[12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2008.

[13] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*, 2022.

[14] T. Agrawal, *Hyperparameter Optimization in Machine Learning*, 2021.

[15] 50Hertz, Amprion, TenneT, TransnetBW. "Präqualifizierte Leistung in Deutschland." Accessed: Dec. 29, 2023. [Online]. Available: https://www.regelleistung.net/xspproxy/api/staticfiles/regelleistung/pq-leistungindeutschland(stand01.01.2023).pdf