Estimating PV Curtailed Power as a Voltage Support Service using Data-Driven Approaches

Verhoeven, Gijs; Vergara , Pedro P.; Salazar Duque, Edgar Mauricio; Kok, Koen

DOI
10.1109/SEST53650.2022.9898486

Publication date
2022

Document Version
Final published version

Published in
Proceedings of the 2022 International Conference on Smart Energy Systems and Technologies (SEST)

Citation (APA)
Verhoeven, G., Vergara , P. P., Salazar Duque, E. M., & Kok, K. (2022). Estimating PV Curtailed Power as a Voltage Support Service using Data-Driven Approaches. In *Proceedings of the 2022 International Conference on Smart Energy Systems and Technologies (SEST)* (pp. 1-6). (SEST 2022 - 5th International Conference on Smart Energy Systems and Technologies). IEEE. https://doi.org/10.1109/SEST53650.2022.9898486

2022 International Conference on Smart Energy Systems and Technologies (SEST) | 978-1-6654-0557-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/SEST53650.2022.9898486

Estimating PV Curtailed Power as a Voltage Support Service using Data-Driven Approaches

Gijs Verhoeven[1], Pedro P Vergara[2], Edgar Mauricio Salazar Duque[1], Koen Kok[1]
[1] Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands
[2] Department of Electrical Sustainable Energy, Delft University of Technology, Delft, the Netherlands
g.verhoeven.1@tue.nl, p.p.vergarabarrios@tudelft.nl, e.m.salazar.duque@tue.nl, j.k.kok@tue.nl

***Abstract*—To guarantee a successful deployment of a droop-based control strategy to mitigate overvoltage problems caused by solar photovoltaic (PV) generation, Distribution System Operators (DSOs) will need to estimate the amount of active power curtailed by the PV inverters for billing purposes. This paper provides a structural elaboration on the development of data-driven approaches in Python to estimate the PV curtailed power as a provision of voltage support services by residential users using droop-based voltage control strategies. The use of the total input data, available for a DSO, would be impractical for an all-regression approach for the estimation of the PV curtailed power. Since in the majority of the data no active power is curtailed, the data-driven models would in this case partly be trained and fitted for situations where there is no active power curtailment. The regression models for the curtailed power prediction are therefore preceded by a classification model. The developed combined classification-regression model was able to estimate the PV curtailed power with an error of less than 4%, for test data from the network on which the model was trained.**

***Index Terms*—Low voltage distribution systems, overvoltage, droop control, PV curtailment, machine learning**

NOMENCLATURE

η	Explained variance	POC	Point of Connection
σ	Standard deviations model accuracy	PV	Solar photovoltaic
ACC	Model accuracy	$PV_{capacity}$	Capacity of the PV installation
APC	Active Power Control	RPC	Reactive Power Control
DSO	Distribution System Operator	T_a	Ambient temperature
G_T	Solar irradiance	V_{droop}	Voltage magnitude at the POC
$KNMI$	Royal Dutch Meteorological Institute	V_{thP}	APC voltage threshold
LV	Low Voltage	y	Actual output value
n	Number of data instances	$\hat{y}$	Predicted output value
p	Probability value	y_{error}	Difference between y and $\hat{y}$
$P_{curtail}$	Active power curtailment	z	Z-score
P_{net}	Net consumer power		

I. INTRODUCTION

A. Motivation and Background

The introduction of decentralized generation, such as solar photovoltaics (PV), in the low voltage (LV) distribution networks, has a considerable influence on voltage regulation. Most of the LV distribution networks today were not designed for the significant increase in reverse power flows due to PV production. During moments of low consumption and high PV generation, the power will not be used by the consumer, but will instead be fed into the network, possibly leading to overvoltage problems. The control of decentralized generation units seems to be the most promising technique

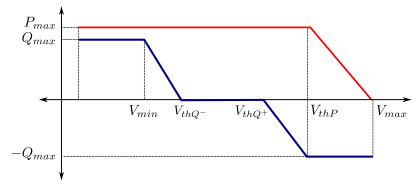

Fig. 1. Droop-based Active Power Control and Reactive Power Control strategies [3].

to deal with these challenges, due to the distributed and easy implementation features of overvoltage strategies which are locally implemented at the PV inverters [1]. Furthermore, since the control of the decentralized generation is only used when it is actually needed, when the voltage level in the network becomes too high, the amount of power that is regularly curtailed is minimized [2]. Dutch Distribution System Operators (DSOs) have started to develop and implement voltage control strategies based on well-known droop control, which could be installed locally at the PV inverters [1]. Standard droop control strategies run continuously using only local measurements, which is why Active Power Control (APC) and Reactive Power Control (RPC) strategies can be implemented in the control system of the PV inverter. Fig. 1 shows the droop-based APC and RPC strategies, where it can be seen that the active and reactive power of the PV installation are plotted as a function of the voltage magnitude at the point of connection (POC) with the network. For the APC, the active power output of the inverter will be set to the current maximum power point when the voltage at the POC is lower than the APC voltage threshold (V_{thP}). In the case of an overvoltage, the PV inverter will reduce the active power injection proportionally to the voltage increase at the POC.

Although in general both strategies are effective to mitigate voltage problems, this paper only focuses on APC. The APC strategy would be beneficial from a DSO perspective, as this strategy raises fewer concerns regarding the possible overloading of distribution transformers due to the reactive power consumption of the PV inverters within coordinated APC-RPC strategies. Although the droop-based control strategy is shown to be effective, the DSOs will be expected to pay a fee to the respective PV owner for the amount of active power that is curtailed by the PV inverter. Therefore, to guarantee the successful deployment of such a control strategy, the DSOs

978-1-6654-0557-7/22/$31.00 ©2022 IEEE

will need to estimate the amount of active power curtailed by the PV inverters for billing purposes. This estimation must be based only on operational data that the DSO has access to, as well as other open-source data, such as weather measurements.

B. Relevant Literature

Within the literature, numerous studies elaborate on the implementation and modeling of PV curtailment and the effects on distribution networks problems [1], [3]–[5]. However, only a few studies focus specifically on the data-driven estimation of PV curtailment approaches [5]–[7]. Within these studies, the PV curtailment is estimated based on only part of the available data (e.g. only smart meter voltage data and the rated power of the PV installation) and/or based on operational data for the PV generation, which is more detailed than a DSO would normally have direct access to. Additionally, these studies only evaluate a limited selection of data-driven approaches based on curve fitting, whereas this study will evaluate on more data-driven machine learning approaches as well as the importance of the different types of available data.

C. Contributions and Organization

Within this paper, a structural elaboration is given on the development of multiple data-driven approaches to estimate the PV curtailed power as a provision of voltage support services by residential users using droop-based voltage control strategies. This work will contribute to the estimation of PV curtailed power purely based on operational data, as well as other open-source data such as weather measurements that a DSO would have access to. Additionally, the developed data-driven models are evaluated and compared using different performance measures and an evaluation of the importance and selection of the available features is given. Finally, a conclusion will be given on the best-performing developed model to estimate curtailed PV power.

II. Methodology

The intended data-driven tools could be used by DSOs to estimate the amount of active power curtailed by the PV inverters for billing purposes. In this sense, this estimation must be based on operational data that the DSO already has access to, such as voltage magnitude at the POC (V_{droop}), the net consumer power (P_{net}) and the capacity of the PV installation ($PV_{capacity}$), as well as open-source data, such as weather measurements for solar irradiance (G_T) and temperature (T_a) from the Royal Dutch Meteorological Institute (KNMI). Within this study, the operational data was obtained from multiple power flow simulations. To realistically model the effects of droop control on voltage problems in LV networks, these power flow simulations are performed using a network model representative for a typical Dutch LV network adapted from [8]. For the initial simulation, the uncurtailed PV production data is used together with the consumer load data obtained from anonymized smart meter data. Whenever the resulting bus voltage for a certain time-step is above the voltage threshold ($V_{thP} = 1.08$ [pu]), the PV production for this time step for the corresponding bus is recalculated for the second power flow simulation. The outcome of this second simulation is representative of the situation in which APC droop control is applied to prevent or mitigate voltage magnitude problems. To analyze the applicability of the machine learning approaches, it is specified that the approaches must be applicable with a margin of error of no greater than 5% and a fair and balanced distribution of the error between the consumer and DSO. The machine learning models are developed based on three techniques, namely: linear regression models, gradient boosted trees models, and neural network models.

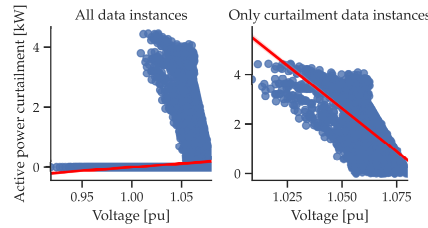

Fig. 2. Operational voltage and active power curtailment output data shown for all data instances and only data instances where active power is curtailed.

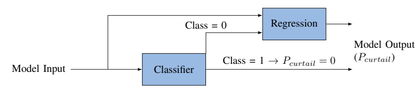

Fig. 3. Flow of data for the machine learning models approach.

As shown in Fig. 2, the use of all available input data would be impractical for an all-regression approach, since the relationships between the estimated active power curtailment ($P_{curtail}$) and the input variables, in this case, would partly be fitted for situations where there is no active power curtailment ($P_{curtail} = 0$). The regression models in the different approaches were therefore preceded by a classifier model, as shown in Fig. 3. Within the linear regression model techniques, for classification, this paper elaborates on a logistic regression model. For regression, this paper elaborates on a linear regression model as well as a polynomial regression model. The linear regression models were trained, optimized, and tested using the Scikit-learn package in Python [9]. Within the gradient boosted trees models, this paper elaborates on both a gradient boosted trees classification and regression model. The gradient boosted trees models were trained, optimized, and tested using the XGBoost package in Python [10]. Regarding the neural network models, this paper will elaborate on both a deep neural network classification and regression model. The neural network models were trained, optimized, and tested using the Tensorflow Keras package in Python [11].

A. Data preparation

Within the different machine learning approaches, the available data set is prepared using the following steps:

1) Splitting the data: To evaluate how well a model performs on new data instances, the total data set is split into two parts: the training set and the test set. The models are first trained using the training set of the data, after which the test set is used to evaluate the performance. The instances which appear in the different sets are picked randomly so that the individual sets will still resemble the full variety of the data, even when the total data set is sorted. To ensure that both data sets contain an equal proportion of the data to represent the different classes, the ratio of the different classes and the ratio of the different capacities of the PV installations are stratified within the random splitting of the data in a 80% training set and a 20% test set.

2) Feature scaling: Due to the different units of the input features, the magnitude of the data from the different features can differ significantly. Most machine learning algorithms do not perform well when using input features on different scales. Therefore, for the linear regression models and the neural network models, standardization is used to scale the input features and new input data. No scaler is used for the XGBoost models since the predictions in these models are based on splitting of the data in decision thresholds where the performance is not affected by the scaling of the data.

B. Model optimization

Within this study, different optimization approaches are used in the development of the machine learning models:

1) Hyperparameter tuning: The optimization of the machine learning models is done using hyperparameter tuning. In this process, a randomized search approach is used as a pre-selection of the best performing values for each of the concerning hyperparameters. Based on a parameter grid, containing the hyperparameters and the values to be evaluated, a grid search is used to evaluate all possible combinations using cross-validation to find the best performing combination of hyperparameter settings for the concerning model.

2) Balancing the training data: For the classification models, an additional optimization approach is used to balance the training data, since there are more instances where no active power was curtailed compared to instances with active power curtailment in the total data set. Within this optimization, the aim is to create a more balanced performance of the classification models, balancing the burden for both the DSO and the consumer for the instances of misclassification. This balancing approach is analyzed using undersampling, oversampling, combined under- and oversampling, and weight scaling techniques, from which the best performing technique is used for the individual classification models.

3) Feature importance and selection: For the initial models, all available features for a DSO are included as input data for the machine learning algorithms. However, from a practical, computational, and data acquisition point of view, it may be beneficial to have a model with the least number of data features necessary. Additionally, models which are fitted to the input data, like linear regression models, will include every parameter in the prediction of the output of the model which can reduce the overall effectiveness and accuracy of the models [12]. Therefore the importance of the input features is evaluated for each of the machine learning models, followed by a feature selection approach to give more insight into the performance of the different models and their dependence on the different input features. For the linear regression and logistic regression models, the importance of the input features is analyzed from their corresponding fitted coefficients. However, the feature importance for the polynomial regression and neural network models can not directly be derived. For these models, the feature importance is determined using a permutation importance approach. In this approach, one of the input features is sequentially shuffled in the total data set, keeping the other input features unaffected. The shuffling of an important feature will result in a significant decrease in the models' performance, whereas the shuffling of a relatively unimportant feature will result in less decrease in the model performance [13]. Within the XGBoost models, the prediction is obtained using data split decisions, the estimates of feature importance can therefore automatically be analyzed using the XGBoost package library [10].

C. Performance measures

The performance of a classifier model can be evaluated using a confusion matrix. From the confusion matrix, the precision and the recall metrics can be determined. To further clarify the naming and meaning of the different predictions in the confusion matrices, the instances of true negative resemble the part of the data where power curtailment takes place which is correctly classified by the classifier model. The instances of true positives resemble the part of the data where no power curtailment takes place which is correctly classified by the classifier model. The false positive instances resemble the part of the data where power curtailment takes place, however, the classifier model classifies these instances as if no power is curtailed. On a more practical approach, this means that consumers will have curtailed active power, however, the model used by the DSO will not predict this and therefore the DSO will not compensate for this amount of curtailed power. The burden of the false positive instances is therefore on the consumers. The false-negative instances resemble the part of the data where no power curtailment takes place, however, the classifier model classifies these instances as if power is curtailed. On a more practical approach, this means that the DSO, based on the prediction of the classifier, will compensate the consumer for power that has not been curtailed. The burden of the false-negative instances is therefore for the DSO. The accuracy for the regression models is calculated based on the so-called explained variance, using Equation 1. In this equation, y_{error} is the difference between the true output value y and the predicted output value $\hat{y}$, $\bar{y}_{error}$ and $\bar{y}$ represent the average of the respective values and n represents the number of instances. Since the regression model is only trained on the part of the training set which is labeled for active power curtailment, it is also only evaluated using the part of the test set labeled for active power curtailment.

$$\eta = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(y_{error} - \bar{y}_{error})^2}{\frac{1}{n}\sum_{i=1}^{n}(y - \bar{y})^2} \cdot 100\% \tag{1}$$

D. Model selection and statistical analysis

To provide a significant conclusion with regard to the best performing model, in addition to the test set performance, a so-called z-score test is performed. The z-score test first formulates a hypothesis, which in this case is chosen to resemble the situation where the performance of two models is the same. This z-score test will thereafter determine if the 95% confidence intervals of the accuracy of the two models are not overlapping, either confirming or rejecting the hypothesis that the performance of both models is equal at a significance level of 5%. The z-score for the different comparisons of the models is calculated using Equation 2, in which ACC_1 and ACC_2 resemble the model accuracies and σ_1 and σ_2 resemble the standard deviations of the model accuracy's [14]. Using this z-score as a threshold, the area under the standard normal cumulative distribution is computed to obtain the probability p-value. If this p-value is smaller than the significance level of 5%, the hypothesis that the performance of the two models is the same can be rejected, meaning that the performance of the model with the highest accuracy is also significantly better [14].

$$z = \frac{ACC_1 - ACC_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{2}$$

The accuracy and standard deviation of the different models are obtained using k-fold cross-validation. The idea of cross-validation is not to evaluate the model on a pre-defined test set, but to create the opportunity to test the model on all parts of the data set. To do this, the data set is split up into k parts, one of the parts is used for validation, the validation fold, and the other $k-1$ parts are combined into a training subset for the evaluation of the model [14]. Within this study, the data set is split up in "$k = 10$" parts. The cross-validation results give an estimate of the performance of the machine learning models with less deviation compared to a single test set performance and are said to be more reliable to estimate the performance of the algorithms on new unseen data [15].

III. Results

A. Test set performance

Fig. 4 shows the confusion matrices for the performance of the different classification models on the test set data. Table I shows the corresponding accuracy, precision, and recall scores for the class predictions of the classification models. It can be seen that the XGBoost model leads to the best performance. In total this XGBoost model leads to the lowest overall instances of misclassification, whilst also leading to the lowest amount of instances of misclassification of an instance of active power curtailment as if no active power is curtailed.

Table II shows the accuracy of the curtailed power estimation of the different regression models on the data in the test set labelled for regression. It can be seen that the neural network model performs slightly better than the XGBoost model, whilst both these models outperform the linear and polynomial regression models.

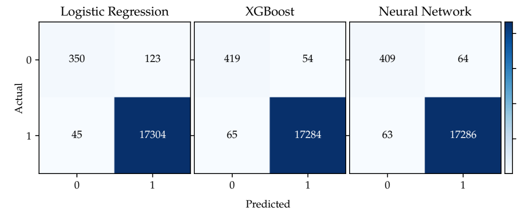

Fig. 4. Confusion matrices of the classification models on the test set data.

TABLE I
Performance of classification models on the test set data

	Logistic Regression	XGBoost	Neural Network
Accuracy	0.9906	0.9933	0.9929
Precision	0.9929	0.9969	0.9963
Recall	0.9975	0.9963	0.9964

TABLE II
Performance of regression models on the test set data

	Linear Regression	Polynomial Regression	XGBoost	Neural Network
Accuracy	95.42 %	95.75 %	98.14 %	98.32%

TABLE III
Performance of the combined classification-regression models on the test set data

	Linear Regression	Polynomial Regression	Gradient Boosted Trees	Neural Network
Accuracy total model	91.82%	92.35%	96.26%	95.71%
Accuracy regression incl. wrong classified	77.59%	79.04%	90.55%	89.29%
Accuracy regression only right classified	91.11%	92.41%	97.75%	98.01%

However, due to the classification error, the regression model will also be used in instances where no active power is curtailed. Table III therefore shows the total combined classification-regression model performance which consists of the three different accuracy values for the curtailed power estimation for the total model, namely: the accuracy for the total combined classification-regression models, the accuracy of the regression models including the misclassified instances and the accuracy of the regression models including only the correctly classified instances. A visualization of the combined classification-regression model performance showing the true (blue) and predicted (red) values of the active power curtailment is depicted in Fig. 5. It can be seen that the XGBoost model performs slightly better than the neural network model, whilst both these models outperform the linear and polynomial regression models on the accuracy of the total model. When only looking at the instances which were classified for the regression model, including wrongly classified instances, the XGBoost model still slightly outperforms the neural network model. Regarding the instances which were correctly classified for the regression model, it can be seen that similar to individual regression performances, the neural network model performs slightly better than the XGBoost model.

B. Model selection and statistical analysis

Since the individual performance of the models on the defined test set does not provide a significant and unambiguous conclusion concerning the best performing model, a so-called z-score test has been performed. The accuracy and standard

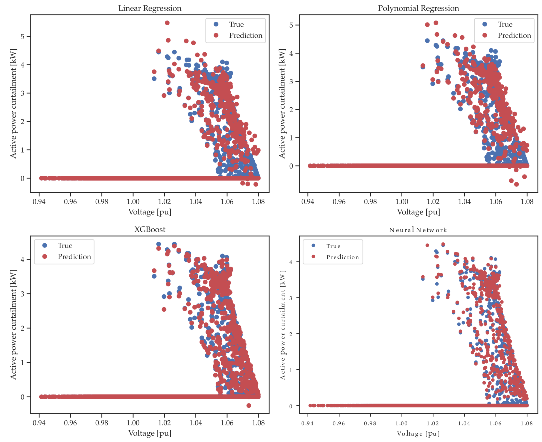

Fig. 5. Performance visualisation showing the true (blue) and predicted (red) values of the active power curtailment for the different combined classification-regression models.

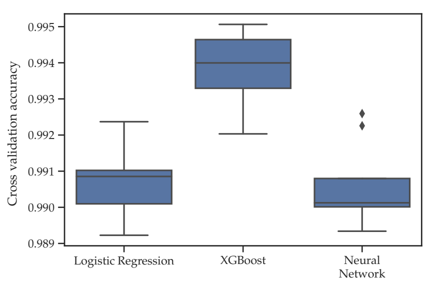

Fig. 6. Boxplot for the cross-validation performances of the different classification models.

deviation of the different models are obtained using the 10-fold cross-validation performances, which would give a better estimate of the performance of the machine learning models compared to the previous shown test set performances. Fig. 6 shows a boxplot for the cross-validation performances of the different classification models. It can be seen that the mean accuracy of the XGBoost model is higher than the mean accuracy of the logistic regression and neural network model. Evaluating the resulting probability p-values for the z-score test, it can be stated that the probability of obtaining these results, assuming the hypothesis where the performance of the logistic regression model and the neural network model is the same as that of the XGBoost model, is 1.7% and 2.4 % respectively. Since the obtained p-values are both well below the predefined significance level of 5%, it can therefore be stated that the XGBoost classification model performs significantly better than the other discussed models.

Fig. 7 shows a boxplot for the cross-validation performances of the different regression models. It can be seen that the mean accuracy of the XGBoost model is higher than the mean accuracy of the linear regression, polynomial regression, and neural network models. Evaluating the resulting probability p-values for the z-score test, it can be stated that the probability of obtaining these results, assuming the hypothesis where the

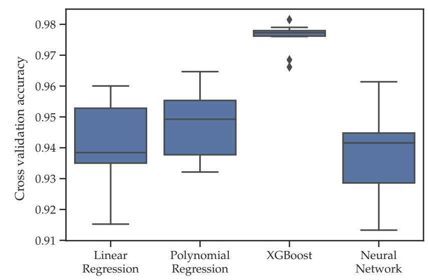

Fig. 7. Boxplot for the cross-validation performances of the different regression models.

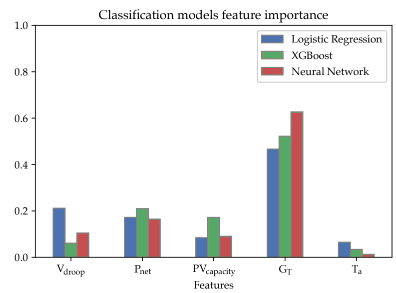

Fig. 8. Feature importance for the different classification models.

performance of the linear regression model, the polynomial regression model, and the neural network model is the same as that of the XGBoost model, is 1.4 %, 1.5 %, and 0.7% respectively. Since this is all well below the predefined significance level of 5%, it can be stated that the XGBoost regression model performs significantly better than the other discussed models.

C. Feature importance and selection

Fig. 8 shows the relative feature importance for the different classification models. It can be seen that for all classification models, the global irradiance, G_T, is the most important feature for the class prediction. Fig. 9 shows the cross-validation performance of the classifier models for the different feature selections, based on the relative feature importance. It can be seen that the performance of the model decreases with the decreasing number of selected features. However, the performance of the XGBoost and neural network classifier models remains quite stable when only the two most important features are included in the selection.

Fig. 10 shows the relative feature importance for the different regression models. It can be seen that for all regression models, the net consumption of active power by the consumer, P_{net}, is significantly the most important feature for the prediction. Fig. 11 shows the cross-validation performance of the regression models for the different feature selections. Similar to the classification models, it can be seen that the performance of the models decreases with the decreasing number of selected features. However, apart from the neural network model, the performance of the regression models is quite stable, even with a decreasing number of features.

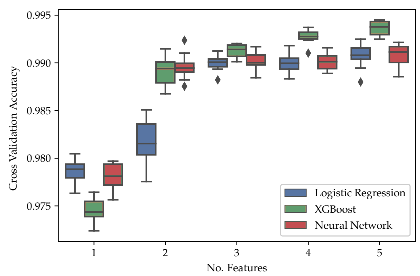

Fig. 9. Cross-validation performance for feature selection of the different classification models.

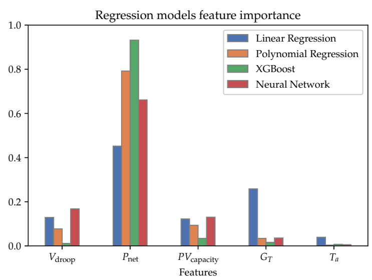

Fig. 10. Feature importance for the different regression models.

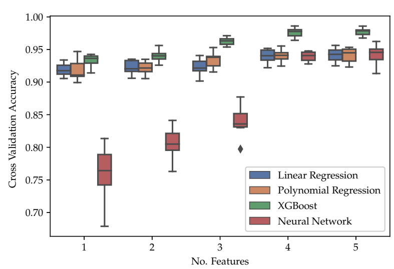

Fig. 11. Cross-validation performance for feature selection of the different regression models.

The required parameters for the modeling, training, and optimization of the data-driven approaches, and the importance of each feature depends on the respective data-driven approach. However, it can be stated that if the ambient temperature data, T_a, is not available, the overall decrease in performance of all the models is minimal and remains under 0.3%. This indicates that even if this data would be missing or unavailable at times, the DSO would still be able to implement the different models to estimate the amount of curtailed PV power without decreasing the accuracy significantly. Nonetheless, a combination of all discussed parameters (V_{droop}, P_{net}, $PV_{capacity}$, G_T and T_a) results in the highest performance for all data-driven approaches.

IV. Conclusions

Regarding the training and optimization of the data-driven approaches, the study shows that the use of the total input data available for a DSO would be impractical for an all-regression approach for the estimation of the PV curtailed power. Since in the majority of the data no active power is curtailed, the data-driven models would in this case partly be trained and fitted for situations where there is no active power curtailment. The regression models for the curtailed power prediction are therefore preceded by a classification model. Regarding the classification models, a combination of a confusion matrix together with the precision, recall, and accuracy metrics, as well as the z-score model selection on the cross-validation performance, results in an unambiguous conclusion on the best performing classification approach. For the regression models, the conclusion on the best performing data-driven approach on the used data set is based on the test set performance, as well as the z-score model selection on the cross-validation performance. Based on these performance measures, it can be concluded that, from the developed models, the choice for a combined classification-regression gradient boosted trees approach, as used in the XGBoost models, for the estimation of curtailed PV power is sufficiently substantiated for data of the network on which the model is trained. This approach meets the specified requirements with an error of less than 4% and is shown to perform significantly better than the other options while looking at the cross-validation performance.

References

[1] P.P. Vergara *et al.*, "Feasibility and Performance Assessment of Commercial PV Inverters Operating with Droop Control for Providing Voltage Support Services," *Proc. of IEEE PES ISGT-Europe 2019*, pp. 2–6.

[2] R. Tonkoski, L. A. C. Lopes, and T. H. M. El-Fouly, "Coordinated active power curtailment of grid connected pv inverters for overvoltage prevention," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 2, pp. 139–147, 2011.

[3] P.P. Vergara *et al.*, "A comprehensive assessment of PV inverters operating with droop control for overvoltage mitigation in LV distribution networks," *Renewable Energy*, vol. 159, pp. 172–183, 2020.

[4] J. Seuss *et al*, "Advanced inverter controls to dispatch distributed pv systems," in *2016 IEEE 43rd PVSC*, 2016, pp. 1387–1392.

[5] A. Latif, W. Gawlik, and P. Palensky, "Quantification and mitigation of unfairness in active power curtailment of rooftop photovoltaic systems using sensitivity based coordinated control," *Energies*, vol. 9, 6 2016.

[6] M. Emmanuel *et al*, "Estimation of solar photovoltaic energy curtailment due to volt-watt control," *IET Renewable Power Generation*, vol. 14, pp. 640–646, 3 2020.

[7] N. Stringer *et al*, "Fair consumer outcomes in the balance: Data driven analysis of distributed pv curtailment," *Renewable Energy*, vol. 173, pp. 972–986, 8 2021.

[8] G. Verhoeven, P. P. Vergara, and K. Kok, "Assessing the operation of typical dutch distribution systems with large penetration of low-carbon energy technologies," *Proc. of CIRED 2021*, pp. 2114–2118.

[9] F. Pedregosa *et al*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[10] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proc. of the ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, vol. 13-17-August-2016, pp. 785–794, 2016.

[11] F. Chollet *et al.* (2015) Keras. [Online]. Available: https://github.com/fchollet/keras

[12] M. Kuhn and K. Johnson, *Applied predictive modeling*, 5th ed. Springer, 2013.

[13] M. Cerliani, "Feature Importance with Neural Network," 2019. [Online]. Available: https://towardsdatascience.com/feature-importance-with-neural-network-346eb6205743

[14] S. Raschka, "Model evaluation, model selection, and algorithm selection in machine learning," *arXiv*, 2018.

[15] J. Brownlee, *Gradient Boosted Trees with XGBoost and scikit-learn*. Machine Learning Mastery, 2018.