Forecasting Electricity Consumption in Industrial Kitchens

Jorge Ruiz Amantegui

jorge.amantegui@tecnico.ulisboa.pt

Instituto Superior Técnico, Universidade de Lisboa, Portugal

September 2022

Abstract- Even though Industrial Kitchens (IKs) are among the highest energy intensity spaces, very little work has been done to forecast their consumption. This work explores the possibility of increasing the accuracy of the forecast in IK by forecasting for the appliances that compose the total load and summing those forecasts, instead of forecasting for the total load (virtual aggregate). To do so, three different methods are used: statistical method (Prophet), classical machine learning (Random Forest) and deep learning (LSTM). The data collected consists of power measurements on the appliances of a Portuguese restaurant roughly for one month.

Keywords: Load forecasting, machine learning, virtual aggregate, industrial kitchen.

I. Introduction

The global demand for electricity keeps increasing, strong economic growth, combined with more extreme weather conditions than in 2020 boosted the demand by 6% since 2021 [1]. These changes in electricity consumption and a prediction for the near future can be seen in *Figure 1*. Despite the efforts made in the recent years regarding the development and implementation of renewable energy sources (RES), electricity generation from coal and gas hit record levels and as a results the global electricity sectors annual carbon dioxide emissions reached the highest value ever recorded [1].

Increasing the renewable energy sources (RES) penetration on the mix is of upmost importance to fight climate change. And implementing RES in one of the highest electricity consuming businesses (IKs) should be a priority.

II. Motivation and Objectives

Given the context mentioned above, this work explores a new methodology to study the forecast of the electricity consumption of IKs. The main tasks developed in this work are:

- Develop methods to forecast the virtual aggregate.
- Develop methods to forecast the individual appliances.
- Compare the performance metrics of the forecast of the virtual aggregate against the performance metrics of the sum of the forecast of the individual appliances.
- Compare the performance of different forecasting algorithms.

Virtual aggregate refers to the load of the restaurant defined as the sum of the loads of each individual appliance used in the restaurant. It is different from the real aggregate (the measured value for the load) because it does not include certain elements such as lighting or air ventilation.

The goal of this study is to explore the possibility to improve the energy consumption forecast in a an unexplored and energy intensive area. Finding suitable forecasting methods for an effective management of energy resources improves the efficiency in energy consumption and decreases its impact on the environment [2].

III. State of the Art

According to Mudie et al. [3], electricity consumption in IKs is vastly greater than the literature suggest. Kitchen and food-related activities greatly outweigh other energy using activities [4], furthermore, the contract catering sector carbon footprint is 80% higher than initial estimates [5].

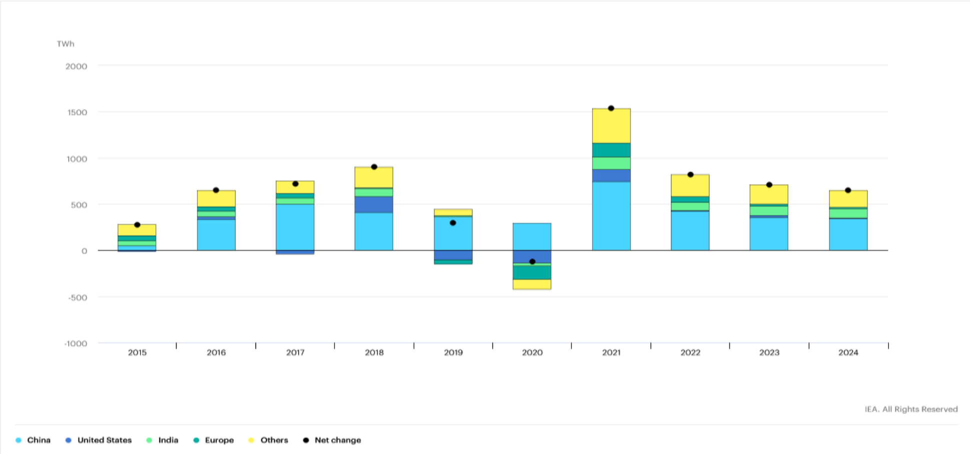

Figure 1 Global changes in electricity consumption [1]

A study developed by Higgins et al. showed that the majority of the literature regarding sustainability of restaurants only engages with parts of sustainability, particularly ecological aspects, rather than holistic sustainability [6]. This limits the possibility to achieve more sustainable restaurant options. There are, however, some exceptions to this, in 2017 a holistic evaluation of the carbon and energy impact of different food preparation options in delivering restaurant's menu was carried out in [7]. After contemplating seven energy reduction scenarios the study determined that up to 58% of energy could potentially be saved.

Regarding the introduction of RES in the operation of IKs, research is very scarce. When using "scholar.google.com" as an academic research engine using the keywords "renewable"," industrial" and "kitchen" almost all of the work that appears is related to developing biomass fuel from industrial kitchen waste or implementing a heat recovery system. In a similar manner, by searching for the exact desired results "implementation of renewable energy in industrial kitchens" there is not any related work in the first four pages. By changing the keyword "industrial" to "commercial" there is more luck, and some research is found.

Among the works that were found, two researches involved buildings composed of a mix of apartments and one commercial kitchen, these building are connected to a small-scale microgrid which supplies renewable energy and electric storage. The microgrids are composed by PV arrays and wind turbines as well as additional back-up, by either being connected to the grid or to a backup diesel generator. Microgrid sizing and costs are generally sensitive to load shape [19]. The research showed that, by implementing a thermal storage there was potential to increase renewable penetration by 7% and decrease renewable curtailment (renewable energy resource that cannot be utilized, and is therefore wasted) by 11 % [20] indicating that multiple types of storage can benefit renewable energy penetration. Even though the goal is to have building with a 100% penetration of renewable energy sources, sometimes this is not the most economically viable option. In the work done by Aldaoubab et al., after doing economic analysis it demonstrates that lowering the desired penetration from 100% to 80% cuts the energy cost significantly [19] due to the required sizing of the microgrid. The economical aspect is a crucial concept to move towards a 100% renewable energy penetration.

IV. Selection of Algorithms

It was decided that three different algorithms were to be used to forecast the consumption of the restaurant and the individual appliances. To have heterogeneity in the algorithms one statistical based model was selected, another classical machine learning algorithm and finally a deep learning algorithm.

For a statistical model, Prophet was selected. Prophet is a modular regression model with interpretable parameters. This algorithm has been used to predict energy consumption in various works and in different ways. In [8] and [9] implement a hybrid prophet-LSTM (Long-Short Term Memory) model to forecast electric load are presented. In [10] the implementation of prophet to use real-time day ahead forecasts.

For the classical machine learning algorithm, random forest (RF) was chosen. RF is considered an ensemble homogenous model. An advantage over conventional DT is that it corrects the tendency of these to overfit their training set. It does so by generating a group of several DT whose results are aggregated into one [11]. They usually implement randomization on two different levels, by having each DT trained with a random subset of observations and then having each tree node divided by considering a random subset of variables [12].

There are many different DL techniques that have been applied to energy building forecasting, such as: autoencoder, recurrent neural network, convolutional neural networks, deep belief networks or recurrent neural networks (RNN). Long Short-Term Memory (LSTM) is a kind of RNN that has been used with success to forecast energy consumption in buildings and was selected as a deep learning method to forecast in this work. Research conducted in 2020 [13] compared a variety of different algorithms in order to forecast the heating demand of a district system. Their research showed that the LSTM models were among the highest performing algorithms. Regarding commercial-residential buildings, forecast using LSTM models has also been used in [14]. In this research LSTM models were compared to other prominent data-driven models and their experiment showed the LSTM model obtained the lowest forecasting error.

V. Methodology

The initial step is to carry out a pre-processing of the data. This includes methods to handle null values and outliers, data resolution resampling and normalization of the data. In the following step, data is further processed to create the virtual aggregate, which consists of the sum of the consumption timeseries of each individual appliance loads.

Since the appliances have unpredictable behaviour and can take a wide range of values, it is not easy to spot outliers. Therefore, a visual inspection of the loads is carried out, no outliers were found. To deal with null values, a forward fill function was carried out (filling a null value with the previous valid value). The next step is to resample the data to a five-minute resolution. At this point the virtual aggregate is created by simply adding the loads of all the individual appliances.

Normalization of the data is carried out now. This helps to evaluate the performance of the algorithms for different appliances (which have very different consumption patterns). This work has taken a max normalization approach where all values will be rescaled between 0 and 1 by undergoing the process in equation 1, where *(x)* is the total dataset and (x_i) is a single value of the dataset. This method allows for easy interpretability of the resulting forecast and its performance metrics. Also, it is easy to rescale to unnormalized values.

$$x_{i,normalized} = \frac{x_i}{\max(x)} \qquad (1)$$

To make the RF and LSTM forecast the data underwent a sliding window process. A window is defined as a set of consecutive datapoint where *n* are input values and *m* are output values. The window "slides" until all values in the dataset are covered. This is illustrated in *Figure 2*, were a dataset consisting of an array of values from 0 to 100 is divided into 96 windows of n=5 and m=1. This is done to establish the way the RF and LSTM models will learn. Relating *n* values to predict the next *m* (the forecast horizon). In this work it was selected *n=288* (one day data), *m=12* (one hour horizon). A different approach is taken with the Prophet algorithm, by taking advantage of how quick it can train, the model is retrained at each window with the new available data and forecasts for the next 12 datapoints (one hour).

Window number	Input Window Values					Output Window Values
1	0	1	2	3	4	5
2	1	2	3	4	5	6
3	2	3	4	5	6	7
4	3	4	5	6	7	8
5	4	5	6	7	8	9
.	.	.	.	.	.	.
96	95	96	97	98	99	100

Figure 2 Sliding window example. n=5, m=1

The data must now be separated into a training and a test set. For this particular case, it was decided the training would be composed of blocks of full consecutive days. The fact that only full days are used to train ensures that the model has had the same amount of data to learn the consumption patterns at different times of the day, and it is important that these are consecutive days since the consumption at the end of one day is always followed by the beginning of the next day and there still could be relationships here, even though this time of the day is of less interest than when the restaurant is operating. Finally, it was decided that the testing data would compose roughly 15% of the available data while the 85% remaining will be used for training or validation. It was decided that a validation set would be used in the training of the LSTM model only. This is because LSTM models have more hyperparameters to train and can benefit more from the implementation of a validation set (at the cost of shortening the training set) and because of the ease to implement this using the keras LSTM model in python.

This means, the RF and LSTM methods will utilize the first 288 datapoints of the test as an input window to predict the next 12, and the window will "slide" 12 datapoints to predict the next 12 points. All the forecasts will finally be chained to create the total forecast. For Prophet, all the dataset will be tested 12 datapoints at a time, retraining each time, all forecast are chained finally.

Two popular performance metrics are root mean squared error (RMSE) and mean absolute error (MAE) shown in equation 2 and 3. Where *n* is the number of samples,y is the target value and $\hat{y}$ is the forecasted value.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (3)$$

While they have both been used to asses model performance for many years, there is no consensus on the most appropriate metric for model errors [15]. While the MAE gives the same weight to all errors, the RMSE penalizes variance as it gives errors with larger absolute values more weight than errors with smaller absolute value.

With all these taken into account and since the dataset is bound to be very diverse (individual appliances are going to have distinct consumption patterns from one another), there is no guarantee the error follows a normal distribution. It is hard to select a performance metric that will satisfy all needs. With all this considered, it was decided to measure the error in the forecast by using both RMSE and MAE to obtain more knowledge of the performance of the algorithms.

VI. Case Study: Virtual Aggregate

The demand of individual appliances was collected using the Egauge circuit-level smart-meter. A total of 18 appliances were individually monitored for up to 30 days at UVA. For each appliance active power is measured.

In order to understand how much of the total demand of the restaurant (i.e., aggregated), can be explained by the individual appliances, the notion of virtual aggregate was introduced. In this context, virtual aggregate corresponds to the sum of the consumption of each of the individual appliances and is given by equation 4. Where $X_a(t)$ corresponds to the measured quantity (e.g., active power) in appliance a at time instant t.

$$VA_X(t) = \sum_{a=1}^{A} X_a(t) \quad (4)$$

Figure 3 shows the virtual aggregate load before pre-processing, a clear gap in the data can be seen right before the 13th of February. This cannot be dealt with a forward fill, therefore all the data before this is day is removed from the dataset. Remaining null datapoints are filled using a forward fill method. Overall, this dataset had over eight thousand missing data points.

The forecast for the virtual aggregate has been done using three different methods: Prophet (with two different time horizons), RF and LSTM. Figure 4 is an illustration of the data used for training, the test set and the forecast obtained.

Table 1 shows the performance metrics obtained from the forecast of the virtual aggregate. RF performed the best, but all methods obtained similar results.

Table 1 Virtual Aggregate normalized RMSE and MAE

	LSTM	RF	Prophet (Three days horizon)	Prophet (One hour horizon)
Virtual Aggregate RMSE	0.11	0.10	0.12	0.10
Virtual Aggregate MAE	0.066	0.066	0.073	0.068

VII. Case Study: Sum of forecast for individual appliances

The same process as for the virtual aggregate was replicated for each individual appliance and the forecasts were stored. They will then be de-normalized and summed together to be compared against the forecast of the virtual aggregate.

Out of the eighteen appliances tested, one did not contain useful information as the test set was composed of zero values. Out of the remaining seventeen appliances, RF showed the better forecast accuracy in ten. The Prophet forecast with a three-day horizon showed the worst forecast in nine out of the seventeen appliances. As expected, the Prophet model improved performance with a shorter horizon and retraining the model as new data was available. This method was much more robust, it was the top performer for four of the appliances.

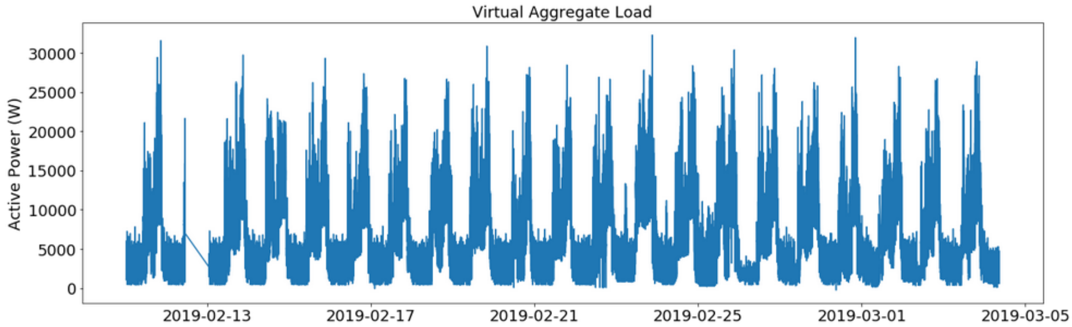

Figure 3 Virtual aggregate load

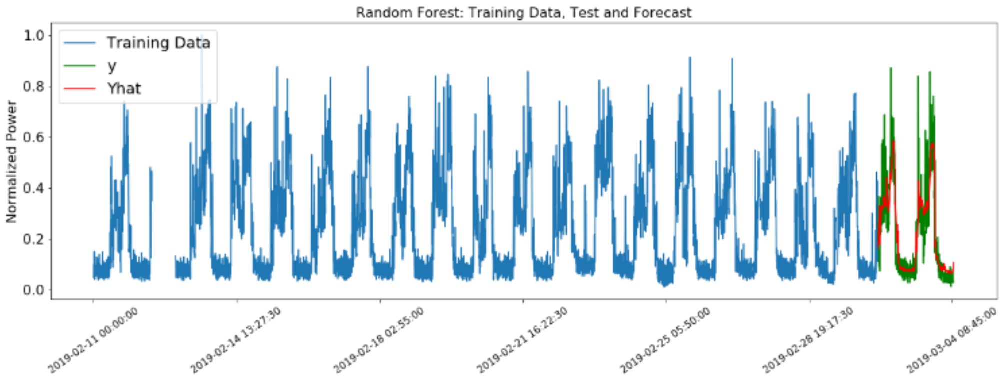

Figure 4 Virtual Aggregate, RF Training set, test set and forecast

For all three algorithms, better accuracy was achieved when forecasting for the virtual aggregate. This may be due to the unpredictability in the behaviour of the appliances.

Figure 5 shows the forecast of the virtual aggregate against the forecast of the sum of the individual appliances for the best performing algorithm, RF. As stated above, the error is bigger when forecasting the individual appliances than for the virtual aggregate. This is especially true when the restaurant is not operating, and the consumption is a valley. Here the forecast of individual appliances is overshooting the forecast and also creating a saw pattern (this occurs for both the RF and the LSTM, not for Prophet).

While these values in the forecast are a combination of the sum of the forecast of every appliance, most of this error comes from appliances with a "binary behaviour" that the model is uncapable of following accurately. This is best exemplified in the forecast of the salamander salamander (a specialized kitchen appliance that one will not commonly find outside of restaurants, it is used to grill, brown and toast faster than a common broiler) seen in Figure 6. Here it can be clearly seen the model is forecasting two different patterns of consumption at the appropriate time (when the appliance is on or off), but those patterns are not correct. While it should be a saturated value or zero it is creating this saw pattern that adds a lot of error when the restaurant is not operating. The LSTM forecast suffers from the same issue but in a lesser manner, the spikes are less steep.

VII. Results

The results for the forecast of the virtual aggregate and the sum of forecast for induvial appliances is shown in Table 2. The error of the forecast for all the algorithms was smaller when forecasting for the virtual aggregate rather than summing the forecast of the individual appliances.

There could be several reasons as to why the forecast performed better for the real load as opposed to the sum of the appliances, and ways to improve the performance.

Firstly, the amount of data available for this thesis was not very big (roughly one month data available). This affects especially the LSTM model which thrives on large datasets. While this issue affects at both the

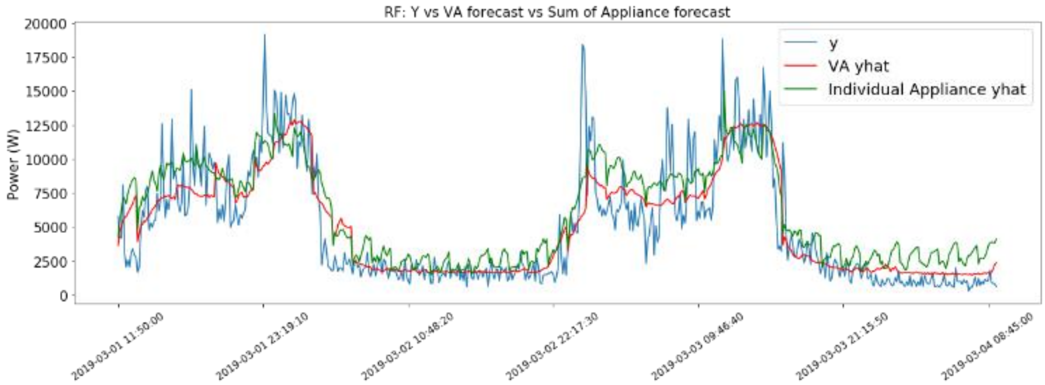

Figure 5 RF virtual aggregate forecast vs sum of individual appliance forecast

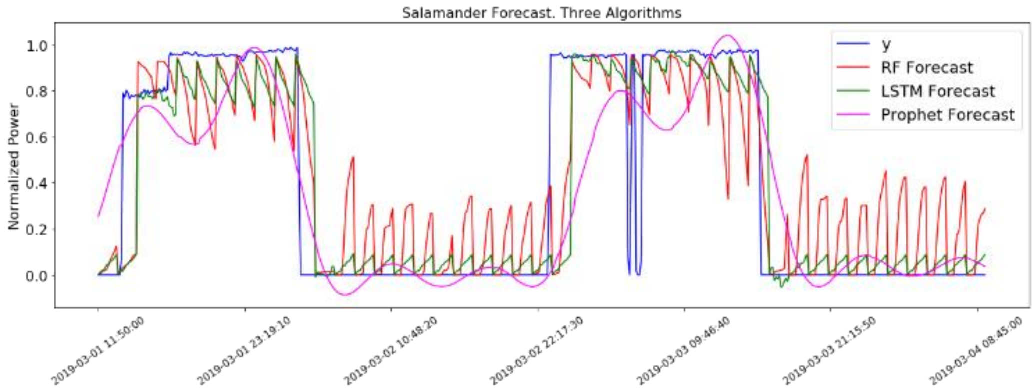

Figure 6 Salamander Forecast: Three algorithms

forecast of the virtual aggregate and the individual appliances the load of the IK is much more stable over time than the load most of the appliances. This load always follows the same pattern: peak values at lunch and dinner time followed by a valley after the next lunch time. While the consumption of many of the appliances are irregular in their behaviour (as they are tied to the specific situations that might arise in an IK).

Table 2 Performance metrics for all algorithms in both methods

	RMSE	MAE
H=1h Virtual Aggregate	2290	1499
H=1h Individual Appliance	2287	1507
RF: Virtual Aggregate	2230	1507
RF: Individual Appliance	2520	1998
LSTM: Virtual Aggregate	2461	1564
LSTM: Individual Appliances	2630	1961

This leads to second issue; the saw pattern seen in Figure 6 when the restaurant is not operating. This is an appliance with an almost binary behaviour that the algorithms fail to predict. All three models seem to predict when this appliance is on or off but do not correctly model the value.

VIII. Discussion: Prophet, Random Forest, and LSTM

All three algorithms used to implement this methodology present their own merits. Prophets' performance is peculiar, it is the easiest algorithm to implement, and its performance is almost identical when forecasting for the sum of individual appliance or the virtual aggregate. This makes it so that making the comparison of forecasts of the individual appliances using the three different methods Prophet outperforms the other two, measured both in RMSE (2287 (W) against 2520 (W) in RF) and MAE (1507 (W) against 1998 (W) in RF). However, comparing the accuracy of the forecast of the virtual aggregate is a tighter call. It presents the smallest MAE (1499 against 1507 in RF) but the second best RMSE (2290 against 2230 in RF).
RF shines when forecasting for the appliances. It has the best accuracy for 10 out of 17 of the appliances, more than the other algorithms combined (Prophet is worst in 10 out of 17). This makes it the best option when one would want to forecast for a concrete set of appliances. Also, Its performance on the virtual aggregate was one of the best.
LSTM was one of the weakest performing algorithms. It has the largest RMSE for both the virtual aggregate and the sum of individual appliances. It does however score better than RF when measured with MAE for the

sum of individual appliances. This is mainly because it does outperform RF when forecasting for some of the appliances with binary behaviour as can be seen in **¡Error! No se encuentra el origen de la referencia.** of the previous chapter. When forecasting for the individual appliance its performance was average, worst for three appliances and best for two. The low accuracy of this model can surely be related to the low amount of available data, on which LSTM thrives.

With the current results, Prophet and RF are the more viable options. Due to its good results and ease of implementation, Prophet seems the better choice and is certainly a good one. But RF has the best performance forecasting on the virtual aggregate and for single appliances. Having the best forecast accuracy for individual appliances can have some benefits that will be expressed in the following section. With a larger dataset it is possible that LSTM outperformed the other algorithms. Another possibility this methodology allows is to select for each appliance the algorithm that best performs. This might even improve the accuracy of the forecast of the individual appliances over the forecast of the virtual aggregate. This could be done in an automatic way if an algorithm was developed to do so, to select the optimal algorithm for each appliance based on historical data.

IX. Discussion: Virtual Aggregate vs Individual Appliances

From the results obtained it is not possible to conclude that the forecast was improved by forecasting for the individual appliances. All three algorithms performed better forecasting for the virtual aggregate. However, this does not mean there are no advantages to this new method of forecasting. To forecast for individual appliances allows the user to manually incorporate new appliances to the forecast (as a restaurant is a changing business) and only need to train a new model for the new appliance, while the data of the already existing appliances remains relevant. If the forecast were done for the virtual aggregate, the inclusion of a new appliance could potentially disrupt the new consumption pattern. This could make the older data and models more harm than good in future forecasts. This advantage is maintained when removing an appliance from the restaurant (for maintenance reasons for example), in this case one can simply remove the forecast for said appliance from the mix.

Another advantage of this kind of forecasting is that one can create forecasts for different consuming sectors of the restaurant. This means that one can agglomerate the consumption of the freezers, or the ovens for example and study consumption in this manner.

Finally, this methodology allows for the forecast of each appliance to be fine-tuned. Each appliance has its own characteristics and distinctive consumption pattern. This means each appliance can benefit from being forecasted with a different algorithm and even using different window sizes to train the models.

X. Limitations and future work

Certain limitations were met during the elaboration of this work. Predominantly the lack of data. The available data was collected in a Portuguese restaurant during roughly three weeks. This posed some problems at the moment of splitting the data into training set, validation set and testing set. A compromised had to be made and the percentage of data destined to test was smaller than usually recommended in the literature, but this allowed the models to train with at least two weeks' worth of data. The lack of data certainly hindered the performance of the LSTM model, which thrives on large datasets. To make this data viable lots of cleaning and resampling had to be done. This work could be expanded and most surely benefit by working on a larger dataset of IK consumption, in this case the forecast accuracy should improve for both the virtual aggregate and the individual appliances.

Another way to improve the accuracy of the forecast that this work did not cover is hyperparameter tuning. As discussed in the previous chapter, this work could also be improved and expanded by treating the forecast of some appliances with binary behaviour as a classification problem rather than regression. Also, different window sizes to train and test the data should be explored.

XI. References

[1] 'Electricity Market Report - January 2022 – Analysis', *International Energy Agency*, Jan. 2022. https://www.iea.org/reports/electricity-market-report-january-2022 (accessed May 20, 2022).

[2] O. Laib, M. T. Khadir, and L. Mihaylova, 'Toward efficient energy systems based on natural gas consumption prediction with LSTM Recurrent Neural Networks', *Energy*, vol. 177, pp. 530–542, Jun. 2019, doi: 10.1016/j.energy.2019.04.075.

[3] S. Mudie, 'Energy benchmarking in UK commercial kitchens', *Building Services Engineering Research and Technology*, vol. 37, no. 2, pp. 205–219, Mar. 2016, doi: 10/f8d998.

[4] S. Mudie, E. A. Essah, A. Grandison, and R. Felgate, 'Electricity use in the commercial kitchen', *Int J Low-Carbon Tech*, vol. 11, no. 1, pp. 66–74, Sep. 2013, doi: 10/gffms5.

[5] *Sector Guide Industrial Energy Efficiency Accelerator Contract Catering Sector*. Accessed: Dec. 26, 2019. [Online]. Available: https://www.carbontrust.com/media/163491/contract-catering-sector-guide-industrial-energy-efficiency-accelerator.pdf

[6] F. Higgins-Desbiolles, E. Moskwa, and G. Wijesinghe, 'How sustainable is sustainable hospitality research? A review of sustainable

restaurant literature from 1991 to 2015', *Current Issues in Tourism*, vol. 22, no. 13, pp. 1551–1580, Aug. 2019, doi: 10.1080/13683500.2017.1383368.

[7] S. Mudie and M. Vadhati, 'Low energy catering strategy: insights from a novel carbon-energy calculator', *Energy Procedia*, vol. 123, pp. 212–219, Sep. 2017, doi: 10.1016/j.egypro.2017.07.244.

[8] M. J. A. Shohan, M. O. Faruque, and S. Y. Foo, 'Forecasting of Electric Load Using a Hybrid LSTM-Neural Prophet Model', *Energies*, vol. 15, no. 6, p. 2158, Mar. 2022, doi: 10.3390/en15062158.

[9] T. Bashir, C. Haoyong, M. F. Tahir, and Z. Liqiang, 'Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN', *Energy Reports*, vol. 8, pp. 1678–1686, Nov. 2022, doi: 10.1016/j.egyr.2021.12.067.

[10] N. Vasudevan, V. Venkatraman, A. Ramkumar, and A. Sheela, 'Real-time day ahead energy management for smart home using machine learning algorithm', *IFS*, vol. 41, no. 5, pp. 5665–5676, Nov. 2021, doi: 10.3233/JIFS-189886.

[11] L. Breiman, 'Random Forests', *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[12] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, 'Modeling and forecasting building energy consumption: A review of data-driven techniques', *Sustainable Cities and Society*, vol. 48, p. 101533, Jul. 2019, doi: 10.1016/j.scs.2019.101533.

[13] G. Xue, C. Qi, H. Li, X. Kong, and J. Song, 'Heating load prediction based on attention long short term memory: A case study of Xingtai', *Energy*, vol. 203, p. 117846, Jul. 2020, doi: 10.1016/j.energy.2020.117846.

[14] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, 'Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network', *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019, doi: 10.1109/TSG.2017.2753802.

[15] T. Chai and R. R. Draxler, 'Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature', *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014, doi: 10.5194/gmd-7-1247-2014.