Int. J. Adv. Eng. Pure Sci. 2022, 34(3): 415-425
DOI: 10.7240/jeps.1127844

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

Demand Forecasting with Integration of Time Series and Regression Models in Pharmaceutical Industry

İlaç Sektöründe Zaman Serisi ve Regresyon Birleşik Modeller ile Talep Tahmini Uygulaması

Salih İMECE[1], **Ömer Faruk BEYCA**[1*]

[1]*İstanbul Teknik Üniversitesi, Endüstri Mühendisliği Bölümü, 34367, İstanbul, Türkiye*

Abstract
Accurate demand forecasting is crucially important to reduce inventory and backlogging cost. In this study, we analyze how promos, holiday statements, price changes, stock availability and date-time features (weekdays, months etc.) affect the demand by using several forecasting methods. Data sets were collected for the products of the global pharmaceutical company providing services in Turkey. Actual daily sales data for 2016, 2017 and 2018 were used in the construction of this data set. In order to predict the next periods demand, we used four different models which are Holt Winters, Ridge Regression, Random Forest and Xgboost. We also ensemble those models to improve forecasting accuracy. Next, by weighting inversely proportional to the error rates of the models, binary, triple and quadruple combinations of the single models were compared with themselves and the single models. Our numerical results show that the lowest forecasting error rate was obtained in ensemble models. Particularly, the lowest error rate in individual models was obtained in Random Forest with 15.7% RMSPE (Root Mean Square Percentage Error) value, and the lowest error rate was obtained with 10.7% RMSPE value in Holt Winters & Xgboost models combination. Results show that ensemble of several models can increase the forecasting accuracy.
Keywords: demand forecasting, production planning, ensemble models, inventory

Öz
Doğru talep tahmini, karşılanmayan talep ve stok miktarını azaltmak için büyük önem taşımaktadır. Bu çalışma, yapılan promosyonların, yıl içi tatil günlerinin, ürünün fiyatında olan değişikliklerin, ürünün stokta bulunup bulunmamasının ve bazı tarih özelliklerinin (haftanın günleri, aylar, yıllar vb.) birden çok tahmin modelinde kullanılarak talebi nasıl etkilediğinin analiz edilmesini amaçlamaktadır. Çalışma için, Türkiye'de uzun yıllar hizmet veren bir global ilaç şirketine ait bir ürün incelenmiştir. Veri seti için 2016, 2017 ve 2018 yıllarına ait günlük satış verileri kullanılmıştır. Gelecek dönemlerin talebini tahmin etmek için; Holt Winters, Ridge Regression, Rastgele Orman ve Xgboost olmak üzere dört ayrı model kullanılmıştır. Ayrıca tahmin doğruluğunu arttırmak için dört modelin birbiriyle olan kombinasyonlarından oluşan modeller de kullanılmıştır. Sonrasında, modellerin hata oranları ile ters orantılı şekilde ağırlıklandırma yapılarak, tekli modellerin ikili, üçlü ve dörtlü kombinasyonları elde edilmiş ve hata oranları hem kendi aralarında hem de tekli modellerle kıyaslanmıştır. Sonuçlar, en düşük tahminleme hatalarının birleştirilmiş modellerden elde edildiğini göstermiştir.Oluşturulan tüm modeller hata oranı bakımından kıyaslandığında, hata oranı en düşük modelimiz %10,7 RMSPE (Kök ortalama Kare Yüzde Hata) değeri ile Holt Winters ve Xgboost modellerinin kombinasyonlarından oluşan kombinasyon olmuştur. Sonuçlar, birden çok modelin birlikte kullanılarak talep tahmininin doğruluk oranının artırılabileceğini göstermiştir.
Anahtar kelimeler: talep tahmini, üretim planlama, birleştirilmiş modeller, stok

I. INTRODUCTION

Production planning is the one of the critical process of pharmaceutical, which has a vital impact in competitive environment. Forecasting methods are widely used for decision making when uncertainty is present. Forecasting methods are utilized in production planning to support decision makers (Goodwin, 2005; Moskowitz, 1972). Moreover, demand forecasting also affects finance, controlling sales, price policy etc. (Cook, 2016).

Although pharmaceutical companies have fairly data, they do not effectively utilize these data to assist decision makers. For example, data mining can be used to analyze datasets in order to discover hidden relations, and find models and trends (Hand David & Heikki, 2001). Another important utilization of recorded data is demand forecasting. Demand forecasting in the pharmaceutical industry is a complex task in terms of different effective elements (Smits & Boon, 2008). There are many elements that might affect consumption quantity of an exclusive kind of drugs such as pharmacologic class and practices of different kinds of medicine, seasonal effects, geographic variety of the patients, patients' characteristics regarding on the cultural background, new launch, and drug prices.

Corresponding Author: Ömer Faruk BEYCA, Tel: +90 551 109 4103 e-posta: beyca@itu.edu.tr

Submitted: 09.06.2022, **Revised:** 22.09.2022, **Accepted:** 22.09.2022

The predictive modelling provides an estimation about unknown data by using historical data. Some of such models are time series analysis, regression, gradient boosting algorithms, neural networks etc. Forecasting is the steps of investigating the relationship of current and past observations and prediction of future values. The regression is the learning function from data to predict future values. In time series analysis, time-dependent changing of data is examined. In time series analysis, many statistical methods are used to analyze the data like Holt Winters, ARIMA etc. (Padhy et al., 2012).

Demand forecasting is a substantial and complex sub-process of supply chain management. Demand forecasting can be defined as the process of developing models and calculating the related quantities in a timely manner to help accurate planning of all processes in the supply chain. For example, by precise forecasting of demand, unnecessary logistics costs, storage costs and supply of unnecessary raw materials can be reduced.

The objective of this research is to introduce new variables and develop models to predict the demand amounts of the products from the wholesalers for a pharmaceutical company. The actual sales of a pharmaceutical company for 4 years period was considered in order to forecast demands. A nutritional product that has dynamic demand activities due to promotions and seasonal effects is chosen. Figure 1 shows the flow for production from company to the patient.

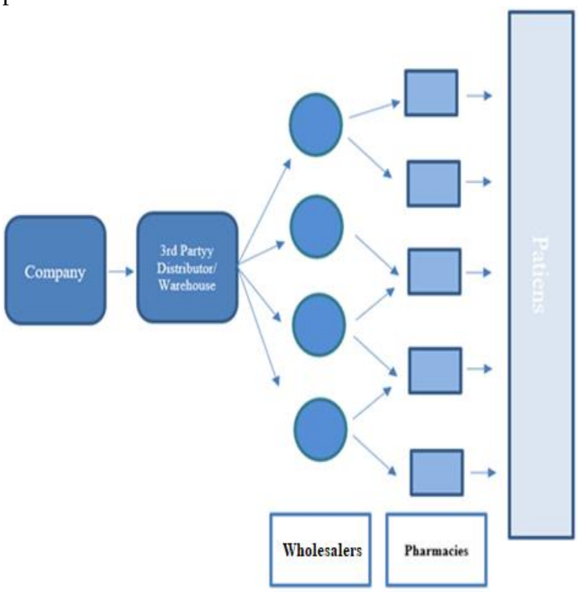

Figure 1. Supply Flow For the Company

The contributions of this study are three-fold.

- New variables are introduced that has not been used in the literature before for pharmaceutical industry such as price change and stock availability.
- Various machine-learning models are employed to time series prediction.
- Ensemble methods that combine multiple models are proposed.

This paper is organized as follows. In second section, the literature review is presented. In the Section 3 and Section 4, proposed models are presented. Section 5 discuss the numerical results. Finally, Section 6 highlights the main conclusion and future work.

II. LITERATURE REVIEW

Demand forecasting for wholesalers requires the prediction of how many products any wholesaler will sell in the planning periods. In the literature, prediction methods are mainly in two groups (single-model and hybrid models). Methods consisting of a single model can be divided into statistical models and machine learning models.

Merkuryeva et al. (2019) suggested different demand forecasting cases using simple moving average, multiple linear regression, and symbolic regression for a pharmaceutical company. They used some features “distributor price-list; the discounted selling price of the product” for regression models. Micajkova et al. (2018) studied about annual sales of a pharmaceutical company from 2001 to 2015 in Macedonia. They proposed models of moving average, exponential smoothing and regression analysis to predict sales of 2016. Anusha et al. (2014) introduced six month moving average, simple exponential smoothing and Winters’ exponential smoothing models to predict demand of two product family of a pharmaceutical company. Box-Jenkins methodology is another commonly used statistical method. For example, time series analysis and prediction were conducted using Box-Jenkins seasonal ARIMA (Autoregressive Integrated Moving Average) models (Liu et al., 2001).

Artificial intelligence algorithms gain popularity in pharmaceutical demand forecasting due to the complexity of supply chain problems. Ali et al. (2009) stated that statistical methods are successful for non-promotion periods but insufficient in case of promotion. In the study, they showed that regression trees are more successful than traditional statistical methods in the case of promotion. Hasin et al. (2011) tested the performance of traditional Holt-Winters model and artificial neural networks on the dataset of a supermarket in Bangladesh. They showed that, based on MAPE (Mean Absolute Percentage Error) metric, artificial neural networks are superior to traditional statistical models. Ghousi et al. (2012) employed Neural Networks and Decision Tree algorithms to forecast the quantity of various drugs monthly consumptions. They stated that both Decision Tree and Neural Networks are stable models in this type of cases with several important factors utilized, while

Decision Tree would perform a little better than Neural Networks in most of these cases.

Some recent studies show that, different models are used together to increase model's estimation performance. Pavlyshenko (2019) indicated that sales estimation is a regression problem rather than a time series problem. They showed that regression models can usually provide better outcomes compared to time series models. They ensembled the individual models for better accuracy. Tugay and Ögüdücü (2020) studied a sales forecasting model on an e-commerce web site in Turkey. They proposed a stacked generalization method with sub-level regressors. They also tested performance of individual methods separately, and then compared with the general method. Scenarios showed that stacking method estimates sales at least as good as individual methods with less training data (only %20 of the dataset). Moreover, stacked generalization was more successful when more data was used. Ganesh et al. (2019) studied a building forecasting model of the Air Quality Index. The accuracy of estimation has been increased by the stacked ensemble of individual predictors. Their results implied that the ensembled methods provide promising results for forecasting the Air Quality Index compared to individual neural network predictors and regression models. Table 1 summarizes the related literature review.

III. EXPERIMENT DESIGN AND RESULTS

In this section, we first describe the data sets attributes. Then, we present proposed models. Finally, we use the described data sets to asses performance of the proposed models and the results of the proposed models are discussed.

3.1. Dataset

In this study, the data set is obtained from actual sales of a global pharmaceutical company in Turkey. The data set consists of actual daily sales for 2016, 2017 and 2018. Selected features for sales forecasting are promotions, holiday statement, price changes, stock availability, days of the week and month, quarter, year are obtained from the datetime features.

The monthly sales distribution of the studied product is shown in the Figure 2. Monthly sales numbers were not given in the Figure 2 due to data privacy.

Figure 2. Monthly sales of the product

The features used between 2016 and 2018 is listed in the table below.

Table 2. Features used in the model

No	Variables	Measurement Scale	Possible Values
1	Promo	Nominal	0 (No Promotion), 1 (Promo 1), 2 (Promo 2), 3 (Promo 3)
2	Holiday Statement	Nominal	0(No), 1(Yes)
3	Price Changes	Nominal	1(Before 1 Month), 2(After 1 Month), 0 (No Changing)
4	Stock Availability	Nominal	0(Yes), 1(No)
5	Day of Week	Nominal	1, 2, 3, 4, 5, 6, 7
6	Date	Interval	1/1/2016 to 12/31/2018
7	Sales	Ratio	0 to 23.235

3.1.1. Promo

The company conducts some activities to increase sales. The correlation values of different promotion types with sales is showed in Table 3. Since the activities are related to the daily sales of the products, it is considered appropriate to add them to the model as a feature.

Table 3. Correlation matrix of promotion types

Promotion Types	Values
Promo 1	0.53
Promo 2	0.19
Promo 3	0.06
No Promo	-0.23

3.1.2. Holiday Statement

Sales to the pharmacies from the wholesalers on official holidays in Turkey are seriously affected. The correlation matrix of these days with sales is showed in Table 4.

Table 4. Correlation matrix according to the holiday statement

Holiday Statement	Values
Yes	-0.33
No	0.24

Table 1. Studies in the literature related to Demand Forecasting.

Author(s)	**Feature(s)**	**Methodological Approaches**	**Industries**	**Data Type & Size**
(Merkuryeva et al., 2019)	● Distributor price-list; The discounted selling price of the product; ● A week number of sales in a month; ● Weekly average currency rate.	● Multiple Linear ● Regression, ● Symbolic Regression, ● Simple Moving Average	Pharmaceutical	Historical Weekly Sales Data (41 Data points)
(Micajkova et al., 2018)	-	● Moving Average, ● Exponential Smoothing, ● Simple Regression	Pharmaceutical	Annual sales of the company from 2001 to 2015.
(Lakshmi Anusha et al., 2014)	● Number of items remaining in inventory	● Moving Average, ● Exponential Smoothing, ● Winter’s Exp. Smoothing	Pharmaceutical	Daily Sales Data (From December 2010 to November 2013)
(Liu et al., 2001)	● Identifying special events includes that holidays or festivals, sports activities, other scheduled local events, as different outlier groups	● Seasonal ARIMA	Fast Food Restaurants	Daily sales data from April 7 1997 to May 18 1998 (407 observations)
(Ali et al., 2009)	● Promotions (TV, Radio), ● Price and Discount for the SKU-store combination, ● SKU and subcategory dummies, ● The week-number and the actual unit sales for the current and last four weeks.	● Stepwise Linear ● Regression, ● Support Vector Regression, ● Regression Tree	Grocery Retailing	76 weeks period (first 51 weeks for training data, after 25 weeks for testing data)
(Ghousi et al., 2012)	● Year, Season, Month, ● Disease, ● Price, ● Population, ● Educated Women/Men, ● Marriage/Divorce ratio	● Regression, ● Neural Networks, ● Decision Trees	Pharmaceutical	3 year dataset of a drug distribution center (407.000 records)

(Ferreira et al., 2016)	• Price, • Events start date/length, • Initial inventories, • Product's brand, size, color, hierarchy	• Regression Trees, • Principal Component • Regression, • Least Squares Regression	Online Retailer	Sales transaction data from the beginning of 2011 through mid-2013
(Hasin et al., 2011)	• Promotions, • Holiday statement, • Festival Periods, • Availability of items on the shelf, • Consumption rate, • Price range of the item.	• Holt Winters, • Neural Networks	Retailer	Including five years daily sales data
(Martin & others, 2019)	• Household Real Disposable Income (*HRDI*), • Total Headline Consumer Price Index (*CPI*), • Gold price in US Dollars (*gold*), • Oil price in US Dollars (*oil*) etc.	• Elastic-net Regression, Random Forests, Support Vector Machines, Recurrent neural networks.	Economics	Quarterly South African GDP dating from Q2 1992 to Q4 2016.
(Pavlyshenko, 2019)	• Promotions, • Holiday statement, • Store types, • Distance from store to competitor's store	• ARIMA, • Lasso, • Random Forest, • Neural Networks, • Stacking (Ensemble)	Retailer (Supermarket)	Daily sales data each store & product (2 years)
(Tugay & Oguducu, 2020)	• Year, Month, Week, Day and Weekday information. • Popularity of the product/good.	• Linear Regression, • Decision Tree, • Random Forest, • Gradient Boosting, • Stacking (Ensemble)	E-commerce Web Site	Weekly sales data (Containing 1925 instances)
(Ganesh et al., 2019)	• NO_2, CO, O_3, $PM_{2.5}$, SO_2 and PM_{10}	• Conjugate Gradient descent, • Neural Networks, • Support Vector Regression, • Stacking (Ensemble)	Air Quality	Data containing 2000 samples (1500 /500 train & test)

3.1.3. Price Changes

We observed that sales increased in the month prior to price increase for a product due to tendency of buying more products at a low price. Consequently, the month after the price increase, sales are low due to the purchase of large quantities of products in previous month. This feature is also added to the dataset because it affects sales.

3.1.4. Stock Availability

Even if there is a sales demand, the demand is not met due to lack of inventory. Although this is rarely the case in this product, it is included in the data set as a feature which is a binary variable.

3.1.5. Day of Week

Depending on the day of the week, sales from wholesalers to pharmacies are varying. Sales increase at the beginning of the week, and significantly decrease at the end of the week. Details are provided in the correlation matrix in Table 5.

Table 5. Correlation matrix of weekdays

Weekdays	Values
Weekday 1	0.42
Weekday 2	0.02
Weekday 3	0.10
Weekday 4	-0.06
Weekday 5	-0.01
Weekday 6	-0.17
Weekday 7	-0.28

3.1.6. Date

For the tree based and regression models; month, quarter, year features were obtained from Date feature and added to dataset. The addition of these features is important in sales forecasting, as well as trying to reflect the characteristics of time series data to regression and decision tree models. The correlation table between the months of the year and sales can be seen in Table 6.

Table 6. Correlation matrix of months

Months	Values
Month 1	-0.12
Month 2	-0.18
Month 3	0.54
Month 4	0.08
Month 5	-0.06
Month 6	-0.06
Month 7	-0.33
Month 8	-0.11
Month 9	0.31
Month 10	-0.07
Month 11	-0.02
Month 12	0.02

3.1.7. Sales

It consists of daily sales from wholesalers to pharmacies. Due to returns, days of the sales that below zero were ignored. In order to reduce the effect of outlier data, the logarithmic transformation was used in the models except Holt Winters.

3.2. Models

We utilize Holt Winters, Ridge Regression, Random Forest and Xgboost methods for sales forecasting. In Holt Winters model, we only use historical sales data. In remaining three models we implement the whole data discussed in Section 4.1. We use 70% of data for training and 30% for testing. In experimental design, we first test individual models to assess their performances. The error rates of the applied methods were calculated by RMSPE.

3.3. Results

The RMSPE values of single models are given in Table 7.

Table 7. RMSPE values for each models

Models	RMSPE
Holt Winters	31.6%
Ridge Regression	25.8%
Random Forest	15.7%
Xgboost	17.0%

We observed that the lowest RMSPE values achieved by Random Forest model followed by Xgboost, Ridge Regression, and Holt Winters respectively.

We next construct ensemble models from these four algorithms. Weighted average of the model results is used for ensemble models. Weights of each model are inversely proportional to the error rates of the models and calculated as follows:

$$w_i = \frac{\frac{1}{e_i}}{\sum_i \frac{1}{e_i}}$$

where w_i and e_i is the weight and error rate of the i^{th} model respectively. Final prediction value is calculated as by taking weighted average of individual model predictions as follows:

$$\hat{y}_{final} = \sum_i w_i \, \hat{y}_i$$

where $\hat{y}_{final}$ is the ensembled results and $\hat{y}_i$ is the individual result of the model i.

In particular, dual combinations of singular models are formed and examined. The results for the dual combinations are illustrated in Table 8.

Table 8. RMSPE values for dual combinations of models

Models	RMSPE
Holt Winters & Ridge Regression	24.0%
Holt Winters & Random Forest	19.8%
Holt Winters & Xgboost	10.7%
Ridge & Random Forest	20.5%
Ridge Regression & Xgboost	15.2%
Random Forest & Xgboost	12.3%

RMSPE values given in Table 8 shows that the combinations Holt Winters & Xgboost algorithm achieves the best RMSPE value. Holt Winters & Xgboost algorithm is followed by Random Forest & Xgboost, Ridge Regression & Xgboost, Holt Winters & Random Forest, Ridge & Random Forest and Holt Winters & Ridge Regression respectively. We note that, single performances of the algorithms are not correlated with combined models' performances. Particularly, Random Forest and Xgboost are the best performing algorithms individually, but their combination is outperformed by Holt Winters & Xgboost. This result may arise since combined algorithms exploit different advantages of each algorithm.

We also inspect triple combinations of four selected algorithms. The error rates obtained from the models created by these combinations are shown in Table 9. We observe that the combination of Holt Winters, Random Forest and Xgboost has the lowest RMSPE value.

Table 9. RMSPE values for triple combinations of models

Models	RMSPE
Holt Winters & Ridge Regression & Random Forest	20.1%
Holt Winters & Ridge Regression & Xgboost	13.7%
Holt Winters & Random Forest & Xgboost	11.2%
Ridge Regression & Random Forest & Xgboost	13.8%

Finally, we create a model by using all four algorithms used in this study. RMSPE value of this model is given as seen in Table 10.

Table 10. RMSPE values for the combination that include all individual models

Models	RMSPE
Holt Winters & Ridge Regression & Random Forest & Xgboost	13.1%

We also present the best RMSPE values for each of four cases in Table 11. Holt Winters & Xgboost achieves the lowest RMSPE value among all algorithms.

Table 11. The best RMSPE values for each of four cases

Models	RMSPE
Random Forest	15.7%
Holt Winters & Xgboost	10.7%
Holt Winters & Random Forest & Xgboost	11.2%
Holt Winters & Ridge Regression & Random Forest & Xgboost	13.1%

IV. CONCLUSIONS AND FUTURE DIRECTIONS

We develop models demand prediction for a pharmaceutical company. We introduced new variables to be considered for pharmaceutical industry such as price change and stock availability. Four different models are considered for demand forecasting. We also construct ensemble models of the studied models by weighting the models error rates. Our results show that ensemble models' performance is superior to single models. We note that, individual performance of the algorithms is not corelated with combined models' performance. This result may arise because combined algorithms exploit different advantages of each algorithm. While regression algorithms can learn the relationship of the exogenous inputs to output Holt Winters method may capture the temporal relationship of the time series better than regression models. This study can guide both the academicians and practitioners working in the pharmaceutical industry for demand forecasting.

This study has a few limitations that may provide an opportunity for future research. First, in this study we only use the data pharmaceutical industry in Turkey. Demand of pharmaceutical products may vary for different countries. Second, the study is conducted for a single product. An improvement can be made by including other products. Future studies may include grouping (sales volume, target audience, sales behavior, etc.) by using various data mining methods and including these groups in the forecasting model. In addition, various data mining studies can be carried out by considering the market shares and profitability in the region in the classification of products. As a result of these studies, different forecasting models can be included in the model communities and tests can be performed on how the performance will change.

REFERENCES

[1] Al-Hafid, M. S., & Hussein Al-maamary, G. (2012). Short term electrical load forecasting using holt-winters method. *Al-Rafidain Engineering Journal (AREJ)*, *20*(6), 15–22.

[2] Al-Hassan, Y. M. M., & Al-Kassab, M. M. (2000). *A comparison between ridge and principal components regression methods using simulation technique*. Al Al-Bayt University.

[3] Ali, Ö. G., Sayin, S., Van Woensel, T., & Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*, *36*(10), 12340–12348.

[4] Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*, *13*(1), 1063–1095.

[5] Boulesteix, A.-L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(6), 493–507.

[6] Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

[7] Chen, C., Liaw, A., Breiman, L., & others. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley*, *110*(1–12), 24.

[8] Chen, T. (2014). Introduction to boosted trees. *University of Washington Computer Science*, *22*(115), 14–40.

[9] Cook, A. G. (2016). *Forecasting for the pharmaceutical industry: models for new product and in-market forecasting and how to use them*. Gower.

[10] Ferreira, K. J., Lee, B. H. A., & Simchi-Levi, D. (2016). Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing \& Service Operations Management*, *18*(1), 69–88.

[11] Ganesh, S. S., Arulmozhivarman, P., & Tatavarti, R. (2019). Forecasting air quality index using an ensemble of artificial neural networks and regression models. *Journal of Intelligent Systems*, *28*(5), 893–903.

[12] Ghousi, R., Mehrani, S., Momeni, M., & Anjomshoaa, S. (2012). Application of data mining techniques in drug consumption forecasting to help pharmaceutical industry production planning. *Proceedings of the 2012 International Conference on Industrial Engineering and Operations Management*, 3–6.

[13] Goodwin, P. (2005). Providing support for decisions based on time series information under conditions of asymmetric loss. *European Journal of Operational Research*, *163*(2), 388–402. https://doi.org/https://doi.org/10.1016/j.ejor.2003.10.039

[14] Hand David, J., & Heikki, M. (2001). *Principles of data mining (adaptive computation and machine learning)*. Bradford Books.

[15] Hasin, M. A. A., Ghosh, S., & Shareef, M. A. (2011). An ANN approach to demand forecasting in retail trade in Bangladesh. *International Journal of Trade, Economics and Finance*, *2*(2), 154.

[16] Karadavut, U., Aşır, G., Tozluca, A., Kınacı, İ., Aksoyak, Ş., Palta, Ç. and Pekgör, A. (2005). Nohut Cicer arietinum L. Bitkisinde Verime Etki Eden Bazı Karakterlerin Alternatif Regresyon Yöntemleriyle Karşılaştırılması. *Journal of Agricultural Sciences*, *11*(03), 328–333.

[17] Lakshmi Anusha, S., Alok, S., & Shaik, A. (2014). Demand Forecasting for the Indian Pharmaceutical Retail: A Case Study. *Journal of Supply Chain Management Systems*, *3*(2).

[18] Liu, L.-M., Bhattacharyya, S., Sclove, S. L., Chen, R., & Lattyak, W. J. (2001). Data mining on time series: an illustration using fast-food restaurant franchise data. *Computational Statistics \& Data Analysis*, *37*(4), 455–476.

[19] Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). *Forecasting methods and applications*. John wiley \& sons.

[20] Martin, L.-C., & others. (2019). *Machine learning vs. traditional forecasting methods An application to South African GDP*.

[21] Merkuryeva, G., Valberga, A., & Smirnov, A. (2019). Demand forecasting in pharmaceutical supply chains: A case study. *Procedia Computer Science*, *149*, 3–10.

[22] Micajkova, V., Georgieva Svrtinov, V., Petkovski, V., & Esmerova, E. (2018). *Selecting appropriate forecast method on the basic of forecast accuracy-pharmaceutical company case study*.

[23] Moskowitz, H. (1972). The Value of Information in Aggregate Production Planning—A Behavioral Experiment. *A I I E Transactions*, *4*(4), 290–297. https://doi.org/10.1080/05695557208974865

[24] Padhy, N., Mishra, D., Panigrahi, R., & others. (2012). The survey of data mining applications and feature scope. *ArXiv Preprint ArXiv:1211.5723*.

[25] Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting. *Data*, *4*(1), 15.

[26] Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*, 93–104.

[27] Smits, R. E. H. M., & Boon, W. P. C. (2008). The role of users in innovation in the pharmaceutical industry. *Drug Discovery Today*, *13*(7–8), 353–359.

[28] Tosun, T. (2006). *Veri madenciliği teknikleriyle kredi kartlarında müşteri kaybetme analizi* Fen Bilimleri Enstitüsü.

[29] Tugay, R., & Oguducu, S. G. (2020). *Demand Prediction Using Machine Learning Methods and Stacked Generalization*. arXiv. https://doi.org/10.48550/ARXIV.2009.09756

[30] Yıldırım, N. (2010). En küçük kareler, Ridge regresyon ve Robust regresyon yöntemlerinde analiz sonuçlarına aykırı değerlerin etkilerinin belirlenmesi.. *Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tez, Adana*.

APPENDIX

Holt Winters Method

Holt winters method can directly analyze the seasonal effect in the data. It is one of the exponential correction methods. The formulation of the method consists of three separate equations; trend, level and seasonality. The seasonal equation must be either additive or multiplicative. In the planned analysis, it is recommended to use the Multiplicative-Seasonal method if the seasonal effect shows a steady increase or decrease with the trend, and additive seasonal method if the seasonality effect shows an increase or decrease with an irregular tendency (Al-Hafid & Hussein Al-maamary, 2012).

Multiplicative Seasonal Holt Winters Method

The equations of the multiplicative-seasonal Holt Winters method are as follows:

Level: $L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1})$ (1)

Trend: $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$ (2)

Seasonality: $S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-s}$ (3)

Prediction: $F_{t+m} = (L_t + b_t m)S_{t-s+m}$ (4)

Where α, β, γ are order parameters, t is time period, y_t is observation at time t, s is seasonal length, L_t is level component at time t, b_t is trend component at time t, S_t is seasonal component at time t and F_{t+m} is forecast value in m period.

Optimum α, β, γ parameters are found as 0.32, 0.11, 0.62 respectively.

Additive-Seasonal Holt Winters Method

The seasonal component of the Holt Winters method rarely can be additive. The equations for the additive-seasonal Holt Winters method are as follows:

Level: $L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1})$ (5)

Trend: $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$ (6)

Seasonality: $S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-s}$ (7)

Prediction: $F_{t+m} = (L_t + b_t m) + S_{t-s+m}$ (8)

The main difference between the additive-seasonal Holt Winters technique and the multiplicative-seasonal Holt Winters technique is the calculation of seasonal indices. In particular, in the additive-seasonal Holt Winters technique seasonal indices are calculated as addition and subtraction while in the multiplicative-seasonal Holt Winters technique seasonal indices are calculated as multiplication and proportioning (Makridakis et al., 2008).

Optimum α, β, γ parameters are found as 0.23, 0.12, 0.45 respectively

Ridge Regression

Ridge regression is one of the biased estimation methods in case of multiple linear connections, that allows all necessary variables to be modeled. This method aims to obtain parameter cross-sections with smaller variance than OLS (Ordinary least squares) estimations in case of multiple linear connections and to remove unnecessary variables from the model (Karadavut et al., 2005)

Ridge regression is generally used when there is a high degree of relationship between two or more independent variables in the model.

In multi-regression model, when the explanatory variables are related to each other, the Ridge regression method can estimate the coefficient with smaller variance than the OLS estimator. It can be used to show the uncertainties in the regression coefficients on the graph with strong multi-link effect and to remove unnecessary variables in the model (Yildirim, 2010). In the Ridge regression penalty term is set as 1.5.

Random Forest Regression

Decision tree is a supervised learning method that can be employed for both regression and classification. The decision trees consist of two objects, namely decision nodes and leaves. The leaf is the value of the target attribute and the decision node is the test value that will be applied in one attribute, followed by all possible attribute values belonging to that attribute, and these values form the branches of the tree (Tosun, 2006). In other words, the node refers to the questions, the branch refers to the answers to these questions and the leaf refers to the class in which the decision is made. With the first node of the tree, questions begin to be asked and continue until the nodes or leaves without branches. It is easy to express the decision tree by rules such as if - then structures. Rule consists of number of leaves.

Let x_i be a point in the feature space X and y_i be a class label in a set of class labels Y. We can define each of the samples in the dataset (D) as the tuple (x_i, y_i). Random Forest Regression aggregates many weak classifiers h_t into a regressor $H_{regress}$. A random forest regression is an estimator occuring of a gathering of randomized root regression trees (weak predictors) $h_t(x, \theta_m)$, in which $m \geq 1$ and $\theta_1, \ldots, \theta_N$ are independent and equally distributed outputs of a random vector θ (Biau, 2012; Breiman, 2001).

The result of these random trees is amalgamated to shape the gathered regression prediction by the following equation:

$$H_{regress}(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x, \theta) \quad (9)$$

in which $x \epsilon D_t$, training samples D_t and T stands for the number of trees in the forest.

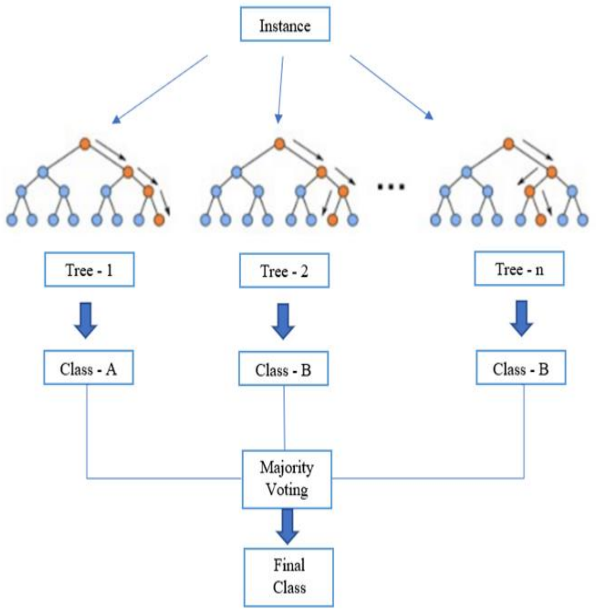

Figure 3. A Simplified Example Of A Random Forest (Chakure, 2019)

The random forest has some important advantages such as a non-parametric structure and ensuring higher classification accuracy than other classifiers. Moreover, considering the appropriate features for the characterization of training forms, it has skilled of detecting the significance of the features (Rodriguez-Galiano et al., 2012), and it is powerful against unbalanced class distribution (C. Chen et al., 2004).

In Random Forest algorithm values for number of trees and maximum depth are set as 100 and 5.

Gradient Boosting

Gradient Boosting is a tree-based Machine Learning method that utilizes boosting structure. Such as random forests Gradient Boosting also consists of ensemble of weak learners. In boosting algorithm, a strong learner is constructed from weak learners in an iterative fashion. Let $F_k(x)$ be a function that predicts output y using input x. A new weak estimator $h_x(x)$ is added to model to predict the output more accurately as shown in the following equation.

$$y = F_{k+1}(x) = F_k(x) + h_k(x) \quad (10)$$

Therefore, the new weak learner will fit the error generated by the model $F_k(x)$

$$h_k(x) = y - F_k(x) \quad (11)$$

In this study we used a variant of Gradient Boosting method called Xgboost.

In Xgboost we set the number of models to 100, max depth of each tree to 4 and learning parameter alpha to 0.01

Error Function

In order to assess the performance of studied methods, we use RMSPE (Root Mean Squared Percentage Error) metric. The RMSPE calculated as follows :

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i-\hat{y}_i}{y_i}\right)^2} \quad (12)$$

where y_i denotes the actual values, $\hat{y}_i$ denotes the corresponding prediction and n denotes the number of samples.