

Benchmarking of Regression and Time Series Analysis Techniques for Sales Forecasting

C. CATAL, K. ECE, B. ARSLAN and A. AKBULUT

Abstract— Predicting the sales amount as close as to the actual sales amount can provide many benefits to companies. Since the fashion industry is not easily predictable, it is not straightforward to make an accurate prediction of sales. In this study, we applied not only regression methods in machine learning but also time series analysis techniques to forecast the sales amount based on several features. We applied our models on Walmart sales data in Microsoft Azure Machine Learning Studio platform. The following regression techniques were applied: Linear Regression, Bayesian Regression, Neural Network Regression, Decision Forest Regression and Boosted Decision Tree Regression. In addition to these regression techniques, the following time series analysis methods were implemented: Seasonal ARIMA, Non-Seasonal ARIMA, Seasonal ETS, Non -Seasonal ETS, Naive Method, Average Method, and Drift Method. It was shown that Boosted Decision Tree Regression provides the best performance on this sales data. This project is a part of the development of a new decision support system for the retail industry.

Index Terms—Machine learning, regression, sales forecasting, time series analysis.

I. INTRODUCTION

THE IDENTIFICATION of the number of stocks and the replenishment strategy are significant activities for many companies in the retail industry.

If the number of the products is insufficient at a given time, the customer demand cannot be satisfied at that time which causes the company to lose the customer. If there are a

CAGATAY CATAL, is with Information Technology Group Wageningen University, Wageningen, The Netherlands, (e-mail: cagatay.catal@wur.nl).

 <https://orcid.org/0000-0003-0959-2930>

KAAN ECE, is with Department of Computer Engineering Istanbul Kültür University, Istanbul, Turkey, (e-mail: 1101020039@stu.iku.edu.tr).

 <https://orcid.org/0000-0001-7225-047X>

BEGUM ARSLAN, is with Department of Computer Engineering Istanbul Kültür University, Istanbul, Turkey, (e-mail: 1201020009@stu.iku.edu.tr).

 <https://orcid.org/0000-0002-4794-0791>

AKHAN AKBULUT, is with Department of Computer Engineering Istanbul Kültür University, Istanbul, Turkey, (e-mail: a.akbulut@iku.edu.tr).

 <https://orcid.org/0000-0001-9789-5012>

Manuscript received December 10, 2018; accepted January 17, 2019.
DOI: [10.17694/bajece.494920](https://doi.org/10.17694/bajece.494920)

sufficient number of products, but there is no potential customer to buy them, then the product stays in stocks. In addition, the fashion industry is volatile. Trends change quickly. Fashion must be closely followed to increase the sales amount.

A company which is out of the trends is not preferred by the customers. Therefore, features which affect the fashion can be examined to increase the sales. In this study, our goal is to predict the actual sales amount accurately by using different machine learning algorithms. Machine learning is a very active research field that helps to learn from the data and uses the data to make a prediction for the future. There are many application areas of machine learning algorithms. For example, Facebook's News Feed feature, which applies the EdgeRank algorithm, can be used for the personalization of the feeds. The algorithm identifies the user interests by using statistical and predictive analysis.

Walmart, one of the best retailers in the world, dramatically increased its online sales and revenue by using advanced data mining techniques. The data prior to the sales and after the sales have been extensively analyzed by the data scientists to change the e-commerce strategy of this retail company. Also, Walmart has changed its shipping policy for the products based on the data analysis performed on big data. According to Walmart's new shipping policy, the minimum amount for free shipping was raised from \$45 to \$50.

In this study, Walmart's public data was analyzed with different regression algorithms in Azure Machine Learning (ML) Studio. In addition to these algorithms, several time series analysis methods were implemented by using R packages which are available from Azure ML Studio. Since there is no way to add time series analysis methods into the experiment screen graphically, these methods were implemented in the R programming language manually. Later, the best model was transformed into a web service and this web service was deployed on the Azure cloud platform. A client application was implemented to consume this web service. Azure sends the results in JSON format.

The following regression algorithms were applied: Linear Regression, Bayesian Regression, Neural Network Regression, Random Forest Regression, and Boosted Decision Tree Regression. Also, the following time series analysis techniques were applied: Seasonal ARIMA, Non-Seasonal ARIMA, Seasonal ETS, Non-Seasonal ETS, Naive Method, Average Method, and Drift Method. According to the experimental results, the best method was identified. It was

observed that the regression techniques provide better performance compared to the time series analysis approaches. In this study, it was demonstrated that building an end-to-end sales prediction system on Azure ML Studio is an easy and very efficient task and there are many algorithms to apply.

The remainder of this paper is organized as follows: Section II presents the related work. Section III shows the methodology and the Section IV explains the initial results of this system. Section V shows the conclusion and future work.

II. RELATED WORK

There are many studies on the development of sales forecasting models, but they did not evaluate many models in one study and use Azure Machine Learning Studio platform to build an end-to-end sales prediction system so far.

Kuo [1] applied the fuzzy-neural network for sales forecasting and demonstrated that this model's performance is superior to traditional neural networks. Chen and Ou [2] used a new approach called Gray extreme machine learning with Taguchi method [2]. The system performance was better than the performance of artificial neural networks. Zhao et al. [3] utilized from clustering, regression, and time series analysis techniques for the electricity sales forecasting. Tian et al. [4] applied seasonal time series analysis for auto sales in China. The seasonal effects were calculated by using the exponential weighted moving average. Later, calculated effects and the counted frequency are combined for the linear regression technique. Zhang [5] combined ARIMA and neural network for forecasting. Pandey and Somani [6] implemented a cloud computing based sales forecasting system and applied time series analysis with the moving average method. They deployed the system on Azure cloud and used MySQL database and PHP programming language. Moving average methods are being applied to forecast the sales for a long time in the literature [7]. Vijayalakshmi et al. [8] implemented a sales forecasting engine based on genetic algorithms. Yeo et al. [9] developed a new customer model which uses customer-browsing behavior and tested the model on an e-commerce website. Choi et al. [10] combined the SARIMA and wavelet transform method for sales forecasting. They demonstrated that the new hybrid model provides better performance than the single methods. Chang et al. [11] designed a hybrid model, which combines k-means clustering and fuzzy neural network for sales prediction of circuit boards. Wong et al. [12] implemented a new model, which uses extreme learning machine and harmony search algorithm for sales prediction of retail supply chains and showed that the new model provides better performance than the ARIMA models. Katkar et al. [13] used fuzzy logic and Naive Bayes classifier for sales forecasting. Müller-Navarra et al. [14] applied Recurrent Neural Networks to forecast the sales. Gao et al. [15] used extreme learning machine algorithm.

Omar et al. [16] examined the Back Propagation Neural Network for sales forecasting. Lu et al. [17] proposed a hybrid method based on MARS and SVR techniques for sales prediction of information technology products. They

demonstrated that the new model provides better performance than the single SVR. Stojanović et al. [18] used several features such as fuel price, holiday, unemployment, temperature, store, and date to forecast the weekly sales in Walmart and showed that Support Vector Machine provides the best performance.

As we see in these studies, each study suggests a single model, but we need a comparative assessment of machine learning models to evaluate their performance on the same public dataset. In this study, we performed our experiments to satisfy this goal.

III. METHODOLOGY

In this study, several regression algorithms in machine learning and time series analysis methods were applied for sales forecasting. In this section, these methods are introduced. First, regression algorithms will be explained and then, the time series analysis methods will be introduced.

A. Regression

Linear Regression

Linear Regression is used to create a mathematical equation to produce the relation between independent variables (x) and a dependent variable (y).

B: Coefficients

E: Residue

Linear regression method uses the Formula 1:

$$Y = B_0 + B_1 + B_2 X_2 + \dots + B_k X_k + E_i \quad (1)$$

The slope of the line is B, and E represents the intercept. Therefore, Y is the response variable which is also called the dependent variable, B's are the weights that are the model parameters, the values of the predictor variables are represented with X's, and finally E is the error term signifying the random sampling noise.

Bayesian Linear Regression

Unlike linear regression, Bayesian Inference is used in the Bayesian approach. The normal distribution in Bayesian approach is calculated based on the Equations 2-3-4-5. Since w is a continuous-valued random variable in R^d , Bayes rule says that the posterior distribution of w is given by y .

$$P(w|y) \propto P(y|w) P(w) \quad (2)$$

$$P(w|y) \sim N(\mu, S) \quad (3)$$

$$S^{-1} = S_0^{-1} + \frac{1}{\sigma^2} x^t \quad (4)$$

$$\mu = S (S_0^{-1} \mu_0 + \frac{1}{\sigma^2} x^t y) \quad (5)$$

In Bayesian linear regression, the predictive distribution is calculated based on Equation 6:

$$P(y_{new}|y, X, x_{new}, \sigma^2) = \int P(y_{new}|w, X, x_{new}, \sigma^2)P(w|X) dw \quad (6)$$

Predictive distribution evaluates the likelihood of a value y_0 given x_0 for a particular w , by means of likelihood by current belief about w given data (y, X) . Finally, sum up all possible values of w .

Neural Network Regression

The neural network regression uses Equation 7:

$$\sum_{i=0}^M w_i x_i = w.x \quad (7)$$

In Neural Networks, a perceptron is used to take a vector of real-valued inputs to calculate the linear combinations of inputs. If the output is greater than some threshold, then it outputs 1, in the other case it produces -1. The weights have to be calculated according to the perceptron training rules shown in Equation 8 and 9:

$$w_i \leftarrow w_i + \Delta w_i \quad (8)$$

$$\Delta w_i = \eta(t - o) x_i \quad (9)$$

The symbol η refers to the neural network learning rate.

Decision Forest Regression

In the Random Forest algorithm, the primary aim is to make a classification by using several trees. In Random Forest, to obtain the last class of the tree, Gini value is used. Gini value is calculated based on Formula 10:

$$\text{Gini}(T) = 1 - \sum_{j=1}^n (p_j)^2 \quad (10)$$

The T dataset is split into T_1, T_2 subsets with N_1, N_2 dimensions, then Gini split value is calculated based on Formula 11:

$$\text{Gini}_{\text{split}(T)} = \frac{N_1}{N} \text{Gini}(T_1) + \frac{N_2}{N} \text{Gini}(T_2) \quad (11)$$

Boosted Decision Tree Regression

Boosted decision trees use gradient boosting algorithm. This algorithm applies the optimization of differentiable loss function by using the weighted sum of functions. The $F(x)$ is calculated based on Equation 12:

$$F_0(x) = \arg \min \sum_{i=1}^n L(y_i, \gamma) \quad \text{and} \quad (12)$$

The predictions are calculated based on Equation 13:

$$\gamma_m = \arg \min \sum_{i=1}^n L(y_i, F_{m-1}(x_i) - \gamma \frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}) \quad (13)$$

B. Time Series Analysis

ARIMA

ARIMA (autoregressive integrated moving average) model uses the following Equations 14-15 and describes the autocorrelations in the data.

$$Y_t = (1 - L)^d X_t \quad (14)$$

$$Y_t \left(1 - \sum_{i=1}^p \Phi_i L^i\right) = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t \quad (15)$$

The parameter L is the lag operator, p is the order which represents the number of time lags of the autoregressive model, and q is the order of the moving-average model. θ are the parameters of the moving average part and Φ are the parameters of the autoregressive part of the model.

Seasonal ARIMA

Seasonal ARIMA (SARIMA) is similar to ARIMA, but it has different elements. SARIMA uses Formula 16-17-18.

$$\Phi(B) \Delta^d X_t = \theta(B) \alpha_t \quad (16)$$

$$\Phi_s(B^s) \Delta_s^D \alpha_t = \theta_s(B^s) \alpha_t \quad (17)$$

$$\Phi(B)_s \Phi(B^s) \Delta_s^D \Delta^d X_t = \theta(B_s) \theta(B^s) \alpha_t \quad (18)$$

SARIMA model has the same structure as the non-seasonal (ARIMA) model: it may have an AR factor, an MA factor (corresponds to α), and/or an order of differencing.

Seasonal ETS

The Seasonal Exponential Smoothing (Seasonal ETS) applies three sub-pass filters recursively with special exponential window functions. The simplest moving averages are the weighting of past observations, but the exponential window functions are multiplied over time to reduce the weight. The simplest formula of seasonal ETS is shown in Formula 19 where α is the smoothing factor and s is the seasonal period.

$$S_t = \alpha X_t + (1 - \alpha) S_{t-1} \quad (19)$$

Non-Seasonal ETS

Non-seasonal time series includes a trend component. To estimate the trend component, the simple moving average function is used as shown in Equation 20.

$$\text{SMA} = \frac{1}{n} \sum_{i=0}^{n-1} p_{m-i} \quad (20)$$

Naive Method

This method works quite well for economic and financial time series. This approach sets each prediction to be equal to the last observed value of the same season. Equation 21 is used for the calculation.

$$\hat{Y}_{t+h/t} = Y_t \quad (21)$$

Naive forecast is the model that calculates in the simplest way using the actual demand for the past period as the expected demand for the future period with an assumption of the past will repeat.

Average Method

Estimates of all the future values are equal to the average of the historical data. This approach can be used with any kind of data where historical data is available. Formula 22 shows this simple approach.

$$\hat{Y}_{t+h/t} = (y_1 + \dots + y_t) / T \quad (22)$$

To make a forecast using averaging, this formula (22) simply takes the average of selected periods of the past data by summing each period and dividing the result by the number of periods. Therefore, forecast of all future values ($\hat{Y}_{t+h/t}$) is equal to mean of historical data. It is worth pointing out that this technique is very effective and useful for short term forecasts.

Drift Method: This method is the variation of the Naive method, but it provides an increasing or decreasing over time which is called drift for the historical data. Equation 23 shows how the method works.

$$\hat{Y}_{t+h/t} = y_t + h \left(\frac{y_t - y_1}{T-1} \right) \quad (23)$$

Forecasts are equal to last value plus average change which is an equivalent expression to the extrapolation of a line drawn between the first and last observations.

IV. EXPERIMENTAL RESULTS

Datasets were obtained from Walmart Recruiting Store Sales Competition page of Kaggle website [19]. There are three different files called features.csv, stores.csv and train.csv. In features.csv, there are several features such as temperature, fuel price, and unemployment. In stores.csv, there are features such as store id, store type, and store size. Finally, train.csv has some historical data and real sales amount. The first step is to create a new experiment. Then, the features.csv, stores.csv and train.csv documents are joined to create a single dataset. The next step is to perform feature selection and feature extraction.

Feature.csv file includes store, date, temperature, fuel_price, Markdown1, Markdown2, Markdown3, Markdown4, Markdown5, CPI, Unemployment, IsHoliday features. Store.csv includes the store, type, size features, and train.csv includes the store, department, date, weekly_sales, IsHoliday.

These three files were imported into MS SQL Server Management Studio and were combined with an inner join operation. The features used in the system are the store, department, date, isHoliday, size, fuel_price, isHoliday. The class label is weekly_sales. As a result of the feature selection, the following features were identified: week of the month, monthOfYear, prevWeekSales, season, store size,

tempCategory. Also, economical features were investigated and used for the prediction such as PPI, TreInf, TNF. These economical features were added to the previous feature set. Therefore, the following features were created in the dataset: Store, department, week of the month, MonthOfYear, PrevWeekSales, isHoliday, period, season, size, PPI, TNF, TreInf, fuel_price.

PPI is the acronym of Producer Price Index and sets the inflation rate incurred by the purchase of goods and services. TreInf indicates the amount of income per household. TNF is the acronym of Total Nonfarm Payrolls which is a monthly report showing how many employers provide employment in private or government sectors for the previous month, or how much reduction occurs in the employment in the United States. Season and period were extracted from the date feature. 1 is used to represent the autumn season, 2 for the winter season, 3 for the spring season and 4 for the summer season.

The Period feature is selected based on the company's yearly plan. Most of the companies start their yearly plan in January. Therefore, 1 represents the first three months, 2 is used for the next 3 months. The week of the month and the month of the year were also extracted from the date feature. The store size feature was extracted from the size feature and the temperature category was extracted from temperature. 1 represents below 15 C°, 2 is between 15 C° - 25 C° and 3 means over 25 C°. Also, previous week sales feature was extracted from the weekly_sales feature by using previous week sales. IsHoliday includes the following days: Super Bowl, Christmas, Labor Day, Thanksgiving. WeekNo feature was extracted from the date feature. Dates on the training dataset were numbered to weeks. Temperature, Markdown1, Markdown2, Markdown3, Markdown4, Markdown5, CPI, Unemployment, store size, TempCategory were observed to have negative effects on the models. Therefore, these features were removed. The data which is up to 2012-01-01 were used for training and the remaining four months were used for testing.

During the experiments, parameters of the regression algorithms were optimized by using "tune hyper model parameters" feature of Azure ML Studio platform. After the best model was selected, "Set Up Web Service" button with "Predictive Web Service" option was used to produce the web service. Web service was deployed by selecting "Deploy Web Service" button. Web Services are displayed on the "Web Services" tab on the Azure ML Studio. Web service C# code for a client application is displayed on the REQUEST/RESPONSE page and its API key is displayed on the dashboard screen of the selected web service. The API key is needed to authenticate the user. In this study, MAE and RMSE error parameters were used.

Root Mean Squared Error: It is the square root of the average error. The mean squared error is shown based on Equation 24.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{obs,i} - X_{model,i})^2}{n}} \quad (24)$$

X_{obs} : observed value

X_{model} : predicted values at time/place i .

Mean Absolute Error (MAE): It takes the average of absolute errors. Equation 25 is used to calculate this parameter. MAE calculates the average absolute difference between y_i and x_i which are the coordinates of point i .

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_t| \tag{25}$$

Initial experiments were performed for only one store and one department of the Walmart Company because of the time series analysis techniques. Test results were evaluated by several evaluation parameters such as Mean Absolute Error and Root Mean Squared Error. Decision Forest Regression was the best approach based on RMSE and MAE value as shown in Figure 1.



Fig. 1. Evaluation of experiments

Table 1 shows that four methods provide better performance. The difference between Decision Forest Regression, Non-Seasonal ARIMA, Boosted Decision Tree Regression, and Seasonal ARIMA's accuracy is not too much. Decision Forest Regression provides the highest performance.

TABLE 1. RMSE AND MAE RESULTS

Method	RMSE	MAE
Seasonal ARIMA	3650.08	2923.35
Non-Seasonal ARIMA	3635.49	2916.03
Seasonal ETS	8274.58	7582.88
Bayesian Linear Regression	4771.36	3980.12
Linear Regression	4613.01	3666.66
Decision Forest Regression	3439.48	2750.21
Boosted Decision Tree Regression	3637.61	2907.69
Neural Network Regression	9622.62	7745.15

To extend the experiment with the entire dataset, all the departments were tested with these regression methods because the use of time series analysis techniques is not appropriate for the entire dataset. Parameters were optimized with Tune Model Hyper Parameters module in Azure ML

Studio. This time, the best approach is Boosted Decision Tree Regression. In Table 2, results are shown.

TABLE 2. ALL THE DEPARTMENTAL RESULTS

Method	MAE	RMSE	Coefficient of Determination
Bayesian Linear Regression	2469.54	4361.40	0.96
Linear Regression	2480.12	4365.04	0.96
Neural Network Regression	14951.09	22499.33	0.00
Boosted Decision Tree Regression	1669.10	3696.59	0.97

The coefficient of determination, which is also known as R^2 (equation 26) represents the predictive power of the model as a value between 0 and 1.

$$R^2 = \frac{\sum (y - \bar{y})^2 - \sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \tag{26}$$

The coefficient of determination, R^2 , signifies the proportion of the total sample variation in y which is measured by the sum of squares of deviations of the sample y values about their mean \bar{y} , is attributed to the linear relationship between x and y . It is a standard way of evaluating how well the model fits the data. It can be interpreted as the proportion of variation explained by the model. A higher proportion is better, where 1 indicates a perfect fit. We used the coefficient of determination factor to express our accuracy. Moreover, the highest prediction success in estimates where all stores are used is 0.97 with Boosted Decision Tree Regression method as shown in Table 2.

As explained in the Related Work section, several studies applied the times series analysis approaches [4, 6] and artificial neural network algorithms [14, 16, 20] for predicting the sales amount. In this study, we demonstrated that regression algorithms (i.e., Boosted Decision Tree Regression) work much better than these algorithms used in the literature and it does not require much optimization to provide high performance. We developed our NN model using a standard topology, and therefore this might affect the performance of the model. This can be considered as one of the threats to the validity of this study. Also, our experiments were performed on a specific dataset and hence, the performance might be different on other kinds of datasets. We suggest practitioners apply the Boosted Decision Tree Regression algorithm for sales forecasting.

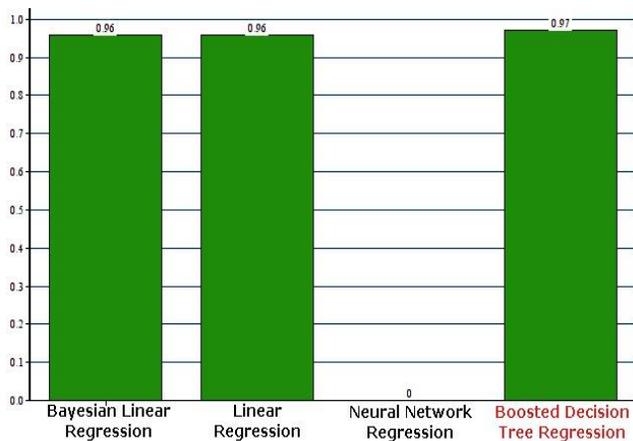


Fig. 2. The coefficient of determination results

V. CONCLUSION

It is crucial to predicting the sales amounts as close as to the actual sales amounts for enterprises to increase their profits [20]. Unless an accurate forecasting model is built, cash flow problems are inevitable. Therefore, building this kind of prediction models for sales forecasting has a high priority for the organizations. In this study, we investigated the effect of Regression and Time Series Analysis methods on the sales forecasting problem. Our experiments show that the regression techniques provide higher performance and accuracy compared to the time series analysis techniques. Boosted Decision Tree Regression algorithm was the best predictor for sales forecasting with the 0.97 coefficient of determination. Prediction results were obtained for weekly sales quantities. In the future, a hybrid model using ARIMA and Boosted Decision Tree Regression techniques will be investigated to solve this problem. In addition, the implementation of another hybrid model with an SVM model is planned. New experiments will be performed if the new public data is obtained. Also, deep learning algorithms [22, 23] can be investigated for sales prediction problem.

ACKNOWLEDGMENT

The authors thank Wageningen University and Istanbul Kültür University for the infrastructure support to complete this scientific research.

REFERENCES

- [1] Kuo, R. J. (2001). A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm. *European Journal of Operational Research*, 129(3), 496-517.
- [2] Chen, F. L., & Ou, T. Y. (2011). Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry. *Expert Systems with Applications*, 38(3), 1336-1345
- [3] Zhao, J., Tang, W., Fang, X., Wang, J., Liu, J., Ouyang, H., ... & Qiang, J. (2015, September). A Novel Electricity Sales Forecasting Method Based on Clustering, Regression and Time Series Analysis. In *Proceedings of the 2015 International Conference on Artificial Intelligence and Software Engineering*.
- [4] Tian, Y., Liu, Y., Xu, D., Yao, T., Zhang, M., & Ma, S. (2012, April). Incorporating Seasonal Time Series Analysis with Search Behavior Information in Sales Forecasting. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 615-616).

- [5] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.
- [6] Pandey, A., & Somani, R. K. (2013). A Cloud Computing Based Sales Forecasting System for Small and Medium Scale Textile Industries. *computing*, 3(4).
- [7] Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3), 324-342.
- [8] Vijayalakshmi, M., Menezes, B., Menon, R., Divecha, A., Ravindran, R., & Mehta, K. (2010, October). Intelligent sales forecasting engine using genetic algorithms. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 1669-1672). ACM.
- [9] Yeo, J., Kim, S., Koh, E., Hwang, S. W., & Lipka, N. (2016, April). Browsing2purchase: Online Customer Model for Sales Forecasting in an E-Commerce Site. In *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 133-134). International World Wide Web Conferences Steering Committee.
- [10] Choi, T. M., Yu, Y., & Au, K. F. (2011). A hybrid SARIMA wavelet transform method for sales forecasting. *Decision Support Systems*, 51(1), 130-140.
- [11] Chang, P. C., Liu, C. H., & Fan, C. Y. (2009). Data clustering and fuzzy neural network for sales forecasting: A case study in printed circuit board industry. *Knowledge-Based Systems*, 22(5), 344-355.
- [12] Wong, W. K., & Guo, Z. X. (2010). A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. *International Journal of Production Economics*, 128(2), 614-624.
- [13] Katkar, V., Gangopadhyay, S. P., Rathod, S., & Shetty, A. (2015, January). Sales forecasting using data warehouse and Naïve Bayesian classifier. In *Pervasive Computing (ICPC), 2015 International Conference on* (pp. 1-6). IEEE
- [14] Müller-Navarra, M., Lessmann, S., & Voß, S. (2015, January). Sales Forecasting with Partial Recurrent Neural Networks: Empirical Insights and Benchmarking Results. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on* (pp. 1108-1116). IEEE.
- [15] Gao, M., Xu, W., Fu, H., Wang, M., & Liang, S. (2014, July). A novel forecasting method for large-scale sales prediction using extreme learning machine. In *Computational Sciences and Optimization (CSO), 2014 Seventh International Joint Conference on* (pp. 602-606). IEEE.
- [16] Omar, H. A., & Liu, D. R. (2012, January). Enhancing sales forecasting by using neuro networks and the popularity of magazine article titles. In *2012 Sixth International Conference on Genetic and Evolutionary Computing (ICGEC)* (pp. 577-580).
- [17] Lu, C. J., Lee, T. S., & Lian, C. M. (2010, December). Sales forecasting of IT products using a hybrid MARS and SVR model. In *2010 IEEE International Conference on Data Mining Workshops* (pp. 593-599).
- [18] Stojanović, N., Soldatović, M., & Miličević, M. (2014, June). Walmart Recruiting-Store Sales Forecasting. In *Proceedings of the XIV International Symposium Symorg 2014: New Business Models and Sustainable Competitiveness* (p. 135). Fon.
- [19] Kaggle.com | Kaggle Datasets | Open Datasets for Any Project, "https://www.kaggle.com/" Access date: 25.02.2019.
- [20] Bohanec, M., Borštnar, M. K., & Robnik-Šikonja, M. (2017). Explaining machine learning models in sales predictions. *Expert Systems with Applications*, 71, 416-428.
- [21] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- [22] Tsoumakas, G. (2018). A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review*, 1-7.
- [23] Luce, L. (2019). *Deep Learning and Demand Forecasting*. In *Artificial Intelligence for Fashion* (pp. 155-166). Apress, Berkeley, CA.

BIOGRAPHIES



CAGATAY CATAL received the B.S. and M.S. degrees in computer engineering from Istanbul Technical University, Istanbul, and the Ph.D. degree in computer engineering from Yildiz Technical University, Istanbul, in 2008. He worked 8 years at the Scientific and Technological Research Council of Turkey as Senior Researcher.

Later, he worked 6 years in Istanbul Kültür University, Department of Computer Engineering as Associate Professor and was the Head of the Department for the last 3 years. On January 2018, he joined the Information Technology Group of Wageningen University in the Netherlands. His research interests are data science, machine learning, software engineering, and experimental software engineering.



KAAAN ECE received the B.S. degree in computer engineering from Istanbul Kültür University and has been working in Eteration software company for 2 years. His research interests are machine learning and software engineering.



BEGÜM ARSLAN received the B.S. degree in computer engineering from Istanbul Kültür University. Her research interests are machine learning, data science, and software engineering.



AKHAN AKBULUT received the B.S. and M.S. degrees in computer engineering from Istanbul Kültür University and the Ph.D. degree in computer engineering from Istanbul University, Istanbul. Later, he worked 2 years in the software industry. In 2004, he joined the department of computer engineering in Istanbul Kültür

University. He has been working at the department of computer science in North Carolina State University since 2017 as Postdoctoral researcher. His research interests are machine learning, cloud computing, software engineering, and internet architectures.