Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 375–384 DOI: 10.15439/2024F2959 ISSN 2300-5963 ACSIS, Vol. 39

Learning from the COVID-19 Pandemic to Improve Critical Infrastructure Resilience using Temporal Fusion Transformers

Jakob Jenko
0009-0007-8655-4838
XLAB
Ljubljana, Slovenia

Joao Pita Costa*
0000-0001-5745-1302
XLAB
Ljubljana, Slovenia
joao.pitacosta@xlab.si

Daniel Vladušič
0000-0001-6032-8575
XLAB
Ljubljana, Slovenia

Urban Bavčar
0009-0009-4046-0312
ELES
Ljubljana, Slovenia

Radoš Šabarkapa
0009-0007-5202-4572
EKC
Beograd, Serbia

Abstract—**During the COVID-19 pandemic, traditional demand prediction models drastically failed mostly due to altered consumption patterns. Accurate forecasts are essential for ensuring grid stability.**

This paper analyzes the performance of the Temporal Fusion Transformer (TFT) model during the COVID-19 pandemic aiming to build resilient demand prediction models. Through detailed analysis, we identify which features may contribute to improved performance during large-scale events such as pandemics. During lockdowns, consumption patterns change significantly, leading to substantial errors in existing demand prediction models.

We explore the impact of features such as mobility and special day considerations (e.g., lockdown days) on enhancing model performance. We demonstrate that periodic updates on a monthly basis make the model more resilient to changes in consumption patterns during future pandemics.

Moreover, we show how improvements in prediction accuracy translate to real-world benefits, such as enhanced grid stability and economic advantages, including reduced energy waste. Additionally, we discuss the implications for energy-critical infrastructure, considering disruptive scenarios like future pandemics.

I. Introduction

As the integration of renewable energy sources into power grids intensifies, and the accuracy of energy demand predictions becomes increasingly crucial. Effective energy management requires that energy production aligns closely with demand to minimize the losses often associated with overproduction. Therefore, reliable and precise demand forecasts are essential for optimally adjusting production levels.

Moreover, extreme situations like global pandemics can drastically change consumption patterns overnight, underscoring the importance of having adaptable and responsive demand prediction models. These models must quickly incorporate new data and adjust to shifting consumption dynamics to ensure energy efficiency and grid stability. By doing so, they help maintain a balance between production and demand, preventing inefficiencies and promoting sustainable energy use. Moreover, the discussion on the trustworthiness and certification of AI systems and, in particular, of neural networks as in [1] is essential to move forward with a fast changing technological landscape.

Maintaining the alignment between production and demand is critical for grid stability, as deviations can lead to significant issues, including a drop in the grid frequency below 50 Hz, potentially causing grid collapse or separation. In scenarios where demand exceeds supply, gas-powered peaker plants are typically utilized to provide the necessary additional capacity. However, some of this demand can also be mitigated through the use of pumped-storage hydroelectricity, which contributes stored, often renewable, energy back into the grid. In situations where it is not feasible to meet high demand exclusively through increased production, load shedding [2] is implemented as a controlled process to prevent total grid failure. This involves selectively disconnecting parts of the grid—such as entire neighbourhoods—to reduce the overall electrical load, ensuring that the grid does not exceed its capacity.

Accurate forecasting models are indispensable to grid management, particularly in anticipating and responding to demand surges. This capability becomes even more crucial during unforeseen critical events, such as pandemics, which can abruptly and drastically alter usage patterns. Effective models must rapidly adapt to new consumption patterns, providing timely forecasts that reflect current consumption trends to maintain grid stability. The pertinence of machine learning methods in the study of energy efficiency in particular pandemic scenarios gains much from the data collected in the most recent COVID-19 global pandemic. The intersection of energy efficiency and artificial intelligence (AI) has gained unprecedented significance, as the crisis reshaped global energy consumption patterns and highlighted the urgency of sustainable practices. The most recent machine learning methods emerged as key enablers in adapting to these changes.

This paper discusses how AI-driven solutions can be instrumental in optimizing energy use during the pandemic, ensuring efficient operations while addressing the environmental challenges exacerbated by the health crisis. By examining AI's role in mitigating energy consumption in a time of fluctuating demand and promoting sustainable practices in the face of adversity, this analysis illuminates the critical role of technological innovation in navigating the energy challenges posed by COVID-19 taking into consideration mobility. We have a closer look at the mobility data in the context of a pandemic, based on the data collected during the COVID-19 incidence between 2020 and 2022. Particularly, we look at the number of unique connections to the cell tower in Slovenia measuring how much people migrate. The consumption curve can describe the behaviour of people in regard to mobility. Taking into account that Telecom data is usually expensive, we compare the relevance of that data for the forecasting model in relation to the usage of labels for special days (e.g., considering lockdown days as holidays).

The research question addressed in this paper regards if the input of mobility data is comparable to the input of special days, particularly during a pandemic scenario learning from the experience (and data) from the COVID-19 pandemic. In particular,

- Can we improve model accuracy and performance during an impactful large-scale event (such as a pandemic) with additional features (e.g. mobility and other special day features)?
- Can we make the model more resilient by periodically updating on a monthly basis
- How do the improvements in prediction accuracy reflect in the real world? Benefits to grid stability, benefits to economic aspects as less energy is wasted etc.
- What should energy Critical Infrastructure (CI) take into account in disruptive scenarios like future pandemics?

The main contribution of this paper is a new methodology based on the Temporal Fusion Transformer (TFT), and its initial evaluation, which shows how past energy consumption, weather forecast and energy-saving features can impact the prediction of energy consumption.

Results will be presented in Section III-A, where we demonstrate the model's performance in predicting Serbian national consumption. This analysis will validate and illustrate how the initial version of our model performs against the established EKC model. Following this comparison, we will build upon this model in subsequent subsections, focusing on data from Slovenia.

The COVID-19 pandemic has significantly altered consumption patterns due to increased home stays, underscoring the need to integrate mobility data into forecasting models for enhanced accuracy. This approach is supported by [3], demonstrating the effectiveness of incorporating mobility features from publicly available Google data into their predictive models, significantly improving forecast precision. While authors in [3] were able to demonstrate this phenomenon across multiple states and continents using mobility data for the US and EU, our study is limited to Slovenia due to data constraints. However, we anticipate that this phenomenon will be applicable to other cultures and states, as observed in the cited paper. In line with these findings, our methodology involves deriving a "mobility factor" from data provided by a Slovenian national telecommunication provider. By analyzing the number of unique connections to cell towers, we can infer mobility patterns: fewer cell connections typically indicate that residents are staying home while connecting to multiple cells suggests movement to different locations. The total number of unique connections across all cells in a given area reflects the overall mobility, serving as a valuable predictor in our models.

We enhance model performance during periods of rapid consumption changes (e.g., lockdowns) by incorporating a "special day flag." This flag is activated on weekends, holidays, or days with enforced curfews/lockdowns. The advantage of this flag lies in its simplicity and availability for day-ahead forecasting scenarios, providing a straightforward yet effective method to account for unusual consumption patterns. This paper discusses the implemented state-of-the-art deep learning models, with particular focus on the TFT approach [4]. Moreover, we build on decision tree-based models such as XGBoost [5] and CatBoost [6]. Linear regression was used as a baseline, and the energy data was sourced from ELES and EKC, represented in this paper by the respective coauthors. The models are further refined by including mobility data as an additional input, which is expected to bolster their accuracy, particularly during periods like the pandemic when typical consumption patterns are disrupted. By adapting these advanced models to incorporate new, relevant data inputs, we propose a new standard in forecasting precision, ensuring optimal energy management even in the face of significant behavioural shifts induced by global crises.

This research work builds on [7] and [8], in the context of CIs as addressed by the Horizon Europe project SUNRISE building resilience in cases of unforeseen events, such as pandemics. It particularly focuses on the needs of CIs (like railway, water distribution operators, hospitals, etc.), however, it is clearly evident that the main dependency of all CIs is electricity. Ensuring stable electricity availability (stability of the network) is dependent on many factors, the most prominent being patterns of energy consumption and in recent years, renewable energy production (typically solar production).

II. Methodology

A. Temporal Fusion Transformer and self attention

Transformer-based models have been shown to surpass traditional architectures like recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) in performance, making them an attractive option for a wide range of applications. The methodology we consider in this study reapplies the TFT approach [9], leveraging the transformers' architecture and self-attention mechanisms inherent to this architecture. The TFT model accommodates the input of various variables using a variable selection network (VSN), which assesses the significance of each input. This system enhances the influence of more impactful inputs while diminishing the effects of less relevant and noisy data. Based on these evaluations, inputs are combined and subsequently processed further. The merged inputs are sent in a LSTM, used to make sense of temporal relations between the time stamps recurring to past and future covariates. The subsequent phase applies the static enrichment layer, which is particularly beneficial when dealing with numerous categories, such as price, carbon emissions, and load demand. This layer enhances the model's handling of such diverse classes. Following this, the temporal self-attention mechanism comes into play, where the model prioritizes (i.e. focuses attention on) the most critical time positions. This mechanism is crucial for identifying both long-term and short-term dependencies within the observed and known time-varying inputs. In the training phase, the model minimizes the loss function by tweaking the weights, which in the VSN amplifies the impact of significant features while suppressing the noisy ones. Similarly, the attention mechanism concentrates on time positions that have a substantial effect on predictions. A clear example is in energy data analysis, where the model gives precedence to past weekends to enhance the accuracy of weekend consumption forecasts.

To train the models we've used Optuna, a hyper parameter optimization framework [10], and for the loss function we utilised Quantile loss [11].

TABLE I
TABLE OF HYPER-PARAMETERS FOR THE 24H MODEL

Hyper-parameter	**Value**
attention head size	32
dropout	0.28
hidden size	92
hidden continuous size	64
learning_rate	0.001
batch_size	64
lstm_layers	2
max_encoder_length	24

The optimal hyper-parameter set for our 24h model is shown in Table I. Note that for the larger 168h model, more parameters are required. Please refer to the original TFT paper [9] for those values. Given these hyper-parameters total number of trainable parameters was roughly 900,000.

B. Data Collection and Processing

The study utilizes historical weather and energy data that have been publicly shared by Transmission System Operators (TSOs) in two countries, Slovenia and Serbia, in the context of the SUNRISE project. This data collection underpins the research, providing a foundational dataset for the TFT-based analysis in this paper. The research employs historical weather measurements sourced from the open Meteostat platform [12] and solar irradiation data from Open Meteo [13]. These sources provide freely available data for research purposes, with the exception of historical forecasted weather data, which was procured in bulk from OpenWeatherMap [14] for the cities of Belgrade and Ljubljana. This forecasted data is crucial for accurate evaluations as weather predictions are updated several times a day, and using historical forecasts helps to prevent data leakage.

Specifically from the Open Meteo database, only shortwave radiation data was utilized as it was not available from other sources. This dataset is accessible via the Open Meteo API and is licensed under the open-source Creative Commons 4.0 license. It includes comprehensive meteorological data for Slovenia, detailing parameters such as temperature, relative humidity, dew point per square meter, apparent temperature, precipitation levels, rainfall, snowfall, snow depth, atmospheric pressure, surface pressure, cloud coverage, wind speed, wind direction, wind gusts, and notably, shortwave radiation.

Conversely, the Meteostat dataset comprises measured meteorological data for Serbia, which includes temperature, dew point temperature, actual humidity, precipitation, snowfall, wind direction, wind speed, peak wind gusts, pressure, daily sunshine duration, and weather condition codes. This dataset is pivotal for the Serbian energy consumption forecast tool and the benchmarking forecasting model. It is also publicly accessible and can be retrieved via API.

The electricity consumption dataset encompasses historical data on national electricity usage. This data was sourced from the ENTSO-E Transparency Platform [15] and enhanced with data from EKC. They provided baseline modelled forecasts for energy based on demand/consumption in megawatts (MW) with an hourly resolution. This comprehensive data collection allows for highly accurate comparisons, facilitated by using the same training cutoff date. Training and cutoff dates are specified in each experiment separately. If not, the training start date was the start of 2017 for "long" models and 2019 for the rest. The 2019 cutoff is related to the mobility data cutoff date.

Additionally, mobility data was supplied by Telekom Slovenije, the national telecommunications provider in Slovenia. This dataset tracks the daily number of unique connections to each cell within the network, excluding connections from hosted users. Each cell tower is divided into multiple cells, and the number of unique connections per cell serves as an indicator of mobility. Essentially, the more frequently users move and change cell towers, the higher the total count of unique connections, which in turn provides a measure of the mobility factor.

Figure 1 illustrates the variations in the mobility factor from 2019 to 2023, with annotations for the three lockdown periods. The data indicates that the reductions in mobility during the summer holidays are similar to those observed during the first lockdown. Notably, the initial lockdown had the most profound impact on mobility, with each subsequent lockdown having a progressively lesser effect; the third lockdown, in particular, shows a minimal influence on mobility patterns.

To properly analyse the performance of the models during the times of altered consumption patterns, we plan to focus on the year 2020. All the data used was sampled at 1 sample per hour and was normalised using z-score normalisation.

C. Implementation

To develop the service and train the model, we utilized Python 3.10, supplemented by several key libraries aimed at data manipulation and mathematical operations. Specifically, we used Numpy [16] for numerical computations, Pandas [17] for data analysis, Matplotlib [18] for plotting graphs, and Scipy [19] for additional scientific computations. For the deep

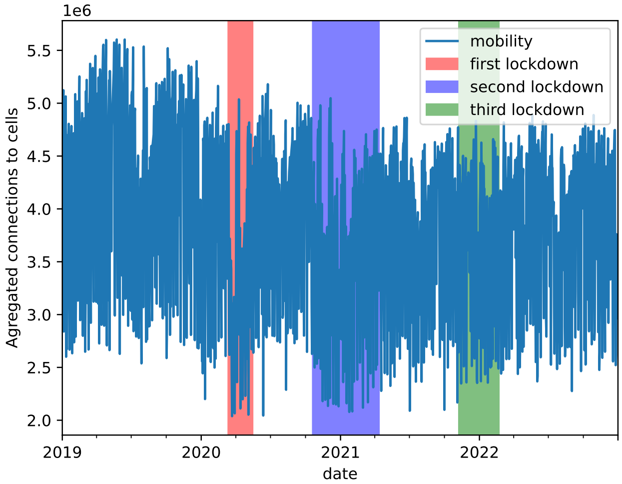

Fig. 1. Mobility between 2019 and 2023

learning components, we employed PyTorch [20], a popular framework for deep learning applications. Alongside, we used PyTorch Forecasting [21], an extension of PyTorch designed specifically for time series forecasting. This combination of tools provided a robust environment for developing complex predictive models efficiently.

D. Data Challenges

Efforts have been concentrated on understanding the basic processes involved in each of the use cases, especially in the context of the COVID-19 pandemic, which presented unprecedented challenges to urban management worldwide. The pandemic severely disrupted daily routines and behaviors as individuals who were exposed to or contracted the virus had to isolate themselves, inhibiting their ability to perform normal activities. Communities enforced social distancing measures to mitigate transmission risks. These widespread disruptions contributed to significant societal and economic impacts, including a substantial loss of life. The adjustments made during the pandemic have highlighted the importance of adaptive strategies in managing public health crises.

The dataset preliminarily consist of aggregated and fully anonymized data concerning people's activity levels, as recorded by the telecommunication provider. This primarily includes the number of individuals present in a specified area (e.g., a municipality) at a given time. The data is aggregated both spatially (to the level of municipalities) and temporally (to hourly intervals), ensuring that it is impossible to extract any privacy-sensitive information. This approach is similar to the methodologies used in the Google COVID-19 Community Mobility Reports and Apple COVID-19 Mobility Trends, which provided public access to mobility data during the pandemic. However, these sources are no longer updated and suffered from limited regional coverage and resolution. The current dataset aims to fill these gaps by providing more detailed and continuously updated information.

E. Evaluation

To assess the performance of our models, we utilized both the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE) as metrics. MAPE is particularly valuable for its intuitive interpretation, making it easier to understand forecasting accuracy. Our primary focus was on predicting load demand, a standard benchmark that enables comparison with other methodologies. Additionally, we extended our analysis to include predictions on energy prices and carbon emissions, demonstrating the versatility and broad applicability of our models in various contexts. This comprehensive evaluation helps highlight the models' effectiveness across different domains.

$$MAE = (\frac{1}{n})\sum_{i=1}^{n}|y_i - x_i| \tag{1}$$

The Mean Absolute Error (MAE) is beneficial for quantifying the actual prediction errors, which can be particularly useful when analyzing individual signals. However, for comparing performance across different models or tasks, the Mean Absolute Percentage Error (MAPE) 2 tends to be more suitable. This metric, expressed as a percentage, provides an intuitive measure of a model's accuracy.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_i - x_i}{x_i}\right| \tag{2}$$

III. Results

To demonstrate that the TFT model can improve upon current state-of-the-art methods, we will compare it to EKC and ELES in-house models currently used in production settings. Next, we will analyze the performance of the TFT model during the first COVID-19 lockdown in Slovenia and investigate the effects of additional features on the TFT, as well as on gradient boosting methods like XGBoost and CatBoost. Finally, we will perform a detailed analysis of the feature importance of the TFT model.

A. Comparing TFT and EKC models

First, we will demonstrate how our base model compares against the EKC, to predict Serbian national consumption. The evaluation was performed between 1.1.2022 and 15.1.2023, and the model was trained on data between 1.1.2019 and 31.12.2022. In Table III we present the comparison between the XLAB TFT model and the EKC model. As we can observe in Table II, we have used future forecasted weather as well as a bigger input window size of 168h instead of 24h. Here we demonstrate that by increasing the amount of information we managed to gradually improve the model's performance.

In Table III we can can further explore the results above in Table II, but in higher detail. Both the MAPE and MAE metrics improved by up to 20%, accompanied by a reduction in standard deviation and a lower maximum error.

TABLE II
PERFORMANCE COMPARISON FOR 2022 BETWEEN EKC BASELINE AND XLAB TFT MODEL

Model	MAPE [%]	Forecasted Weather	Window Size [h]
EKC_baseline	2.34	No	24
XLAB_TFT 1	2.18 (-7 %)	No	168
XLAB_TFT 2	1.93 (-17 %)	Yes	24
XLAB_TFT 3	**1.87 (-20 %)**	**Yes**	**168**

TABLE III
MAPE AND MAE COMPARISON BETWEEN EKC BASELINE AND FOR BEST PERFORMING TFT VARIANT

Metric	MAE		MAPE [%]	
Model	EKC	XLAB	EKC	XLAB
mean	92.24	**73.58 (-20 %)**	2.34	**1.87 (-20 %)**
std	91.00	**67.53**	2.075	**1.65**
max	3540.00	**860.00**	29.60	**20.05**

By extending the input window size to 168 hours, we achieved improved results, however, this also led to an increase in training time. In the next section, we will utilise the 24h model, as it offers a better ratio between time to converge and performance.

B. Comparing Iterative TFT and ELES Models

In the case of ELES, the specific training cut-off date was not known, making direct comparisons potentially unfair since one model might have more recent knowledge updates than the other. To address this issue, we developed an iterative model that updates its knowledge on a monthly basis, ensuring a fairer comparison. The model was trained on data from 1.1.2018 up to 31.12.2019. The last 14 days of this period were used for validation during the training loop. Subsequent months were used for testing, progressively incorporating more data into the training set.

In regards to the iterative model, Table IV shows the comparison with the ELES in-house model and the improvements we achieve with the approach proposed in this paper. As we can observe in the Table for 2020, the ELES model outperformed our approach, whereas, for 2021 and 2022, we were able to improve their approach by up to 10.59 % on average. The reason for the decreased performance in 2020 could be attributed to multiple factors where one of them could be related to the model not having enough data and of course big change in consumption patterns due to the COVID-19 pandemic.

Overall, an iterative approach inherently results in a more resilient model. When new consumption patterns emerge, they will be automatically incorporated into the next model iteration. An example is expressed in the data of the COVID-19 pandemic. Since no lockdowns or similar events were present in the training data, we cannot expect our model to make accurate predictions during the first COVID-19 lockdown. Even if we do not pass any information about mobility or lockdown, the model should adapt to new consumption

TABLE IV
MAE AND MAPE METRICS FOR ITERATIVE LEARNING (2020-2022)

Year	Metric	MAE		MAPE [%]	
	model	ELES	XLAB	ELES	XLAB
2020	mean	**41.77**	43.26 (+3.57%)	**2.81**	2.92 (+3.91%)
	std	**37.94**	40.37	**2.56**	2.72
	max	**689.00**	730.00	30.66	**29.55**
2021	mean	43.46	**37.34 (-13.87%)**	2.73	**2.37 (-13.1%)**
	std	42.82	**34.26**	2.58	**2.10**
	max	319.00	**300.00**	21.57	**20.00**
2022	mean	43.62	**40.07 (-8.14%)**	2.92	**2.69 (-7.88%)**
	std	44.32	**35.75**	3.21	**2.49**
	max	1214.00	**340.00**	91.97	**47.22**

patterns in the next iteration. Alternatively, we could assume that the lockdown day consumption pattern is similar to that of a weekend or a holiday. Setting a flag that would treat lockdown days as holidays would not confuse the model as much as it would without such a flag. In the next Section III-C, we will include this feature to inspect its impact.

C. Model Performance During COVID-19 Lockdown

In this section, we will utilise the very same models as in Section III-B and dive deeper into analysis during COVID-19 lockdowns in 2020 for Slovenia. By adding new features, we demonstrate how we can improve the prediction capabilities of existing models.

The Table V is calculated based on Table VI. It demonstrates that the relative difference between model predictions during lockdown and non-lockdown periods. The first two models, *ELES* and *XLAB base long* are the same models as in Table IV. The remaining models are similar but include additional features; for example, *XLAB specday long* incorporates a special day feature. The models *base, mobility,* and *mobility specday* follow a similar pattern but use less training data, specifically starting from January 1, 2019—approximately one year less than the other models. Consequently, this may result in poorer performance for these models, as they were trained with roughly 14 months of data by the first lockdown.

In Figure V we can observe a noticeable decrease in performance during the first COVID-19 lockdown. After removing the outliers, the change is roughly **32.66%**, which is in line with the literature in [3] observing a similar impact. While the analysis of the change is not relevant, it demonstrates a pattern that models utilising either mobility or special days have lower differences i.e. performing better during the first lockdown.

Most relevant is the analysis of Table VI, where we are focusing on the first column *lockdown 1* presenting performance during the first lockdown period in Slovenia. The interval can be visually observed in Figure 1.

The best-performing model is *XLAB spec day long*. Based on the comparison with its baseline, it yielded a relative improvement of roughly 18.10% and 10.13% compared to the model that was trained on data from 2019 onward. The improvement compared to the ELES model was less than 1%.

TABLE V
RELATIVE DIFFERENCES BETWEEN MAPE DURING LOCKDOWN NO-LOCKDOWN FOR VARIOUS MODELS IN 2020

Model	lockdown 1 MAPE [%]	no-lockdown MAPE [%]	Relative Difference
ELES	3.55	2.68	32.46%
XLAB base long	4.31	2.65	62.64%
XLAB specday long	**3.53**	2.71	30.26%
XLAB base	3.95	2.91	35.74%
XLAB mobility	3.90	2.95	32.20%
XLAB mobility specday	3.65	2.94	24.15%

TABLE VI
MAPE FOR DIFFERENT MODELS UNDER VARIOUS LOCKDOWN CONDITIONS WITH RELATIVE DIFFERENCES

Model	Metric	lockdown 1 MAPE [%]	lockdown 2 MAPE [%]	no-lockdown MAPE [%]
ELES	mean	3.55 (-17.63%)	2.38 (+3.48%)	2.68 (+1.13%)
	std	3.07	2.12	**2.44**
	max	16.42	15.45	30.66
XLAB base long	mean	4.31 (0.00%)	2.30 (0.00%)	**2.65 (0.00%)**
	std	3.38	2.09	2.48
	max	18.55	13.40	29.55
XLAB specday long	mean	**3.53 (-18.10%)**	2.38 (+3.48%)	2.71 (+2.26%)
	std	**2.67**	1.93	2.54
	max	16.10	12.28	**27.53**
XLAB base	mean	3.95 (-8.35%)	2.39 (+3.91%)	2.91 (+9.81%)
	std	3.14	2.07	2.90
	max	18.85	12.28	30.53
XLAB mobility	mean	3.90 (-9.51%)	2.70 (+17.39%)	2.95 (+11.32%)
	std	3.15	2.14	2.59
	max	18.24	14.91	27.94
XLAB mobility specday	mean	3.65 (-15.31%)	**2.25 (-2.17%)**	2.94 (+10.94%)
	std	2.81	**1.78**	2.64
	max	**13.97**	**12.23**	28.74

Their model performed impressively well for 2020, as we demonstrated in Table IV. Comparison to the ELES model is not relevant here, as we are focusing on assessing the impact of adding features in a controlled environment. As mentioned the only difference between the *XLAB* models are features.

With that in mind, when further observing Table VI, a pattern is observed demonstrating that models utilising either mobility or special day performed better compared to those not using it.

Another thing to notice is that the best performance was actually during the second lockdown, which can be observed over all models. This could be attributed to various facts, where patterns could have stabilised by that point and become more predictable. Additionally, models were updated with new consumption patterns by then.

The last observation from Table VI is that the mobility and special day features did not significantly enhance performance on a typical non-COVID-19 day. To confirm this observation, we will perform an extensive study of a variety of features in the next Section III-D.

D. Effect of Additional Features to Model Performance

In this section, we utilize linear regression, XGBoost, and CatBoost models to assess feature importance across a variety of input features. The results were evaluated and averaged for the years 2020 to 2023, and are presented in Table VII.

The first experiment, "None", does not include any additional features besides the signal itself. The next two experiments contain the signal and date-time features, along with off-time features such as weekends and holidays. Together, these features form a base, which is used in subsequent experiments to study the impact of individual features.

The first subgroup of experiments includes base features and weather features. As shown in Table VII, adding future weather improves performance for more complex models, where experiments utilising base and future weather yielded the best overall results for XGBoost and CatBoost.

The next group of experiments examines the addition of mobility data to the base set of features. For linear regression models, the mobility group outperformed the weather features group, whereas, for more complex models (XGBoost and CatBoost), the performance was worse compared to using weather features. Overall, more complex models performed better than the simple linear regression model. Here, we have to keep in mind, that errors are a lot higher for linear regression.

TABLE VII
EFFECT OF ADDING NEW FEATURES BETWEEN 2020 AND 2023.

Experiment Name	linreg MAPE[%]	xgb MAPE[%]	cbm MAPE[%]
none	6.20 (+0.00%)	5.46 (+0.00%)	5.13 (+0.00%)
datetime	6.66 (+7.31%)	3.71 (-32.07%)	3.37 (-34.28%)
base	5.67 (-8.60%)	3.60 (-34.01%)	3.39 (-33.85%)
base_weather	5.44 (-12.35%)	3.69 (-32.36%)	3.36 (-34.43%)
base_weatherfut	5.63 (-9.22%)	**3.51 (-35.67%)**	**3.17 (-38.29%)**
base_mob	5.42 (-12.63%)	3.67 (-32.87%)	3.36 (-34.58%)
base_specday	5.64 (-9.01%)	3.64 (-33.40%)	3.36 (-34.49%)
base_mobspecday	**5.31 (-14.38%)**	3.67 (-32.80%)	3.31 (-35.45%)
base_mobfuture	6.94 (+11.95%)	3.67 (-32.86%)	3.31 (-35.52%)
all_specday	5.40 (-13.00%)	3.64 (-33.37%)	3.23 (-37.01%)
all_mob	5.38 (-13.19%)	3.63 (-33.49%)	3.23 (-37.06%)
all_mobfuture	7.31 (+17.79%)	3.63 (-33.47%)	3.23 (-37.03%)
all	6.06 (-2.27%)	3.64 (-33.42%)	3.19 (-37.73%)

**base stands for date-time and holiday features used together*

In the final set of experiments, combining all features, including special days, mobility, and future mobility, resulted in overall solid performance. The complex models, XGBoost and CatBoost, showed consistent improvement and leveraged the extensive feature set effectively. This indicates that the integration of a diverse range of features allows these advanced models to extract and utilize information more effectively, enhancing their prediction accuracy. Notably, the models in the "all" group demonstrate that these models performed quite well on average, achieving significant reductions in error rates compared to the baseline models. Even though the best results were achieved using future weather data, it makes sense to include data related to mobility, if available.

In the next Section, we focus on the effect of given features during the COVID-19 lockdown period in 2020.

E. Effect of Additional Features on Model Performance during COVID-19

To evaluate the impact of additional features during the COVID-19 pandemic, we must analyze their effects specifically during lockdown periods. Let's first examine Table VIII presenting a performance of the models during normal or non-lockdown days. In comparison to the findings in Table VII (from the previous Section), we observe a consistent pattern: models incorporating mobility and special day features offer limited performance improvements.

TABLE VIII
EFFECT OF ADDING NEW FEATURES FOR NON-LOCKDOWN DAYS IN 2020 (MAPE)

Experiment Name	linreg MAPE[%]	xgb MAPE[%]	cbm MAPE[%]
none	6.83 (0.00)	5.99 (0.00)	5.55 (0.00)
datetime	9.81 (43.70%)	4.22 (-29.54%)	3.68 (-33.82%)
base	6.94 (1.61%)	3.99 (-33.31%)	3.78 (-31.82%)
base_weather	6.25 (-8.43%)	4.14 (-30.87%)	3.69 (-33.52%)
base_weatherfut	6.78 (-0.68%)	**3.93 (-34.37%)**	**3.50 (-37.00%)**
base_specday	6.95 (1.90%)	3.99 (-33.31%)	3.76 (-32.36%)
base_mob	6.83 (0.14%)	4.32 (-27.84%)	3.94 (-29.07%)
base_mobfuture	12.16 (78.10%)	4.33 (-27.68%)	3.91 (-29.57%)
base_mobspecday	6.85 (0.32%)	4.32 (-27.86%)	3.93 (-29.29%)
all_specday	**6.12 (-10.35%)**	4.12 (-31.23%)	3.69 (-33.18%)
all_mob	6.77 (-0.81%)	4.25 (-28.93%)	3.75 (-32.38%)
all_mobfuture	12.59 (84.47%)	4.29 (-28.32%)	3.79 (-31.75%)
all	6.95 (1.89)	4.25 (-29.13%)	3.78 (-31.99%)

**base stands for date-time and holiday features used together*

We must keep in mind that approximately **68.2%** of our data represents normal, non-lockdown days, with the remainder being lockdown days. This context is crucial for interpreting model performance. Similar to our observations with TFT, we face the challenge of limited data (only one year), which may affect prediction accuracy. Nonetheless, patterns forming across many models and feature combinations should still yield relevant results to be able to confirm or not confirm our hypothesis.

Table IX confirms a significant drop in model performance during the first lockdown compared to normal days, aligning with the results from Table V. Furthermore, linear regression was the least accurate of the models, making it almost unuseful in some cases. While we cannot expect lockdown prediction accuracy to match non-lockdown periods, our goal should be to minimize this performance reduction.

What we can notice, is that experiments including mobility data and special day data do contribute to improvements to better performance across all models and combinations. Several factors support this observation. The first one is that all best-performing experiments for every model include mobility or special days. For the second one, let us focus on the CatBoost model. Based on the results, it is the best-performing model.

Before we analyse the results more in-depth. It is worth clarifying that only experiments incorporating special days and mobility futures contain information on possible big changes in consumption patterns. While the 'mobility future' scenario demonstrates potential gains with perfect mobility forecasts, this is not realistically achievable. For the CatBoost model best-performing experiment was the 'mobility future', since it is not realistically possible, it is presented in italic.

TABLE IX
EFFECT OF ADDING NEW FEATURES FOR THE FIRST LOCKDOWN DAYS IN 2020 (MAPE)

Experiment Name	linreg MAPE[%]	xgb MAPE[%]	cbm MAPE[%]
none	8.33 (0.00)	8.19 (0.00)	7.57 (0.00)
datetime	8.68 (4.17%)	**6.67 (-22.61%)**	5.86 (-18.56%)
base	8.20 (-1.58%)	6.98 (-14.81%)	6.78 (-10.44%)
base_weather	7.43 (-10.81%)	7.25 (-11.58%)	6.99 (-7.64%)
base_weatherfut	8.41 (1.06%)	6.99 (-14.75%)	6.36 (-15.93%)
base_specday	6.98 (-16.21%)	6.98 (-14.82%)	6.17 (-18.50%)
base_mob	8.34 (0.19%)	7.03 (-14.15%)	6.40 (-15.48%)
base_mobspecday	6.92 (-16.90%)	7.03 (-14.15%)	**5.64 (-25.53%)**
base_mobfuture	12.23 (46.89%)	6.99 (-14.75%)	*5.49 (-27.40%)*
all_specday	**6.14 (-26.25%)**	7.30 (-10.97%)	5.78 (-23.66%)
all_mob	8.29 (-0.44%)	7.21 (-11.97%)	6.68 (-11.80%)
all_mobfuture	14.49 (73.93%)	7.18 (-12.38%)	6.65 (-12.12%)
all	13.47 (61.70%)	7.21 (-11.97%)	5.95 (-21.34%)

**base stands for date-time and holiday features used together*

Overall, we can notice that CatBoost models containing either mobility, special day or both performed better compared to those not utilising these features. However, it's important to stress that this advantage is not observed for non-lockdown days.

When analysing the effect of individual features, based on the results we could argue that special day feature has a bigger impact than mobility. Of course, in a world where we would be able to perfectly predict mobility, the best feature would be (future) mobility.

Across the tables VII, IX, and VIII, we observe a pattern: models containing all features often under-perform compared to the best-performing combinations. This suggests the potential impact of the 'curse of dimensionality'. With too many features and limited data, the model may struggle to identify meaningful relationships.

More importantly, the pattern observed is similar to that of Section III-C, where we have demonstrated that models utilising either mobility or special day feature, on average performed much better during the first COVID-19 lockdown, compared to those not utilising. Similar conclusions can be made on effect during normal days, where their effect is present but does not have a big impact, in some cases even causing a curse of dimensionality.

F. Explainability and Feature Importance

To address the explainability part of the model, we provide average feature importance from the iterative model for the past three years as seen in Figure 2. The encoder features represent information from the past that is already known, while the decoder features represent information from the future that we are trying to predict.

The encoder features (in blue) in Figure 2 demonstrate the dominance of real load in shaping end results within the encoder. The second most important feature is mobility, which additionally confirms our observations. Date-time features and

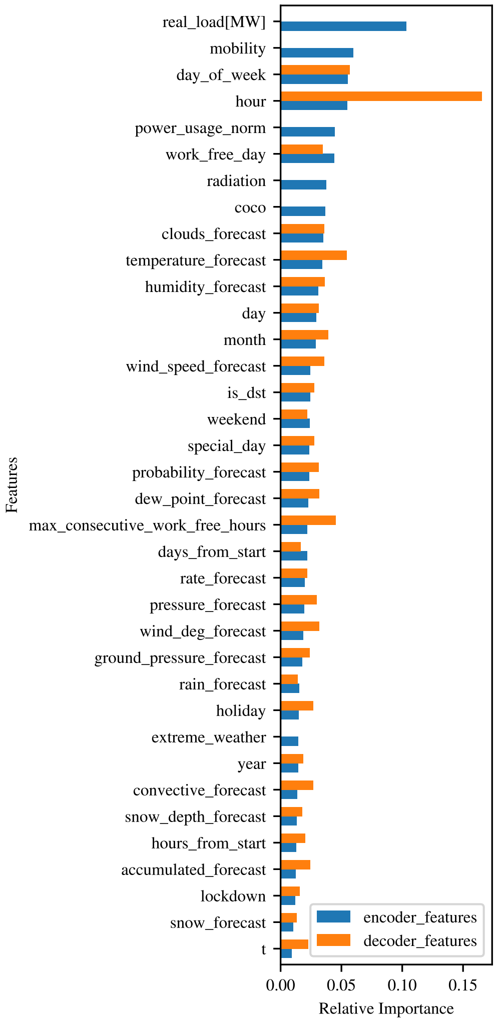

Fig. 2. Encoder feature importance

modelled solar radiation follow in importance. Individual modelled weather features exhibit relatively minor impact; however, their aggregate influence remains significant due to their large number.

The decoder features (in orange) in the same Figure 2 reveal a similar trend to encoder features, particularly regarding the most influential factors. Notably, modelled weather features appear to exert a stronger impact within the decoder context.

While it is expected for target and date-time features to be most important, it was less expected to see a big impact of measured solar irradiation on predictions. This could be explained by an increasing amount of PV installed. In Slovenia, solar energy accounts for less than 10% of total energy (more precisely 7.58% as of 2023 [22]. This growth in solar power could be making accurate solar irradiation data even more important for accurate predictions.

IV. Discussion

The obtained results show, that mobility data and special day features do improve the results during the first lockdown period. This was demonstrated for deep learning models utilising an iterative approach, as well as gradient boosting methods such as XGBoost and CatBoost. In the case of the iterative approach, the performance gains were almost 20 % for the model containing special days only and the model with no context regarding the COVID-19 status, supporting the findings in [3].

On the other hand, we have demonstrated the opposite for normal days, non-COVID-19 days, where the addition of mobility and special day features did not significantly improve the results. Even though no significant improvement was present, feature importance for multiple runs in 2020 suggests that mobility contains a lot of relevant information being ranked the second most important in the feature importance plot in Figure 2. The information it provides might be extracted elsewhere from weather and energy consumption patterns. The overall TFT model, when applied to load demand prediction, outperforms current state-of-the-art approaches. We have demonstrated that the iterative model is able to ingest new consumption patterns, thus improving its performance.

These results verify the hypothesis proposed in the context of COVID-19 pandemic data, demonstrating that model accuracy and performance can be improved during large-scale events. To address the more subjective research question of what energy critical infrastructure (CI) should consider in disruptive scenarios, recent experiences and best practices from the pandemic indicate that the primary focus should be on understanding the impact of demand changes on a transmission system (TS). It is essential for a CI operator, specifically a TSO, to always be prepared for any unforeseen scenarios, such as future pandemics, to maintain operational continuity (flawless, uninterrupted core business activities). At the same time, it is crucial to uphold activities related to grid resilience, employee health and safety, cybersecurity measures, and other vital operations.

The severe disruptions caused by the COVID-19 pandemic in the daily routines and behaviours of consumers led to a change in the shape of the daily consumption diagram. For instance, before COVID-19 the daily peak of demand was in the evening, however during COVID-19 the daily peak shifts to the morning. The previous change has a big influence on the adequacy assessment of a TS. Namely, the adequacy analysis or adequacy assessment of a TS is the most important analysis of a TS which answers the question: ‘'Is there going to be enough energy in the system in each situation including failures of generators”. Based on this analysis, the levels of necessary spinning and non-spinning reserves are determined which are crucial for the secure operation of a TS. An incorrect adequacy analysis leads to an increase in the percentage of loss of load probability (LOLP), which can further lead to significant economic damage. The economic damage when the energy is not served (ENS) is usually estimated with the Value of Lost Load (VoLL).

VoLL represents the economic value associated with not being able to supply electricity to consumers during periods of high demand or supply shortages. It reflects the cost to consumers of being without electricity for a certain period and includes factors such as lost production, inconvenience, and potential damage to equipment or goods. VoLL is usually expressed in currency per MWh and can vary depending on factors such as the type of consumer, the time of day, and the duration of the outage. In the European Union, the specific values for VoLL can vary between countries and regions due to differences in electricity market structures, consumer preferences, and economic conditions. Typically, VoLL values in the EU range from around €1,500 to €23,000 per MWh [23], but they can be higher in some cases, particularly for critical services or industries where the cost of downtime is very high. For comparison, wholesale market prices are usually at the level between 50-100 €/MWh. Other key aspects to be considered are discussed in [24].

The conventional approach takes typical daily load patterns when conducting an adequacy analysis. The results of this research hint at the issues that energy CIs must consider when preparing resilience strategies for disruptive scenarios based on recent experience and good practices in the latest pandemic. They show that the typical daily load patterns can be interrupted for a significant period in the case of large-scale events such as COVID-19. Therefore, to decrease LOLP the CI (i.e. TSOs) should consider the use of untypically daily load patterns while conducting an adequacy analysis. These load patterns can be generated by the model described in this research.

Regarding further analysis and discussions for future disruptive scenarios, TSOs could also consider using industrial and residential power consumption predictions as encoder and decoder features (for total consumption forecast), taking into account that these two consumption categories could perform significantly differently in specific situations. Currently, most of the TSOs do not have separate metering information for industrial and residential power consumption in real-time (usually they get data for residential consumption at the end of the month), as well as their forecasts, although these values could be estimated well.

Also, the number of prosumers or active customers (solar installations "behind the meter") is increasing rapidly, and taking into account that analysis shows solar irradiation as a very important feature, future models could also consider the installed power capacity of active customers/prosumers (at least on yearly level) as a feature (because it is changing/growing over the time, and solar generation has the impact on the measured net consumption).

V. Conclusions

The goal was to develop a resilient model, that enables us to make better predictions during times of large-scale events that have a significant effect on energy demand prediction models. In this work we show how we can use the mobility and lockdown flag to improve model accuracy and performance during impact large-scale events, taking as a basis the data collected during the COVID-19 pandemic.

We achieve these improvements with additional features, particularly by considering special day features. The biggest impact of these features can be observed in the first lockdown period, whereas for normal days, improvements were harder to notice. Overall it makes sense to utilise both mobility and special day functions, if available. They offer insights in case of large-scale events, even if they are not common. When taking into account the cost of individual features, mobility data may turn out to be relatively expensive, here special day features are much more cost-efficient.

Moreover, the periodic monthly update of the model shows great benefit for the predictions computed and the resilience of the models, as we have seen that the effect of COVID-19 was hardly noticeable in the second lockdown.

It was particularly clear, in the cases of EKC and ELES, that the improvements in prediction accuracy have significant real-world benefits. During the COVID-19 pandemic, shifts in daily consumption patterns, such as peak demand moving from evening to morning, impacted the adequacy assessment of transmission systems (TS). Accurate forecasts help ensure there is enough energy to meet demand, reducing the LOLP and preventing economic damage from unserved energy (ENS). Better predictions lead to optimal levels of spinning and non-spinning reserves, enhancing grid stability. Economically, accurate predictions minimize wasted energy, saving costs associated with VoLL.

The investigation of larger, more complex versions of the TFT is crucial as the industry shifts from 1-hour to 15-minute resolution quadrupling the input parameters. We have demonstrated that using a 168h input window improves its performance, but again increases the number of input features by up to seven-fold. Both changes result in a much bigger and more complex model, highlighting a key area for future research on performance impacts.

Research is increasingly focusing on foundational time-series models based on transformers. Authors of research work in [25], [26], [27] and [28] focus on zero-shot forecasting of univariate time-series. Inspired by breakthroughs in natural language processing with models like Gemini, GPT and Claude, time-series prediction, similarly aims to predict the most probable next value based on prior input. Referenced models often outperform TFT in certain applications, showing promising results. These methods do not directly apply to our work, as we are using multivariate time-series, whereas the papers are focusing on univariate time-series. A potential implementation approach for multivariate foundational models could be composed out of multiple foundational models, each fine-tuned for a specific task, then fusing the outputs to create a similar architecture to the TFT. At the time of writing, no studies have been published on this specific approach, but further research in this direction is anticipated. Overall, these approaches will play a significant role in efficiently managing the operations of CIs.

ACKNOWLEDGMENT

This research has received funding from the European Union's Horizon Europe research and innovation programme under the grant agreement No. 101073821 (SUNRISE) and 101070052 (TANGO).

REFERENCES

[1] M. Kwiatkowska and X. Zhang, "When to trust ai: Advances and challenges for certification of neural networks," in *Proceedings of the 18th Conference on Computer Science and Intelligence Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki, and D. Ślęzak, Eds., vol. 35. IEEE, 2023. doi: 10.15439/2023F2324 pp. 25—-37.

[2] S. D. Li, C. and T. Reindl, "Real-time scheduling of time-shiftable loads in smart grid with dynamic pricing and photovoltaic power generation," *In: 2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, pp. 1–6, 2015.

[3] Y. W. Chen, Y. and B. Zhang, "Using mobility for electrical load forecasting during the covid-19 pandemic," 2020.

[4] B. Lim, S. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.

[5] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *CoRR*, vol. abs/1603.02754, 2016. [Online]. Available: http://arxiv.org/abs/1603.02754

[6] A. V. Dorogush, V. Ershov, and A. Gulin, "Catboost: gradient boosting with categorical features support," 2018.

[7] J. Jenko, "Development and analysis of new activation based load profiles," Master's thesis, University of Ljubljana, Faculty of Electical Engineering, 2023.

[8] J. Jakob and J. Pita Costa, "Using temporal fusion transformer predictions to maximise use of renewable energy sources," *In: Proceedings of the International Workshop on Artificial Intelligence and Machine Learning for Energy Transformation (AIE), IEEE*, 2024.

[9] B. Lim, S. O. Arik, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," 2020.

[10] Optuna. (2024) Optuna documentation. [Online]. Available: https://optuna.readthedocs.io/en/stable/

[11] Pytorch. (2024) Quantileloss documentation. [Online]. Available: https://pytorch-forecasting.readthedocs.io/en/stable/api/pytorch_forecasting.metrics.quantile.QuantileLoss.html

[12] Meteostat. (2024) Weather and climate database. [Online]. Available: https://meteostat.net/en/

[13] Open-meteo. (2024) Free weather api. [Online]. Available: https://open-meteo.com

[14] Openweathermap. (2024) Open weather history forecast bulk. [Online]. Available: https://openweathermap.org/api/history-forecast-bulk

[15] ENTSO-E. (2024) Entso-e transparency platform. [Online]. Available: https://transparency.entsoe.eu/

[16] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Ro, M. Wiebe, P. Peterson, P. Gerard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with numpy," *Nature*, vol. 585, no. 7825, p. 357 to 362, 2020. doi: 10.1038/s41586-020-2649-2. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2

[17] W. McKinney *et al.*, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, vol. 445. Austin, TX, 2010, p. 51 to 56.

[18] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science and Engineering*, vol. 9, no. 3, p. 90 to 95, 2007. doi: 10.1109/MCSE.2007.55

[19] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, p. 261 to 272, 2020. doi: 10.1038/s41592-019-0686-2

[20] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An open source machine learning framework," *arXiv preprint arXiv:1802.02611*, 2018.

[21] PyTorch Forecasting Developers, "Pytorch forecasting documentation," https://pytorch-forecasting.readthedocs.io/en/stable/getting-started.html, 2024.

[22] P. Jowett. (2024) Slovenia's new solar additions hit 400 mw in 2023. [Online]. Available: https://www.pv-magazine.com/2024/02/19/slovenias-2023-solar-additions-hit-400-mw/

[23] Swinand, G. Peter, Natraj, and Ashwini, "The value of lost load (voll) in european electricity markets: Uses, methodologies, future directions," in *2019 16th International Conference on the European Energy Market (EEM)*, 2019. doi: 10.1109/EEM.2019.8916400 pp. 1–6.

[24] D. R. M. L. Quang Hieu Vu, Ling Cen, "Key factors to consider when predicting the costs of forwarding contracts," in *Proceedings of the 18th Conference on Computer Science and Intelligence Systems*, ser. Annals of Computer Science and Information Systems, vol. 30. IEEE, 2022. doi: 10.15439/2022F293 pp. 447–450.

[25] K. Rasul, A. Ashok, A. R. Williams, H. Ghonia, R. Bhagwatkar, A. Khorasani, M. J. D. Bayazi, G. Adamopoulos, R. Riachi, N. Hassen, M. Biloš, S. Garg, A. Schneider, N. Chapados, A. Drouin, V. Zantedeschi, Y. Nevmyvaka, and I. Rish, "Lag-llama: Towards foundation models for probabilistic time series forecasting," 2024.

[26] A. F. Ansari, L. Stella, C. Turkmen, X. Zhang, P. Mercado, H. Shen, O. Shchur, S. S. Rangapuram, S. P. Arango, S. Kapoor, J. Zschiegner, D. C. Maddix, H. Wang, M. W. Mahoney, K. Torkkola, A. G. Wilson, M. Bohlke-Schneider, and Y. Wang, "Chronos: Learning the language of time series," 2024.

[27] A. Das, W. Kong, R. Sen, and Y. Zhou, "A decoder-only foundation model for time-series forecasting," 2024.

[28] A. Garza, C. Challu, and M. Mergenthaler-Canseco, "Timegpt-1," 2024.