Workload Forecasting of Warehouse Stations: Comparison Between Classical Time Series Methods and XGBoost

İrem Kalafat[*, 1], Mustafa Hekimoğlu[1], Ahmet Deniz Yücekaya[1], Nilay Ay[2], Habib Gültekin[2]
[1]*Department of Industrial Engineering, Kadir Has University, İstanbul, Turkey*
[2] *Dogus Technology, İstanbul, Turkey*

***Abstract*— Effective management of warehouse processes is essential in order to maintain high-level service quality and keep the costs at optimum. Each item passes through numerous workstations during their journey in warehouses from the entrepot to the shipping area. Accurate estimation of workload at stations allows personnel assignment optimization and the increase of the warehouse performance. Otherwise, it causes personnel shortages at stations, delays in shipment commitment dates and disruptions in warehouse activities. In this paper, time series forecasting models are used to estimate the load in each workstation for a better operation. The proposed methodologies are applied to an automotive spare part warehouse in Turkey. The classical time series method, which performs best in estimating the workload of each workstation, is presented and these results are compared with the XGBoost model. Thus, the models that give the best results for each station are shown. The proposed research covers part acceptance, storage, order picking and packaging processes and their sub-stations, which were not considered in previous studies.**

***Index Terms*— Machine Learning, Warehouse Management, Workload Prediction, Time Series, XGBoost**

I. INTRODUCTION

In its simplest definition, the warehouse is an area where items are kept in the time period between acceptance and shipment in accordance with the time-dependent demands. Therefore, it is an essential link between distributor and buyer in the continuous, productive execution of operations from product supply to shipment. Through proper management of warehouses, it is possible to deliver customer orders in a short time with the least cost. The fact that it is associated with cost minimization, customer satisfaction, and market competition, warehouses are strategically crucial for companies.

Fulfilling workload in the shortest time with a quality manner and assigning sufficient personnel to appropriate places are some of the main challenges warehouse managers face constantly. While having more than necessary personnel in stations causes high costs, the insufficient number of staff will also result in low productivity, employee fatigue, and reduced profits [1]. Furthermore, since the demands coming to the distribution centers are uncertain, and the amount of workload in operations tends to change with time, it is difficult to determine the required workforce [2]. To manage these problems, it is necessary to accurately forecast the daily workload in each workstation and schedule the personnel accordingly.

The aim of the paper is to present time series forecasting models and compare the performance metrics of the approaches by taking daily tasks and types of workload as input. To the best of our knowledge, we are the first to forecast the workloads in warehouse sub-stations of the receiving, storing, packaging next to order picking. We have applied different kinds of time series methods, which are considered classical in the literature, to each sub-stations' data and compared the best-performing method with a more advanced machine learning method. Due to the different characteristics of each sub-station, we have concluded that each one performs differently. Also, according to our results, while storage workload prediction is the most difficult among other operations, in general, the packaging is the most accurately predicted station. In addition, we can say that stations with small volumes of data are more difficult to predict. Next, we present a review of existing literature on workload forecasting and problem description. After that, a brief explanation of models, and followed by the results, are given. Finally, we end the paper by showing our conclusions.

II. LITERATURE REVIEW

Information of future workload in each operation will affect the strategic decisions in business processes. The most costly activity on a workstation basis is order picking, accounting for approximately 55% of expenditures. Among the operations, the most common subject in the literature has been order picking [3]. This paper covers the other operations in warehouses such as; part acceptance, storing, order picking, and packaging operations and their sub-stations. Although demand forecasting in the supply chain is a continuously studied area in literature, workload forecasting in warehouse workstations almost has not received any attention. [4] compared support vector machines, neural networks, recurrent neural networks, and more straightforward forecasting methods like naive forecast, trend, moving average, and multiple linear regression to predict supply chain demand. They have concluded that advanced methods give better results. [5] presents a heuristic algorithm that balances the workload

Manuscript received July 19, 2021; accepted June 05, 2022.
*Corresponding author: irem.kalafat@stu.khas.edu.tr
This study is funded by TUBITAK 2244-Industrial PhD Fellowship Program with grant number 119C085.

distribution between order pickers, considering the amount of work in different picking zones. However, this approach is suitable in the long-run estimations. [6] applies bucket-bridge method to the order-picking system to balance the workload between order pickers in the short term. Investigates the performance of the bucket bridges and shows that they can cause efficiency loss. [7] states that actual workload forecasting results in workload balancing. Their work is the first to forecast order pickers' workloads and also proposed several statistical forecasting models and ensemble models of exponential smoothing, SARIMA, and ARIMAX to predict expected order lines for different order picking zone levels. [8] considered bias, the combination of professional forecasts and historical sales information to boost demand forecasts and confirmed an explicit level of forecast bias is enforced to optimize labor resource planning. [9] studied on predicting the needed workload in Maintenance, Repair, and Overhaul (MRO) companies. They have used simple exponential smoothing, Holt's linear method, additive Holt-Winters, and multiplicative Holt-Winters methods and developed a decision support system for capacity planning according to forecasted workloads.

There are also other time series workload forecasting studies in different application areas. [10] proposed batch trained ensemble of an expert model to predict radiologists' workload using time series data of a medical system. [11] presented a Deep Multi-Task Learning approach to forecast the workload of patients in healthcare systems. The motivation of forecasting in their study is to balance the workload between teams to increase efficiency in resource management. An ensembled approach for predicting the 7 days ahead workloads in local health department is proposed in [12]. They have compared the results of combined version of the XGBoost, random forest (RF) and LSTM methods with the separated results of these methods. Power systems load estimation with ensemble methods such as extreme gradient boosting (XGBoost) and random forests (RF) is also quite popular in literature. [13] compared standard statistic methods (SARIMA and SARIMAX) with ensemble algorithms (Random Forests and Gradient Boosting) for predicting electricity load. They present the performance metrics of the proposed techniques, and according to their case, ensemble models perform better than SARIMA and SARIMAX. [14] analyzed three approaches for shot-term power load prediction. The authors made a cluster analysis to obtain similar behaved days and applied the XGBoost technique. Then used prediction result to show that it gives higher accuracy than LSTM (Long Short Term Memory). [15] also proposed a study of forecasting the electricity load with XGBoost on weekly data and concluded that it had not worked well on predicting high loads. The authors of [16] presented an ML-based model to estimate the workload and energy usage in cloud data centers. Linear regression, ridge regression, ARD regression, elastic net regression and gated recurrent unit which is a deep learning algorithm is used for workload forecasting part of their study. [17] studied nature-inspired algorithms for estimating the workload in cloud centers. They have compared the results according to root mean square error (RMSE) and mean absolute error (MAE) error metrics.

Our aim is to extend the studies in literature by applying classical time series forecasting methods and XGBoost on not only order picking operation but also receiving, storing, packaging, and their sub workstations.

III. Problem Description

The case study is applied to an automotive spare parts warehouse in Turkey. Many warehouses divide the works that are performed into two main categories: inbound operations and outbound operations. While the former consists of receiving and storing the items, the latter includes order picking, packaging, and shipping [18].

Inbound operations start with delivering the orders, which have arrived from Original Equipment Manufacturer (OEM) to the warehouse, to the bonded warehouse for completing their import procedures. After completing these operations, they are carried out in cages from bonded to the central warehouse. Next, spare parts are transported and removed from the cages, and the receiving process begins. Finally, the parts whose barcodes are scanned by hand terminals are loaded to the storing vehicles that will go to the locations where they will be placed. Storing is the final process of inbound operations.

The outbound operations start when the orders from the authorized dealer are approved. In order to collect the parts that are sold, picking labels are printed with information about the order type and picking method are on them. The printed labels are collected by picking personnel, and picking processes from the relevant zones are performed.

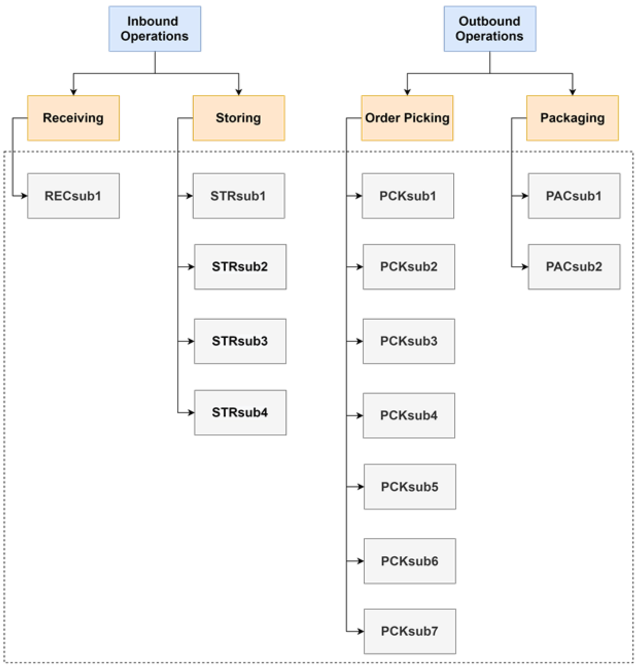

Fig. 1. Hierarchy scheme of operations

Then, the collected parts are filtered by authorized dealers and transported to the packaging areas. Here, the packaged spare parts are loaded on the shipping vehicles and delivered. After each transaction in the warehouse, barcodes are scanned with hand terminals for the relevant invoice item. In this way, the data on which operation was performed on which day at what time is stored. There are zones belonging to different part types in the warehouse.

Considering these areas, all four tasks (receiving, storing, order picking, and packaging) are divided into sub-tasks. See the Figure 1 hierarchy scheme for operations.The operations inside of the dotted line in the figure represent sub-stations. We have given shortcode names to sub-stations for easier understanding. Currently, the daily workload amount forecasting at the stations and the personnel assignments to these stations are made based on experience, depending on past trends by the warehouse managers. This situation causes personnel deficiencies in stations, deviation from the dispatch commitment, and disruption in warehouse activities. Therefore, we aim to present models to predict the daily workload at each sub-station and decide the best performing models. This will lead to optimizing the personnel assignments.

IV. DATA DESCRIPTION AND PRE-PROCESSING

The amount of workload completed varies depending on the month, quarter of the year, day of the week, national or religious holiday, and whether it occurs in the last and first 15 days of the year. Extreme situations can be observed in the warehouse at year-end closings and in the first days of the year due to the delivery of orders and administrative factors. We will later refer to this situation as a busy and regular period. In the feature engineering process, we have added these variables as dummy variables to our data. We updated the raw data so that the actual workload amounts for each workstation are new columns. Thus, the values in the date column are unique for each feature. In addition to this arrangement, we determined extreme observations which distort data. While we extracted these outlier values from the data in naive, SES, Holt's linear, and ARIMA methods, which we will explain in the next section, we used them as dummy variables in ARIMAX and XGBoost models. With these new multivariate time series data frame where each sub-station is a time series, we tested each series' stationarity with Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test [19]. While some sub-stations need differencing to make the given time series stationary, some do not.

V. METHODOLOGY

This section briefly explains some of the classical time series and boosting methods used in predicting the target variable SKU. These methods include (ARMA, ARIMA, ARIMAX, Holt's linear trend method, simple exponential smoothing, and naïve method) and XGBoost. Later, the six days ahead (which is equal to a week, since there is no work on Sundays) forecast results of the models and their performance metrics are shown.

A. *Classical time series methods and XGBoost*

The naïve method is a simple method that does not consider trend or seasonality and forecasts the value based on the last observation. It has been used as a benchmark method in this paper. It can be represented with the following notation.

$$y_{t+1} = x_t \tag{1}$$

x_t:actual value in period t

y_{t+1}:forecasted value in period t+1

Since we forecast next week's workloads on a daily basis, we adjusted the naive method equal to the average of last week's SKUs.

Two exponential smoothing models have been presented in this study. The first one is simple exponential smoothing (SES) which is suitable for stationary series with no clear trend and seasonality [20]. SES expands the naïve method and states that forecasts not only depends on the last observation but also on all previous observations with some decaying importance weight.

$$y_{t+1}=\alpha x_t+\alpha\left(1-\alpha\right)x_{t-1}+\alpha\left(1-\alpha\right)^2x_{t-2}+\alpha\left(1-\alpha\right)^3x_{t-3}\ldots \tag{2}$$

where $0\leq\alpha\leq1$

α:smoothing parameter for level

Since SES is not handy in non-stationary time series data, we will present another method called Holt's linear exponential smoothing . This method is applicable to series which has trend but no seasonality or cyclical patterns. Holt's linear method involves two smoothing parameters, one for level and the other for the trend [21].

$$y_{t+p}= L_t+pT_t \tag{3}$$

$$L_t= \alpha y_t+(1-\alpha)(L_{t-1}+T_{t-1}) \tag{4}$$

$$T_t= \beta(L_t-L_{t-1})+(1-\beta)T_{t-1} \tag{5}$$

where $0\leq\beta\leq1$ and $0\leq\alpha\leq1$

β:smoothing parameter for trend

L_t:level of the series at time t

T_t:trend of the series at time t

The autoregressive moving average method (ARMA) is the combination of AR(p) and MA(q). It is appropriate for stationary univariate time series. AR's p value stands for the order of the autoregressive model, and MA's q value is the order of the moving average model. They can be obtained from partial autocorrelation graphs (AR(p)) and autocorrelation (MA(q)). These orders will respectively tell, in which order the forecasted value y_t can be explained by its past values and error terms. For our case, p and q values were captured as ARMA(6,1), which means that it is modeled by six past values(lag) and one error term.

$$y_{t+1}=c+\sum_{i=1}^{p}\varphi_i x_{t-i+1}+\sum_{j=1}^{q}\theta_j\epsilon_{t-j+1} \tag{6}$$

c:the intercept of the model

ϵ_t:random error in time t (white noise)

φ:coefficients of the AR terms

θ:coefficients of the MA terms

While ARMA only considers stationary series, the ARIMA method can deal with non-stationary data. In addition to ARMA(p,q) method ARIMA has another coefficient d which is the number of differencing to make data stationary and stabilize the mean. ARIMA(p,1,q) can be described with the following notation.

$$y_{t+1} = c + \sum_{i=1}^{p} \varphi_i (x_{t-i+1} - x_{t-i}) - \sum_{j=1}^{q} \theta_j \epsilon_{t-j+1} \tag{7}$$

Aside from lagged values and error terms, there is an extension of the before-mentioned methods called ARMAX and ARIMAX. These techniques take into account the effects of exogenous variables on the target variable.

$$y_{t+1} = x_t + \sum_{k=1}^{u} \rho_k \tau_k + \sum_{i=1}^{p} \varphi_i (x_{t-i+1} - x_{t-i}) - \sum_{j=1}^{q} \theta_j \epsilon_{t-j+1} \tag{8}$$

In our case, we have considered the first six lagged values, month, year, day of the week, Ramadan, and public holidays, and whether the date is in busy or regular period as covariates. These standard statistical methods have been compared to a more up to date method called XGBoost. XGBoost is an ensemble decision tree-based model. Rather than examining every record, XGBoost divides the data into weak learners and works according to those until it evolves to a strong learner. Furthermore, due to its high execution speed and ability to explain the relation between features and target, it is used in load forecasting. More detail about the algorithm can be found in [22].

VI. RESULTS

The explained classical time series models have been applied to each sub-station. We split our data to train and test. The last six observations are considered as the test set, and the rest of the observations are used to train models. A weekly workload forecast is obtained for sub-stations. The root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) values of the best performing models are shown in Table 1 with comparing to naive method's results. We can say that ARIMAX(6,1,1) performed better in most of the sub-stations. As mentioned, SES, Holt's Linear Trend, ARMA, ARIMA, and ARIMAX can only work with univariate time series data. Thus we needed pre-processing step for them. On the other hand, the raw data format is suitable for building XGBoost model cause it handles sub-stations as categorical variables and works fine with multivariate series. Moreover, the exogenous variables used in ARIMAX are also considered in the XGBoost model. One of the significant aspects of this ensemble method is that it enables to see which features in the data have more importance on the target value. We have seen in our XGBoost model that lag values and the dummy variables based on outliers have the highest effect.

The overall test sets performance metrics are shown in Table 2.

TABLE II. TEST SET'S PERFORMANCE METRICS OF XGBOOST

Model	Performance Metrics		
	RMSE	MAE	MAPE
XGBoost	238.16	156.3	18.56

In order to provide more a better comparison between all of the presented methods we have divided the overall XGBoost performance into sub-station levels. Table 3 shows the best performing model among all the methodologies mentioned. Considering the MAPE values, the following conclusions can be made; while storage can be interpreted as the worst among other operations, in general, the packaging is the most accurately predicted station. In addition, we can say that stations with small volumes of data are more difficult to predict.

The training and competencies required for various operations in the warehouse differ from each other. Even if personnel assignment changes are made between stations, the primary purpose is to keep the productivity of the personnel at their main stations at the maximum level. For this reason, the importance of the workload estimation study for each station is essential. Our study contributes to the literature by taking into account the other main stations (receiving, storing, packaging) and their sub-stations that affect the cost in the warehouse apart from order picking and by comparing the performances of the classical time series methods and the XGBoost model through a real case study. Due to the characteristic workload characteristics of the sub-stations, since the time series data of each is different, the classical time series methods were successful

TABLE I. TEST SET'S PERFORMANCE METRICS OF CLASSICAL TIME SERIES MODELS

Sub-Station	Best Performed Method	Best Performed Model's Metrics			Naïve Model's Metrics		
		RMSE	MAE	MAPE	RMSE	MAE	MAPE
PACsub1	ARIMAX(6,1,1)	553.78	470.08	14.03	1036.42	550.56	29.99
PACsub2	ARIMAX(6,1,1)	851.55	654.24	11.03	1979.58	1062.54	47.73
PCKsub1	ARIMAX(6,1,1)	166.44	141.26	8.47	404.29	290.87	22.00
PCKsub2	ARIMAX(6,1,1)	377.29	329.63	13.51	593.87	434.96	24.56
PCKsub3	ARIMAX(6,1,1)	67.24	54.62	17.66	96.06	75.83	28.21
PCKsub4	ARIMAX(6,1,1)	338.17	293.29	7.24	1494.17	906.26	58.83
PCKsub5	ARIMA(6,1,1)	41.68	30.30	12.10	56.35	44.65	18.35
PCKsub6	Holt's Linear Trend Method	15.76	13.11	14.18	22.64	17.52	16.53
PCKsub7	ARIMAX(6,1,1)	45.60	41.53	6.15	185.92	140.68	30.11
RECsub1	ARMA(6,1)	722.47	455.91	21.53	770.20	705.53	41.93
STRsub1	ARMAX(6,1)	90.53	71.85	10.94	272.99	190.33	28.46
STRsub2	Holt's Linear Trend Method	295.37	173.57	22.91	319.25	199.36	26.84
STRsub3	SES	152.86	112.17	24.19	232.67	189.12	39.49

in some, while the more advanced machine learning model gave better results in others.

TABLE III. TEST SET'S PERFORMANCE METRICS OF BEST PERFORMED MODEL'S AMONG CLASSICAL TIME SERIES MODELS AND XGBOOST

Sub-Station	Best Performed Method	Best Performed Model's Metrics		
		RMSE	MAE	MAPE
PACsub1	XGBoost	171.95	134.33	3.70
PACsub2	XGBoost	258.03	216.17	3.97
PCKsub1	XGBoost	154.42	110.33	6.23
PCKsub2	XGBoost	243.73	195.00	9.87
PCKsub3	ARIMAX(6,1,1)	67.24	54.62	17.66
PCKsub4	ARIMAX(6,1,1)	338.17	293.29	7.24
PCKsub5	ARIMA(6,1,1)	41.68	30.30	12.10
PCKsub6	Holt's Linear Trend Method	15.76	13.11	14.18
PCKsub7	ARIMAX(6,1,1)	45.60	41.53	6.15
RECsub1	XGBoost	137.43	108.67	13.06
STRsub1	ARIMAX(6,0,1)	90.53	71.85	10.94
STRsub2	Holt's Linear Trend Method	295.37	173.57	22.91
STRsub3	SES	152.86	112.17	24.19

VII. CONCLUSIONS

In this paper we daily forecasted the workloads in warehouse sub-stations for the next week. We have compared the results between classical univariate time series models and a machine learning advance technique XGBoost. Although previous studies have shown that XGBoost performs higher accuracy than ARIMA or ARIMAX, for this case study, we can conclude from the results that while some stations get better results from the ensemble model, others are not. This study explained the importance of estimating workloads at various warehouse stations, what kind of data we have, and the models we use. Although, this study shows stationary classic time series models, seasonality has not yet been taken into account. As a next step, the results of seasonal models such as SARIMA, SARIMAX and Holt Winters on workload estimation are planned to be added to the study. Also, we aim to enrich our data with new features such as; orders from authorized dealers, number of SKUs pending for acceptance in the entrepot. Further studies also can involve the assignment of the personnel to the stations using the future workload forecasts as input. In order to optimize workforce management in the future, models can be developed to optimize the number of personnel needed at stations and give optimum shift suggestions, using future workload forecasts as input. With these improvements, profitability in warehouse management can be increased by minimizing labor expenses.

ACKNOWLEDGMENT

We offer our respect and gratitude to Dogus Automotive for all their help and contributions.

Part of this study was presented orally in the Fourth International Conference on Data Science and Applications 2021 (ICONDATA'21).

REFERENCES

[1] Smith, J. The warehouse management handbook. Tompkins Press, 1998.

[2] Wiers, S. d. Warehouse manpower planning strategies in times of financial crisis: evidence from logistics service providers and retailers in the Netherlands. Production Planning & Control, 328-337, 2015.

[3] Rene´ de Koster, T. L.-D. Design and control of warehouse order picking:A literature review. European Journal of Operational Research 182, 481–501, 2007.

[4] Real Carbonneau, K. L. Application of machine learning techniques for supply chain demand forecasting. European Journal of Operational Research 184, 1140–1154, 2008.

[5] Chin-Chia Jane, Y.-W. L. A clustering algorithm for item assignment in a synchronized zone order picking system. European Journal of Operational Research 166, 489–496, 2005.

[6] Koo, P.-H. The use of bucket brigades in zone order picking systems. OR Spectrum 31(4), 759-774, 2009.

[7] Teun Van Gils, K. R. The Use of Time Series Forecasting in Zone Order Picking Systems to Predict Order Pickers' Workload. International Journal of Production Research, Vol. 55 No. 21, 6380-6393, 2017.

[8] Thai Young Kim, R. D. Improving warehouse labour efficiency by intentional forecast bias. International Journal of Physical Distribution & Logistics Management, 2018.

[9] Dinis, D., Barbosa-Póvoa, A., & Teixeira, Â. P. Enhancing capacity planning through forecasting: An integrated tool for maintenance of complex product systems. International Journal of Forecasting, 38(1), 178-192, 2022.

[10] Tasquia Mizan, S. T. A causal model for short-term time series analysis to predict incoming Medicare workload. Journal of Forecasting, 228– 242, 2021.

[11] Olya, M. H., Badri, H., Teimoori, S., & Yang, K. An integrated deep learning and stochastic optimization approach for resource management in team-based healthcare systems. Expert Systems with Applications, 187, 115924, 2022.

[12] Piccialli, F., Giampaolo, F., Salvi, A., & Cuomo, S.. A robust ensemble technique in forecasting workload of local healthcare departments. Neurocomputing, 444, 69-78, 2021.

[13] S. Papadopoulos, I. K. IEEE Power and Energy Conference at Illinois (PECI). Short-term Electricity Load Forecasting using Time Series and Ensemble Learning Methods, 1-6, 2015.

[14] Liao, X., Cao, N., Li, M., & Kang, X. International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). Research on Short-Term Load Forecasting Using XGBoost Based on Similar Days, 675-678, 2019.

[15] Raza Abid Abbasi, N. J. Short Term Load Forecasting Using XGBoost. Advances in Intelligent Systems and Computing, 1120-1131,2019.

[16] Khan, T., Tian, W., Ilager, S., & Buyya, R. Workload forecasting and energy state estimation in cloud data centres: ML-centric approach. Future Generation Computer Systems, 128, 320-332, 2022.

[17] Kumar, J., & Singh, A. K. Performance evaluation of metaheuristics algorithms for workload prediction in cloud environment. Applied Soft Computing, 113, 107895, 2021.

[18] Bartholdi, J. a. Warehouse & Distribution Science,2019.

[19] Kwiatkowski, Denis & Phillips, Peter C. B. & Schmidt, Peter & Shin, Yongcheol, Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series

have a unit root?. Journal of Econometrics, Elsevier, vol. 54(1-3), pages 159-178,1992.

[20] Hyndman, R. J., & Athanasopoulos, G. Forecasting: Principles and Practice. 2nd ed. Otexts, 2018.

[21] Holt, C. C. Forecasting seasonals and trends by exponentially weighted averages. O.N.R. Memorandum No. 52, 1957.

[22] Tianqi Chen, C. G. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785-794, 2016.