2ND WORKSHOP “MACHINE LEARNING & NETWORKING” (MaLeNe) PROCEEDINGS

SEPTEMBER 4, 2023

CO-LOCATED WITH
THE 5TH INTERNATIONAL CONFERENCE ON NETWORKED SYSTEMS (NETSYS 2023)
POTSDAM, GERMANY

Replicable Machine Learning Workflow for Energy Forecasting

Stepan Gagin
Chair of Computer Networks and Computer Communication
University of Passau
Passau, Germany
stepan.gagin@uni-passau.de

Adrian Carrasco-Revilla
Data Scientis Senior (Innovation Department)
Inetum
Madrid, Spain
adrian.carrasco@inetum.com

Hermann de Meer
Chair of Computer Networks and Computer Communication
University of Passau
Passau, Germany
hermann.demeer@uni-passau.de

Nuria Sánchez Almodóvar
Head of Innovation Department
Inetum
Madrid, Spain
nuria.sanchez@inetum.com

***Abstract*—Energy forecasting is an essential functionality of energy management systems. In this paper, a machine learning workflow for energy forecasting is proposed. Integration of such forecasting service does not require preliminary knowledge in a machine learning domain. Moreover, the workflow is designed to be replicable. That means that it can be used for the forecasting of different measurements in the energy domain in multiple locations.**

***Index Terms*—machine learning, energy systems, replicability.**

I. Introduction

Energy forecasting is an essential functionality for the energy sector [1]. Many energy management systems (EMSs) use forecasting to control energy equipment and send recommendations to users and other applications to manage their energy consumption. However, a development and training of forecasting models as well as processing data from sensors require familiarity with machine learning (ML) concepts, while energy managers often lack experience in that domain.

The replicability of software systems used in research [2] is not always ensured in the ML domain. To mitigate that, in this paper the replicable ML workflow is proposed. A replicability of the workflow means, that it can be used without additional changes in different locations for time series forecasting in the energy sector. This workflow can be integrated into existing EMSs and does not require the involvement of a ML specialist. For an integration, only historic time series are necessary, however, exogenous variables can be accepted to improve the predictions.

II. Related Work

The concept of automated ML is becoming popular [3]. It aims to automate the training of multiple ML models and chooses the one that is performing best. ML can be encapsulated as a reusable software service [4]. However, the approach proposed in this paper focuses on the forecasting in energy sector, for instance on photovoltaic (PV) generation, building energy consumption in heating and electricity domains. That means that the models can be trained with specific types of measurements and interconnection between them in mind, which improves forecasting performance.

III. Methodology

The replicable workflow introduced in this paper can be connected to the centralized data storage of an EMS as shown in Figure 1. The other EMS services can communicate with the storage too. There could be multiple copies of the workflow running in parallel in the system for different types of forecast data series, for instance, PV generation, electricity consumption of a building or heating consumption of a flat in a building. Moreover, the same workflow can be used for different buildings in multiple locations.

Replicability of a workflow means that it can be reused for forecasting of different time series without any internal changes. An interface between data storage and forecasting workflow consists of three elements:

- measurements - historic data from a sensor with a given timestamp for each measurement;
- weather data - historic data connected to the weather, such as temperature, relative humidity, wind speed and solar radiation;
- metadata - additional data about the type of sensor, physical quantity and units used for a measurement.

That allows for easy usage of this workflow by non-ML specialists because it does not require any adjustments of the parameters of underlying models.

The workflow can be implemented as a separate software service that communicates with other services using REST API. Inputs and outputs of the workflow can be represented

This paper has received funding by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 957845: "Community-empowered Sustainable Multi-Vector Energy Islands - RENergetic".

as JSON objects with predefined values for attributes to model data and metadata.

Subsections III-A-III-E provide an overview of each step of the proposed workflow.

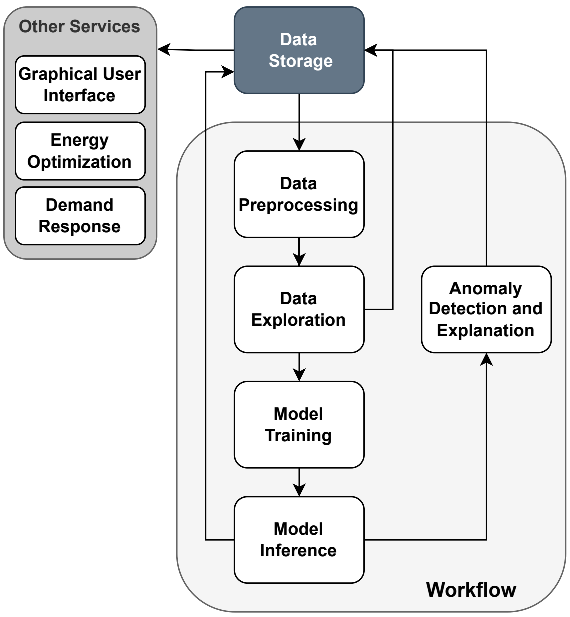

Fig. 1. Forecasting workflow

A. Data Prepossessing

Data collected from sensors and equipment can contain missing data points, thus they should be estimated, for instance, using the interpolation methods. Additionally, the data used in forecasting models should have a constant frequency of data points. However, data from sensors can arrive at the data storage at different time. Thus, data should be transformed and aggregated. Aggregation should be done considering the type of measurements. For instance, PV panels report about their generation power, although for energy optimization the data about energy generated are needed.

B. Data Exploration

In this step, statistics and other information about time series are collected. These can include a median, average and standard deviation of time series over a defined period. Additionally, the identification of patterns in time series is done. The identified shapelets [5] are used in the following steps by forecasting models. Anomaly detection is also performed at this step to discover and handle outliers in the data set before training. The output can be used by the forecasting models or by other services in the system, for instance, to display statistics in a graphical user interface.

C. Model Training

The workflow assumes that the model was already created. This step includes training of pre-defined models or their versions with different hyperparameters that can be used for the forecasting. It is important to note that training of the model on new data is performed, only if needed. It is decided based on the performance metrics of previously predicted data points. If performance is higher than set threshold, the re-training is not needed. If re-training is needed, several models are trained in parallel and the best-performing model is chosen. Such mechanism allows saving computational resources spent on the model training.

The models could be trained in competition (only one is saved) or in cooperation, meaning the final process combines all the results into a final model which is trained to take the benefits of each model and discard the weak spots.

D. Model Inference

In this step, the model resulting from the previous step makes predictions for the next time window, for instance, for 24 hours, 48 or an interval provided by the user.

E. Anomaly Detection and Explanation

Using the same processes introduced in Subsection III-B, the output of forecasting models can be analysed, for instance, allowing quick actions on results, such as notifications when high values are forecasted. Additionally, some models support the explainability of results - from feature importance they provide information on which elements of input data have the biggest influence on the output to seasonality disambiguation, which shows which of the time components impacts the forecast.

IV. Conclusion

In this paper, a ML workflow for energy forecasting is proposed. This workflow includes several steps that enable an integration of energy forecasting functionalities into EMSs without expertise in the ML domain. Replicability of the workflow means, that it can be used for energy forecasting in different locations. In an extended version of this paper, the models that can be used in such workflow are developed and evaluated on real-world data.

References

[1] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy Forecasting: A Review and Outlook," vol. 7, pp. 376–388, 2020, publisher: Institute of Electrical and Electronics Engineers Inc.

[2] Association for Computing Machinery. [Online]. Available: https://www.acm.org/publications/policies/artifact-review-and-badging-currentappendix

[3] X. He, K. Zhao, and X. Chu, "AutoML: A survey of the state-of-the-art," *Knowledge-Based Systems*, vol. 212, p. 106622, Jan. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0950705120307516

[4] M.-O. Pahl and M. Loipfinger, "Machine learning as a reusable microservice," in *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*. Taipei: IEEE, Apr. 2018, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/8406165/

[5] J. Grabocka, N. Schilling, M. Wistuba, and L. Schmidt-Thieme, "Learning time-series shapelets," ser. KDD '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 392–401. [Online]. Available: https://doi.org/10.1145/2623330.2623613