DEMAND FORECASTING USING STATISTICAL AND MACHINE LEARNING METHODS – A CASE STUDY IN AUTOMOTIVE INDUSTRY

EGE AK KAYIŞ

Thesis for the Master's Program in Industrial Engineering

Graduate School
Izmir University of Economics
Izmir

2023

DEMAND FORECASTING USING STATISTICAL AND MACHINE LEARNING METHODS – A CASE STUDY IN AUTOMOTIVE INDUSTRY

EGE AK KAYIŞ

THESIS ADVISOR: PROF. DR. AHMET SERMET ANAGÜN

A Master's Thesis
Submitted to
the Graduate School of Izmir University of Economics
the Department of Industrial Engineering

Izmir
2023

ETHICAL DECLARATION

I hereby declare that I am the sole author of this thesis and that I have conducted my work in accordance with academic rules and ethical behavior at every stage from the planning of the thesis to its defense. I confirm that I have cited all ideas, information and findings that are not specific to my study, as required by the code of ethical behavior, and that not all statements cited are my own.

Name, Surname: Ege Ak Kayış

Date:12.07.2023

Signature:

ABSTRACT

DEMAND FORECASTING USING STATISTICAL AND MACHINE LEARNING METHODS – A CASE STUDY IN AUTOMOTIVE INDUSTRY

Ak Kayış, Ege

Master's Program in Industrial Engineering

Advisor: Prof. Dr. Ahmet Sermet Anagün

June, 2023

Demand forecasting is an important issue in the supply chain. In increasingly competitive market conditions, it gains more and more importance with the effect of demand fluctuations due to various reasons. Demand forecasting not only helps to predict the future, it also enables to reduce the risks that may occur and to use the opportunities. This study focuses on the weekly demand forecasting for the products that have high sales volume of a company in the automotive industry. The aim of the study is to compare the performances of different approaches frequently conducted in studies of time series forecasting in the literature. In the study, Exponential Smoothing, Seasonal Decomposition, ARIMA as statistical methods, and Random Forest (RF), Artificial Neural Network (Multi-Layered Perceptron), Support Vector Regression, Sequential Minimal Optimization Regression as machine learning methods were applied. The performances of the methods were evaluated by means of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). It has been observed that seasonal decomposition (additive model) for statistical methods and random forest for machine learning methods have provided minimum MAE and RMSE values. The RF has the best performance 4,699.29 for MAE and 6,139.28 for RMSE out of all methods applied.

Keywords: demand forecasting, statistical methods, random forest, artificial neural network, support vector regression, sequential minimal optimization for regression.

ÖZET

İSTATİSTİKSEL VE MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TALEP TAHMİNİ-OTOMOTİV SEKTÖRÜNDE BİR UYGULAMA

Ak Kayış, Ege

Endüstri Mühendisliği Yüksek Lisans Programı

Tez Danışmanı: Prof. Dr. Ahmet Sermet Anagün

Haziran, 2023

Talep tahmini, tedarik zincirinde önemli bir konudur. Artan rekabetçi piyasa koşullarında çeşitli nedenlere bağlı talep dalgalanmalarının da etkisiyle giderek daha fazla önem kazanmaktadır. Talep tahmini sadece geleceğin tahmin edilmesine yardımcı olmakla kalmaz, oluşabilecek risklerin azaltılmasına ve fırsatların değerlendirilmesine olanak sağlar. Bu çalışmada, otomotiv sektöründe faaliyet gösteren bir firmanın satış hacmi yüksek olan ürünlerine yönelik haftalık talep tahminine odaklanılmaktadır. Çalışmanın amacı literatürdeki zaman serisi tahmini çalışmalarında sıklıkla kullanılan farklı yaklaşımların performanslarını karşılaştırmaktır. Çalışmada istatistiksel yöntemler olarak Üstel Düzeltme, Mevsimsel Ayrıştırma, ARIMA, makine öğrenmesi yöntemleri olarak Rastgele Orman, Yapay Sinir Ağı (Çok Katmanlı Algılayıcılar), Destek Vektörü Regresyon, Ardışık Minimal Optimizasyon Regresyon uygulanmıştır. Yöntemlerin performansları Ortalama Mutlak Hata (MAE) ve Ortalama Kare Hatanın Kökü (RMSE) ile değerlendirildi. İstatistiksel yöntemler için mevsimsel ayrıştırmanın (toplama modeli), makine öğrenmesi yöntemleri için rastgele ormanın minimum MAE ve RMSE değerlerini sağladığı gözlemlenmiştir. Rastsal orman, uygulanan tüm yöntemler arasında MAE için 4.699,29 ve RMSE için 6.139,28 ile en iyi performansa sahiptir.

Anahtar Kelimeler: talep tahmini, istatistiksel yöntemler, rastgele orman, yapay sinir ağları, destek vektör regresyonu, sıralı minimum optimizasyon.

Dedicated to my family.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my advisor Prof. Dr. Ahmet Sermet Anagün for his endless guidance, insight, and kindness throughout the research.

I would like to thank my family for always being with me and encouraging me at every stage of my life. I am grateful to have them for their endless support, love, and their moral support in completing my thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADF	Augmented Dickey-Fuller
AMT	Alternating Model Tree
ANN	Artificial Neural Network
APE	Absolute Percentage Error
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
CNN	Convolutional Neural Networks
CO	Carbon Monoxide
DT	Decision Tree
ES	Exponential Smoothing
GCN	Graph Convolutional Network
GRNN	General Regression Neural Network
IQR	Interquartile Range
K-LPS	Kernel Partial Least Squares
K-NN	K-Nearest Neighbors
LC	Lee Carter Model
LM	Linear Model
LR	Logistic Regression
LSRF	Long Short-Term Random Forest
LSTM	Long Short Term Memory
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MARS	Multivariate Adaptive Regression Splines
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MMFM	Multiple-Mode Forecasting Model
MSE	Mean Squared Error
NAR	Nonlinear Autoregressive
NB	Naïve Bayes
NMSE	Normalized Mean Square Error

QP	Quadratic Programming
PPR	Projection Pursuit Regression
RAE	Relative Absolute Error
RF	Random Forest
RMSE	Root Mean Squared Error
RR	Ridge Regression
RT	Random Tree
SARIMA	Seasonal Autoregressive Integrated Moving Average
SMFM	Single-Mode Forecasting Model
SMO	Sequential Minimal Optimization
SMOReg	Sequential Minimal Optimization for Regression
SPI	Standardized Precipitation Index
SVM	Support Vector Machine
SVR	Support Vector Regression
TA	Trend Analysis
TDNN	Time Delay Neural Network

CHAPTER 1: INTRODUCTION

1.1.Forecasting

Time series forecasting is a highly active research topic in the field of science, engineering and business. It is aimed to develop a model with time series analysis that can predict future values based on historical (observed) data in series. Due to the difficulty in assessing the exact nature of a time series, it is often quite challenging to generate appropriate forecasts (Khandelwal, Adhikari, and Verma 2015). For instance, some time series data has seasonal trends at certain time intervals of variability, some time series data is not seasonal whereas some time series have a trend, some do not. In the literature, linear and nonlinear methods have been proposed. However, in some studies it is reported that statistical and linear methods give better results whereas nonlinear methods give better results in time series data exhibiting high volatility and multi-collinearity. Today, there are studies conducted with the currently used methods, and as discussed, there are also studies to create new methods that can increase the level of precision compared to these methods.

Exponential smoothing describes a class of forecasting methods historically and some of the most successful forecasting methods are based on concept of exponential smoothing (Hyndman et al., 2008). Therefore, this method is widely included in the studies. Seasonal decomposition is preferred for time series analysis that is affected by time-varying (cyclical/periodic) factors. ARIMA models are known for their remarkable prediction accuracy and flexibility to represent various time series. However, an important limitation is the assumed linear formats of the correlated data, which makes them unsuitable for complex nonlinear time series modeling (Khandelwal, Adhikari, and Verma 2015). Exponential smoothing (ES) models deal with trend and seasonality in time series data, whereas ARIMA deals with autocorrelation in data and historical data explains present data.

Random Forest (RF) creates multiple trees and uses the output of all trees generated (ensemble learning technique). It also has an approach based on the bagging algorithm. With this approach, it reduces the problem of overfitting and variance, thus increasing accuracy and is often used in prediction for these reasons. Artificial Neural Networks require less training. It is among the most preferred prediction methods because it can detect complex, nonlinear and non-stationary relationship structures between dependent and independent variables and can detect the connections between predictive variables. Support Vector Regression (SVR) provides a competent predictive model on non-linear data such as artificial neural network (ANN). In addition, it is resistant to outliers and has high prediction accuracy. Nevertheless, it is suitable for small datasets and performs better when the dataset does not have much noise. Sequential Minimal Optimization for Regression (SMOReg) slightly reduces the difficulty of the Quadratic Programming (QP) problem through decomposition strategies compared to the SVR method, thus offering a faster learning phase to larger training data.

1.2.Problem Statement and Purpose of The Study

Demand forecasting is an important issue in the supply chain. It becomes more and more important due to the effect of demand fluctuation because of various reasons in increasingly competitive market conditions. Demand forecasting does not only help to predict the future, but also allows to reduce the risks that may occur, increase the profit margin with the financial decisions to be taken, ensure the cash flow, use the resource allocation efficiently and evaluate the opportunities that may arise for growth.

This study focuses on the weekly demand forecast for the products with high sales volume of a company in the automotive industry. In the studies in time series forecasting field, statistical methods which are among the classical methods and machine learning methods, which have become increasingly popular recently, have been conducted and interpreted by making comparisons with various error values. In some time series forecasting, statistical methods give more successful results, while in others, machine learning methods give more successful results. The performance comparisons and discussions on different methods are also included in this study.

1.3.Structure of the Thesis

The structure of this thesis is organized into six chapters. The next chapter covers literature review on time series forecasting. The third chapter discusses forecasting methods classified into two parts: statistical methods and machine learning methods. The fourth chapter demonstrates the research methodology composed of data gathering, data cleaning, data preparation, and data analysis. The results obtained from the forecasting methods are given in chapter five. The last chapter contains a conclusion and proposals for future research.

CHAPTER 2: LITERATURE REVIEW

Temizel and Casey (2005) discuss in their study whether hybrid models are better than single models. In their studies, they conduct research on the following topics: the importance of preprocessing of neural networks, the effect of trend abandonment on performance, the extent to which neural networks can model seasonality, the success of neural networks compared to linear autoregressive models, the success of ARIMA neural network hybrids compared to single models. In the training of neural networks, it has been observed that the effect of de-trendization on the result makes a significant difference, so taking a difference gives better results than trend fitting. It is predicted that neural networks can model seasonality and give better predictions when the hidden layer size is equal to the longest loop information obtained from Fourier Analysis and its size is small. The average RMSE of time delay neural networks (TDNNs) outperformed linear AR processes in two of the nine datasets, while when optimal results were compared, TDNNs outperformed AR processes in eight of the nine datasets. It was also seen that linear AR and TDNN models performed better when compared to ARIMA neural networks. This result shows that although hybrid models are highly preferred models, they can be sufficient in single models. Based on this result, they emphasized the importance of model selection because hybrid models are not always good.

Rahman (2008) studies on seasonal product demand forecast and inventory management. Bayesian techniques and probability distribution model, ARIMA and Bayes-based ARIMA models are presented. The presented models are compared in terms of inventory costs and accuracy of results. Data with missing values is studied therefore the Bayes-based ARIMA model is being developed. While developing the ARIMA model, the Box-Jenkins methodology consisting of three phases is used. Although the ARIMA model gives better results on the data set, Bayesian methodology, which is an advantage to use this methodology in non-stationary periods, has been proposed because the model becomes more complicated in non-stationary periods. Models are also compared in terms of inventory cost. MAPE and standard deviation of forecast values are used to calculate the errors. Comparisons are made according to the obtained values.

Herrera et al. (2010) works on nonlinear time series analysis to predict future water demand on an urban area. In the study, artificial neural networks (ANN), projection pursuit regression (PPR), multivariate adaptive regression splines (MARS), support vector regression (SVR) and random forest (RF) methods are presented. In order to compare the results of the models, a Monte Carlo simulation is designed and the variants of the model are compared. The success order of applied and compared machine learning methods are respectively; SVR, MARS, PPR and RF. Although the most successful result is observed with the SVR model, the success of other methods is close to the SVR model.

Wang (2011) propose kernel partial least squares (K-PLS) algorithm for time series forecasting that can be applied to non-linear time series. To evaluate the K-PLS regression approach, the seasonal ARIMA (SARIMA) model is applied and the error values are compared. This model is developed because ARIMA is a benchmark model for linear time series forecasting. However, there is a limitation, which is linearity. ARIMA models assume that the time series can be written as a linear combination of its past. However, since there is nonlinear behavior in many empirical time series, an efficient forecasting model is needed for nonlinear time series. Therefore, in the study, K-PLS based time series forecasting system is thought to meet this need and discussed.

Chen and Wang (2012) work on drought forecasting. In the study, drought prediction is made using the data of the monthly-standardized precipitation index (SPI). Random forest method is proposed for short-term and long-term prediction. According to the results obtained, the RF-based model provides a better predictive capability than the ARIMA model. At the same time, it is stated as an advantage that the RF model creates a drought forecast community rather than an average forecast. Although a few extreme droughts are detected in the prediction of the study that fall outside the 95% confidence interval, it is stated that it covers almost all observations.

Kane et al. (2014) present a retrospective analysis for predicting of avian influenza H5N1 outbreaks, using ARIMA and RF forecasting methods. The comparison of the models is made with MSE values obtained. According to the results, the RF model is more successful than ARIMA. It is also stated that ARIMA may have underperformed due to its inability to include nonlinear relationships.

Mei et al. (2014) work on real-time price prediction in the electricity market. They proposed adaptive forecasting framework for the study. As model, RF, ANN, ARMA and Adaptive RF methods are compared by MAPE values. It is obtained that RF outperformed.

Lahouar and Slama (2015) study on predicting the electrical load demand of day-ahead, by a step of one hour. The focus of the study is to compare the proposed machine learning techniques without resorting to optimization algorithms, namely RF, ANN and SVM. No optimization is applied for a fair comparison of the implemented methods. However, since ANN and SVM are highly influenced by their parameters, minor adjustments are made to get the best results. It is concluded that the accuracy of RF model is higher than the other models without optimization. Nevertheless, it is also stated that the ANN and SVM models may give better results with optimization.

Sarı (2016) studies on sales forecasting in automotive industry by using backpropagation algorithm that is used for forecasting using multiple regression analysis. In order to measure the forecast performance, multiple regression method, moving average method, exponential smoothing methods are used and compared based on MAPE values. Two types of demand forecasting are examined in the study, these are demand forecasting with regression and demand forecasting with time series. All data are normalized between [0.1, 0.9] and transferred to the program. The input and output layer consists of eight and one cell respectively. Attempts are made to find the optimal number of the hidden layer and the geometric pyramid rule is used to find the number. The comparison is made by entering [0.1, 0.9] values for the learning coefficient and the momentum coefficient. In the testing phase of the model, the artificial neural network model that gives the smallest error value from the MSE, MAPE and MAE statistical error values are selected. In addition, regression analysis is performed with eight variables (number of vehicle parks, number of exports, domestic producer price index, interest rate sales, dollar exchange rate, production number, gross domestic product, consumer price index). Afterwards, the error values are calculated using the moving average method, simple exponential smoothing method, Holt linear binary exponential smoothing method and Winters exponential smoothing method. Outputs are compared with the values obtained from the artificial neural network method and discussed.

Li et al. (2016) implement three scenarios by applying five-fold cross-validation for the prediction of Lake water level. The aim of the study is to determine the most efficient model among RF, SVR, ANN, and LM (Linear model). In the study, the effect of time lag and past water levels are considered as input of the model for real-time forecasting. The importance of variables is then analyzed. It is obtained that RF outperformed ANN, SVR, and LM by comparing values of RMSE and R^2.

Yu et al. (2017) work on real-time radar-derived precipitation forecasting by discussing random forest and support vector machine. Single-mode forecasting model (SMFM) and Multiple-mode forecasting model (MMFM) which is one, two and three hours ahead rainfall forecasting are implemented. Obtained results are compared with RMSE values. In the study, it is concluded that the MMFM model is less successful than the SMFM model, and among the SMFM models, the SVM-based model is more successful than the RF-based model.

Matei et al. (2017) investigate whether data mining, which is preferred for predicting oil moisture, is reliable in prediction and whether it can be used in real life. In this way, it is aimed that the farmers take action by using this data. In this study K-NN, SVM, ANN, logistic regression, fast large margin, decision tree (DT), random forest, linear regression are applied to compare accuracy of methods. It is obtained that K-NN gives the most accuracy among the data mining methods followed by rule induction, logistic regression, random forest and fast large margin, neural net and linear regression, support vector machine, decision tree respectively. It is said that the system could be used as early warning for emergencies.

Drisya et al. (2017) present random forest regression among the time series machine learning approached for predicting wind speed variations. In the training phase of the model, it is trained using two weeks of data consisting of the previous 12-hour values as input for every value and the accuracy is measured with the RMSE. It shows that the model trained with two weeks of data can be used to make reliable predictions up to three years later. Therefore, it is stated that random forest is suitable candidate for prediction.

Örnek (2018) presents a demand forecasting study of two phone models. First, Simple Moving Average, Weighted Moving Average (3 periods), Single Exponential Smoothing, Trend Adjusted Exponential Smoothing, Trend and Seasonal Effects and Linear Trend methods are applied and compared with MAE and MSE values. WMA method obtained the lowest error rate for the two phone models. Afterwards, SMOReg, MLP, LR, K-NN and RF methods from machine learning methods are applied and compared with MAE, RMSE and RAE(Relative Absolute Error) values. When the obtained error rates are compared, the lowest error value is obtained with the RF method.

Altınçöp and Oktay (2018) estimate particulate matter 10 (PM_{10}) and carbon monoxide (CO), which are indicators of air pollution, using ANN and RF methods, using meteorological data consisting of air temperature, humidity, wind speed and air pollutant data for Şişli (Istanbul) and Dilovası (Kocaeli) regions. The data for the year 2016-2017 are taken as a basis. Ninety percent of the data is allocated for training and prediction is made. The results show that the predictions made using RF are more successful than ANN.

Yen et al. (2018) propose two different types of solar forecasting schemes for one hour forward estimation of solar power output. SVM and RF are carried out and according to the results by comparing RMSE, MAE and MASE, RF outperforms the SVM prediction results by a significant margin. It is observed that the RF forecasting model more closely follows the actual data trend, whereas both models catch the pattern of the daily solar generation output cycle.

Akdağ (2019) studies on sales forecasting of white goods, which has seasonal, cyclical and trend patterns by using ARIMA and Nonlinear Autoregressive (NAR) Neural Network methods. ARIMA model is applied. However, since data series in the study is nonstationary, Box-Cox transformation and differencing are carried out respectively. NAR neural network is carried out with feedback delays with the values of 1-6 and hidden layer size with the value of 4. The performances of these models are compared with MSE, MAE and MAPE.

Solanki (2019) develop sales forecasting model with a hybrid approach consisting of time series analysis and feed forward ANN. ANN inputs are selected from correlation coefficients with the value more than 0.5, which is obtained from correlation analysis done for external variables. In the study two datasets are used to compare different forecasting methods. To check the stationarity Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests are performed on the time series. To calculate accuracy of the models MAPE, MAPE, MSE, normalized mean square error (NMSE), R^2 are used as performance measures. Since it is obtained that dataset is stationary, it is decided to use AR, MA, ARMA models. For training set, residual from the best result obtained with AR and MA is calculated and used to train ANN model. The proposed hybrid approach of ANN and time series technique are discussed.

Gültepe (2019) studies on air pollution by applying machine learning methods. In the study, RF, K-NN, LR, DT, LR, Naïve Bayes (NB) are compared for forecasting air pollution by applying min-max normalization to data set. Seventy percent of the data is reserved for the training set and 30% for the test set. R^2, MSE, MAE, RMSE values obtained are discussed. The top three algorithm that gives best accuracy is DT, RF and K-NN.

Şahinarslan (2019) works on population estimation with machine learning algorithms. In the study, LR, RR, Holt-Winters and ARIMA methods are used, and the performance results are compared with the calculated RMSE and MAPE values. According to the results obtained, the Holt Winters method gives the best result after ARIMA.

Dedeoğlu (2019) studies on forecasting of the number of outpatients, surgeries and inpatients in health sector. In the study moving average (MA), ES, Holt method, trend analysis (TA) and RF method that is one of the machine learning methods are applied to perform demand forecasting in the study. By comparing the error values of the future health services demand, it is observed that Holt's Linear Method, among the time series methods, gives the best results. With the RF method applied to the surgical department, the first year data and the second year data are estimated, and an estimation of 74% is achieved.

Taghiyeh (2020) studies on lumpy demand forecasting. Three neural models have been proposed: the three-layer backpropagation model, the five-layer backpropagation model and the general regression neural network (GRNN) with genetic adaptive training. In the study, it is proposed to use forecast accuracy of training and validation sets as input for training classifiers. The classifiers are used for model selection that is trained according to the relative normalized performance of the model on training set and out-of-sample performance of the model being analyzed on validation set.

Zhou (2020) studies on multiple interrelated time series forecasting with machine learning. To prevent overfitting Monte Carlo dropout technique is used by applying CNN. Also a hybrid deep learning approach with GCN and LSTM is present. Accuracy of methods applying on different datasets is compared. In the uncertainty evaluation part, LSTM, CNN, SVR, RF and XGBoost are compared. In the first model, which is a basic multilayer perception, 32 units are used as hidden layers. In the second model which is LSTM, two hidden layers with 16 neurons each followed by a final linear layer are used. In the third model, a one-dimensional CNN with 4 layers with 8 neurons followed by a final linear layer are used. These models predict uncertainty by performing Monte Carlo sampling with dropout. The uncertainty estimate constructed with 90% interval based. To train the neural network models, MAE loss function and Adam optimization are used and the models are compared with RMSE, MAE, MAPE values. In the long-term stability part, the performance of three methods are compared.

Öztürk (2020) uses machine learning methods to predict the changing demands in the clothing industry during the season. First, by looking at the performance evaluation criteria (R, MAE, RMSE) for the current season data, an estimation study is made with a method that gave better results, and it is tried to predict which product will be demanded and the number of them in case of changing demand during the season. Afterwards, parameter values such as learning rate, momentum rate and number of hidden layers are compared for MLP and RF methods, and the parameter values that give the best results are selected and the models are tested with the selected parameters. According to the results obtained, it was observed that the RF method gives better results.

Yegen (2020) study on food sales estimation by data mining method. ES and ARIMA models are used in the study. Although there is not much difference between the obtained MAPE values, it is seen that the time series model is more suitable for estimation with 5% error rate.

Tavukcu and Sennaroglu (2021) aim to reduce the cost of inventory of spare parts. Holt-Winters and ARIMA methods are proposed. The results of proposed methods and the methods that is moving average used by the company are compared with the MSE value, holding costs and lost sales costs. It is obtained that Holt-Winters' Multiplicative method is provided the most accuracy with lowest holding cost.

Javeri (2021) conducts a study comparing the traditional statistical approach with Automated ANN that is a machine learning technique for time series forecasting. In the study, data augmentation method is presented to improve the quality of out of sample predictions. In empirical analysis it is obtained that ANN outperforms on intermediate length time series forecasting where the sample size is limited when comparing traditional statistical models.

Ballı (2021) presents the time series forecasting model to predict the course/trend of the epidemic using machine learning methods for the US, Germany and global COVID-19 data. LR, ANN, RF and SVM machine learning methods were compared with RMSE, absolute percentage error (APE), MAPE values. Time series data is based on 35 weeks and it is divided into two groups, 18 weeks as training data, 17 weeks as test data. It is obtained that SVM method outperforms LR, ANN and RF.

The aim of the proposed study of Gothai et al. (2021) is to analyze the COVID-19 data and determine the number of cases that may occur in the future. In the study, LR, SVM, single, double and triple ES methods are presented and compared. According to the results obtained, LR and SVM gave less accuracy than the proposed Holt-Winter method.

Demirezen (2021) studies on gold price forecasting by using ARIMA from time series analysis and RF from machine learning methods. As a result of the estimation, it is determined that the method with the best estimation performance for both the training set and the test set according to the MAE, RMSE criterion was the RF method.

Demirezen and Çetin (2021) use RF and SVR methods to estimate the market swap price, which give a better estimation. With the obtained MAE, MAPE and RMSE values, it is concluded that the RF method predicts the market clearing price better than the SVR method.

Shakri (2021) presents a comparison between five machine learning techniques for predicting Bitcoin returns which is alternating model tree (AMT), RF, MLR, ANN and M5 Tree algorithms. Ten years of data is used in the study. Five error values, namely R, MAE, RMSE, RAE and RRSE, are used to evaluate the accuracy of the models using the color intensity method. The results of this study show that Bitcoin profit can best be predicted using the RF technique.

Budiman and Ifriza (2021) aim to predict earthquakes before occurring using information about the disaster arrival time and amplitude height from the destination station. In the stduy RF and multi RF which is one of the machine learning method are carried out to predict earthquakes. It is obtained that prediction result of two methods are relatively similar which is around 63%.

Nacar (2021) perform a sales forecasting study using LR, K-NN and RF algorithms from machine learning algorithms in the supply chain. The algorithm that gave the lowest error among the algorithms was the RF.

Hong et al. (2021) carry out to prevent the loss of insurance companies or the overpayment of premiums by people who want to take out insurance by estimating the death rates correctly. In this study, three hybrid Lee-Carter (LC) models are applied, LC-ARIMA, LC-ANN and LC-RF to estimate mortality rates in Malaysia. A comparison of the models is made based on selected error criteria such as MAPE, RMSE and AFE. The hybrid model with the best performance in predicting male and female mortality rates in Malaysia is determined to be LC-ANN and LC-ARIMA models, respectively.

Dinçoğlu (2022) studies on the data of market sales. A two-year data set is used. In order to prevent the seasonal market campaigns and discounts from creating a misleading effect on the estimation, outliers are cleaned and data analysis is performed. Regression and time series model ARIMA from sales forecasting modeling are used and error values are compared with MAPE value. It has been found to be highly successful since it gives an error rate of less than 10% for both models. However, the regression model is found to give relatively better results than the ARIMA model.

Doğan (2022) works on an algorithm that can estimate torsional strength for reinforced concrete beams with known material properties and cross-sections and uses MLR, DT, RF and SVM in the study. The success of the models is compared with the R^2, MAE and RMSE values. The SVM algorithm provided the best prediction rate according to the error rates obtained, followed by MLR, RF and DT.

Şengül (2022) aimed to predict the next day's value based on Bitcoin's closing values of previous day. SVR, RF, K-NN, which are machine learning algorithms, are selected and MSE, RMSE and R2 error rates obtained by working on 5-year data are interpreted. In the results, it has been determined that the algorithm that makes the closest estimation to the truth is RF and then K-NN, and there is a lot of deviation in the SVR model.

Çelik (2022) works on air pollution forecasting. A dataset consisting of PM10, SO2, CO, NO2, NOX, O3 pollutants and temperature, humidity, precipitation, wind speed and pressure meteorological values measured between 2016-2021 in Manisa and Zonguldak provinces are used. Modeled with SVM, DT, NB, RF, K-NN and LSTM methods and compared the success of the models with RMSE, MAE and R^2. According to the results obtained, it is seen that LSTM gives better results than other methods, and NB shows the lowest performance.

Akgök (2022) studies on forecasting vehicle price in automotive industry by using DT, RF, SVM, and ANN methods. In the study, the correlation of the features in the data set (price, mileage, model, year, fuel type, engine power, etc.) with each other is measured, and it is observed that performance of machine learning models would increase when the extreme values are removed considering the normal distribution. In the applied methods, it is observed that the RF method gives the best results with the lowest RMSE error values.

Wang et al. (2022) proposed online self-adaptive forecasting method based on random forest that is considering possible fluctuations of market and adapts by maintaining training sets of different sizes. The proposed model is compatible with LR, K-NN and SVM. The results of the simulation show that proposed long short-term random forest (LSRF) performs better than ordinary RF and other learning algorithms.

Çiftçi and Batur (2022) aim to estimate the number of applications to be made to the emergency service by using time series analysis methods SARIMA and Holt-Winters, while using RT and RF methods from machine learning methods. The estimation results are examined according to the model evaluation criteria and it is seen that the SARIMA method achieves a higher success rate than other methods. With the SARIMA method, 83.93%, 70.45% and 0.005% values are reached as R, R^2 and MAPE values, respectively. Machine learning methods show low success compared to time series analysis methods, but it is stated that the performance of machine learning methods on large data sets is high and it can be re-evaluated in a larger data set.

The extensive literature review reveals that demand forecasting may be applied via statistical and machine learning methods. Due to the circumstances of dynamics in automotive industry, certain statistical (i.e. exponential smoothing, seasonal decomposition, and ARIMA) and machine learning methods (i.e. random forest, artificial neural networks and sequential minimal optimization for regression) are applied in the study.

CHAPTER 3: FORECASTING METHODS

Forecasting methods used in this thesis for demand forecasting are examined in two groups: statistical and machine learning methods. The methods are introduced in the following subheadings.

3.1. Statistical Methods

3.1.1. Exponential Smoothing

Exponential smoothing is a widely used time series method for estimating univariate time series data. It is used to make predictions of time series data based on previous assumptions and based on a description of the trend and seasonality in the data. The Exponential Smoothing method estimates past observations by assigning exponentially decreasing weights. In other words, the weight assigned to each demand observation decreases exponentially.

The exponential smoothing was proposed by Brown (1959) without citation to previous work, formed as double exponential smoothing by Holt (1957) and triple exponential smoothing by Winters (1960).

3.1.1.1. Single Exponential Smoothing

Simple exponential smoothing is a suitable method for estimating series without significant trends and seasonality. A single parameter is used, which expresses the rate at which the effect of previous observations decreases exponentially, alpha, which is called the smoothing factor. The closer the alpha value is to 0, the more the forecast is affected by the previous data, the closer to 1 it is by the latest data.

The forecast equation is given by formula:

$$\hat{y}_{t+h|t} = l_t \qquad (1)$$

The simple exponential smoothing is given by the formula:

$$l_t = \alpha y_t + (1-\alpha) l_{t-1} \qquad (2)$$

where

α: the smoothing factor, $0 \leq \alpha \leq 1$

l_t: the level (or the smoothed value) of the series at time t, a simple weighted average of the current observation y_t

y_t: the current observation

l_{t-1}: the previous smoothed statistic

3.1.1.2. Double Exponential Smoothing (Holt's Trend-Corrected)

Holt (1957) developed the simple exponential smoothing method for better estimation of trended data and proposed a method called "double exponential smoothing" or "quadratic exponential smoothing", which is the recursive application of an exponential filter twice. The basic idea behind double exponential smoothing is to introduce a term that takes into account the probability of a series to exhibit some kind of trend. This slope component itself is updated via exponential smoothing.

As with simple exponential smoothing, the level equation below indicates that l_t is the weighted average of the observation y_t and the one-step-ahead training forecast for time t, given by $(l_{t-1} + b_{t-1})$. The trend equation shows that b_t is the weighted average of the trend estimated at time t based on the previous estimate of the trend, $(l_t - l_{t-1})$ and b_{t-1}. The forecast function is no longer flat, but in trend. The h-step ahead estimate is equal to the last predicted level plus h times the last predicted trend value. So the estimates are a linear function of h.

The forecast equation is given by formula:

$$\hat{y}_{t+h|t} = l_t + hb_t \qquad (3)$$

The level equation of double exponential smoothing is given the formula:

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \qquad (4)$$

The trend equation of double exponential smoothing is given the formula:

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \qquad (5)$$

where

α: the smoothing factor for level and $0 \leq \alpha \leq 1$

β: the smoothing factor for trend and $0 \leq \beta \leq 1$

l_t: the level (or the smoothed value) of the series at time t

b_t: the best estimate of the trend at time t

y_t: the current observation

l_{t-1}: the previous smoothed statistic

b_{t-1}: the previous estimated of the trend

3.1.1.3. Triple Exponential Smoothing (Holt-Winters' seasonal method)

Holt (1957) and Winters (1960) introduced the Holt-Winters seasonal method, extending Holt's method to consider seasonality. The Holt-Winters seasonal method includes the prediction equation and three smoothing equations: one for the level l_t, one for the trend b_t, and one for the seasonal component $s_t$with the corresponding smoothing parameters α, β, and γ. m is used to indicate the frequency of seasonality, in other words, the number of seasons in a year, m=4 for 3-month data and m=12 for monthly data.

There are two variants of the seasonal components that differ. The additive method is preferred when the seasonal changes are roughly constant throughout the series, while the multiplication method is preferred when the seasonal changes vary in proportion to the level of the series. With the addition method, the seasonal component is expressed in absolute terms in the observed series scale and seasonally adjusted by subtracting the seasonal component in the level equation. Within each year, the sum of the seasonal component will be approximately zero. With the multiplicative method, the seasonal component is expressed as relative (percentage) and the series is divided by the seasonal component and seasonally adjusted. In each year, the seasonal component will total approximately.

The forecast equation of Holt-Winters' multiplicative method is given by formula:

$$\hat{y}_{t+h|t} = (l_t + hb_t)s_{t+h-m(k+1)} \qquad (6)$$

The level equation of triple exponential smoothing is given the formula:

$$l_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1}) \qquad (7)$$

The trend equation of triple exponential smoothing is given the formula:

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \qquad (8)$$

The seasonal equation of triple exponential smoothing is given the formula:

$$s_t = \gamma \frac{y_t}{l_{t-1}+b_{t-1}} + (1-\gamma)s_{t-m} \qquad (9)$$

where

α: the smoothing factor for level and $0 \leq \alpha \leq 1$

β: the smoothing factor for trend and $0 \leq \beta \leq 1$

γ: the smoothing factor for seasonal change and $0 \leq \gamma \leq 1$

l_t: the level (or the smoothed value) of the constant part for time t

b_t: the the sequence of best estimates of the linear trend that are superimposed on the seasonal changes

s_t: the sequence of seasonal correction factors

y_t: the sequence of observations

m: the number of seasons in a year

l_{t-1}: the previous smoothed statistic

b_{t-1}: the previous estimated of the trend

3.1.2. Seasonal Decomposition

The classical decomposition method emerged in the 1920s. It is a simple procedure and forms the starting point for many other time series decomposition methods. It attempts to construct a sequence of components from a time sequence, each of which has a certain type of behavior (which can be used to reconstruct the original by additions or multiplications).

Time series are usually decomposed into trend component, cyclical component, seasonal component, irregular component (noise). The series has a trend when there is a permanent ascending or descending direction in the series. The trend component reflects the long-term progression of the series and is not necessarily linear. The cyclic component reflects repetitive but non-periodic fluctuations. The duration of these fluctuations depends on the nature of the time series. Seasonal component reflects seasonality or seasonal variation that occurs over a fixed and known period. A seasonality exists when a time series is influenced by seasonal factors. Irregular component describes random, irregular influences that represents the residuals of the time series after the other components have been removed.

There are two forms of classical decomposition: an additive decomposition and a multiplicative decomposition. An additive model would be used when the variations around the trend do not vary with the level of the time series whereas a multiplicative model would be appropriate if the trend is proportional to the level of the time series (Otexts, 2016)

Additive decomposition equation is given by the formula:

$$Y_t = T_t + C_t + S_t + I_t \qquad (10)$$

Multiplicative decomposition equation is given by the formula:

$$Y_t = T_t * C_t * S_t * I_t \qquad (11)$$

where

T_t: the trend component at time t

C_t: the cyclical component at time t

S_t: the seasonal component at time t

I_t: the irregular component at time t

3.1.3. ARIMA (Box-Jenkins Model)

Box and Jenkins (1970) present autoregressive integrated moving average (ARIMA), also known as Box-Jenkins model, which is named after the statisticians George Box and Gwilym Jenkins. ARIMA is widely used approach for time series forecasting that is applied to find the best fit model for the past values of a time series in time series analysis. ARIMA models aim to describe the autocorrelations in the data.

ARIMA is a regression analysis method that measures the strength of a dependent variable relative to other varying variables, and its purpose is to predict future values by examining the differences between the values in the series rather than the actual values. Non-seasonal ARIMA consists of a combination of difference taking and autoregression and moving average model.

Auto-Regression (AR) means that the evolving variable of interest regresses according to its lagged (ie previous) values. Integrated (I) means that the data values are replaced by the difference between their own values and the previous values. In a Moving Average (MA), the regression error is actually expressed as a linear combination of error terms whose values arose simultaneously and at various times in the past.
ARIMA(p,d,q) where the parameters p, d, and q are non-negative integers is given by formula:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (12)$$

where
y'_t : the differenced series (it may have been differenced more than once)
p: the order (number of time lags) of AR model
d: the degree of differencing (the number of times the data have had past values subtracted)
q: the order of MA model

3.2. Machine Learning Methods

3.2.1. Random Forest

Ho (1995) proposed the general method of random decision forests. Tin Kam Ho (1995) created the first algorithm for random decision forest using the random subspace method which is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg which in Ho's formulation. Then Breiman (2001) developed the random forest algorithm using his own idea of "bagging" and random selection of features, introduced first by Ho and later Amit and Geman (1997).

Random forest is a machine learning technique for solving regression or classification problems that uses ensemble learning which is a technique that combines many classifiers. The random forest algorithm consists of many decision trees.

The "forest" produced by the random forest algorithm is trained by either bagging (an ensemble meta-algorithm that improves accuracy) or bootstrapping methods. The random forest algorithm generates the result based on the predictions of the decision trees. It estimates by averaging the output from different trees, so increasing the number of trees increases the precision of the outcome (Breiman, 2001).

Random forests remove the limitations of a decision tree algorithm, avoid the habit of decision trees to over fit their training sets, and random forests generally outperform decision trees.

Random forest equation is given by the formula:

$$x = \{(x_1 + y_1), \dots, (x_N + y_N)\},\ x_i = [x_i^1, \dots, x_i^D]^T, y_i \in \{1, \dots, K\} \qquad (13)$$

Set of tests: $S = \{(g_1(x), \theta_1), \dots, (g_M(x), \theta_M)\}$ (14)

Gain: $\Delta L = L_j - \frac{|j_r|}{|j|} L_{j_r} - \frac{|j_l|}{|j|} L_{j_l}$ (15)

Feature Test: $\mathcal{G} = \{x^1, \dots, x^D\}$ (16)

Hyperplane Test: $\mathcal{G} = \{g_w(x) = w^T x | w \in R^D\}$ (17)

where

x : test sample

j : the number of rows that particular node has

j_r : the number of nodes in right node

j_l : the number of nodes in left node

L : the number of nodes

3.2.2. Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) are computing systems inspired by the biological neural networks that constitute animal brains (Yang and Yang, 2014). One of the most significant advantages of the ANN models over other classes of nonlinear models is that ANNs are universal approximators that can approximate a large class of functions with a high degree of accuracy (Chen et al., 2003; Zhang and Qi, 2005). Their power comes from the parallel processing of the information from the data. No prior assumption of the model form is required in the model building process. Instead, the network model is largely determined by the characteristics of the data. Single hidden layer feed forward network is the most widely used model form for time series modeling and forecasting (Zhang et al., 1998).

ANN consists of connected units or nodes called artificial neurons that model neurons in a brain. Like synapses in a biological brain, each connection can transmit signals to other neurons. An artificial neuron receives signals, then processes them and can send signals to neurons connected to it. The "signal" in a link is a real number, and the output of each neuron is calculated by a nonlinear function of the sum of its inputs. Connections are called edges. Neurons and edges have a weight that changes as learning progresses. Weight increases or decreases the strength of the signal on a link. Neurons can have a threshold such that a signal is sent only when the total signal exceeds that threshold.

Neural Networks have at least one input and one output layer. At the input, as many neurons as the number of inputs are used, in the output layer, as many neurons as the number of the output variable are used. The number of neurons in the hidden layer is determined by trial-and-error method by increasing the number of neurons from the least to the most, giving better results for the training and test groups. The structure of an ANN is given in Figure 1.

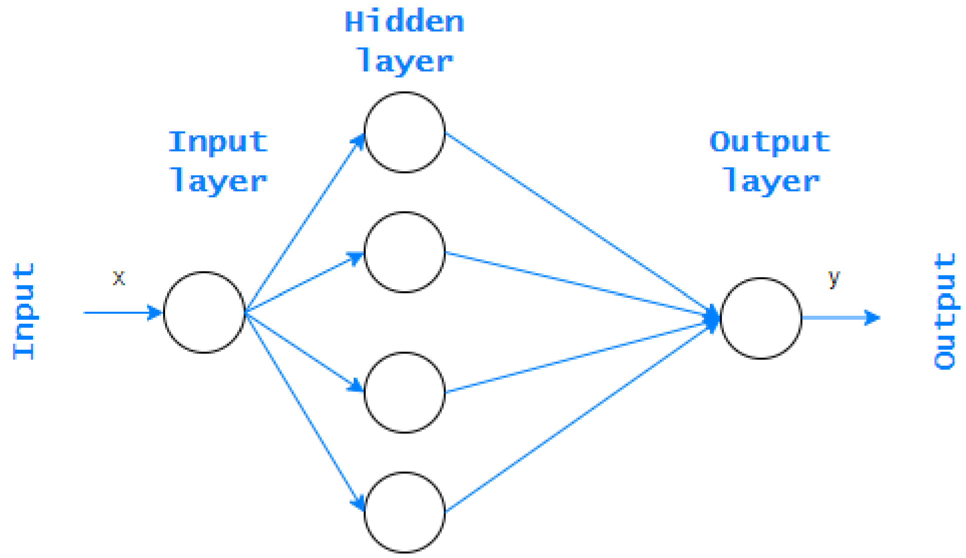


Figure 1. ANN structure

Each neuron consists of an addition function that combines the input values received and an activation function that transmits it to the other neuron. In addition to these, there are different learning algorithms used in artificial neural networks, the most commonly used one is the backpropagation learning algorithm. One of the most important features of this training function is that it reduces the possibility of over-learning, i.e., memorization. Thus, it ensures that the generalization (prediction) ability of the network remains high. It reflects the error value on the output back to the model. Therefore, back propagation algorithms need a differentiable activation function. The most frequently used activation functions are given by Şener (2019):

Logarithmic Sigmoid Function:

$$y = \log_e\left(\frac{1}{(1+e^{-x})}\right) = \ln\left(\frac{1}{(1+e^{-x})}\right) = 1/(1+e^{-x}) \qquad (18)$$

Hyperbolic Tangent Sigmoid Function:

$$y = \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right) \qquad (19)$$

Linear Function:

$$y = ax + b \qquad (20)$$

The input is a vector multiplied by the x weights and added bias.
The equation is given by the formula:

$$y = w * x + b \qquad (21)$$

A sensor produces a single output from real-valued input. In doing so, it creates a linear combination using the input weights. (sometimes its output may be a non-linear activation process). The equation is given by the formula:

$$y = \varphi(\textstyle\sum_{i=1}^{n} w_i + x_i + b) = \varphi(w^T x + b) \qquad (22)$$

where
w: weight vector
x: vector of inputs
b: bias (bias)
φ: nonlinear activation process

3.2.3. Support Vector Regression (SVR)

In machine learning, support vector machines (SVMs) are supervised learning model that analyze data for classification and regression analysis. SVM algorithm is first proposed by Vapnik and Chervonenkis(1964) and developed by suggesting a way to create nonlinear classifiers by applying the kernel trick to maximum-margin hyperplanes by Vapnik, Boser and Guyon (1992). The difference to the formally similar algorithm is that each dot product is replaced by a nonlinear kernel function. This allows the algorithm to fit the hyperplane with the maximum margin in a transformed feature space. A "soft margin" incarnation commonly used in software is developed in 1993 and published by Cortes and Vapnik (1995). A version of SVM for regression which is called support vector regression is proposed by Vapnik and with colleagues Drucker et al. (1996).

The SVM can be used as a classification method as well as a regression method with a few minor differences, while preserving the features that characterize the algorithm (maximal margin). In case of regression, a margin of tolerance (epsilon) is set. Since output is a real number, it becomes very difficult to predict which has infinite possibilities. However, the main idea is always the same, minimizing the error, individualizing the hyperplane, maximizing the margin.

SVM equation is given by the formula:

$$y = wx + b \qquad (23)$$

SVM solution is given by the formula:

$$min\frac{1}{2}\|w\|^2 \qquad (24)$$

Constraint is given by:

$$wx_i + b \geq +1, \quad \forall y_i = +1 \qquad (25)$$
$$wx_i + b \leq -1, \quad \forall y_i = -1 \qquad (26)$$

SVM solution for regression is given the formula:

$$min\frac{1}{2}\|w\|^2 + c\sum_{i=1}^{N}(\xi_i + \xi_i^*) \qquad (27)$$

Constraints are given by:

$$wx_i + b \geq +1 - \xi_i, \quad \forall y_i = +1, \quad \xi_i \geq 0 \qquad (28)$$
$$wx_i + b \leq -1 + \xi_i, \quad \forall y_i = -1, \quad \xi_i \geq 0 \qquad (29)$$

Linear SVR is given by the formula:

$$y = \sum_{i=1}^{N} (a_i - a_i^*)\langle x_i, x\rangle + b \qquad (30)$$

Non-Linear SVR is given by the formula:

$$y = \sum_{i=1}^{N} (a_i - a_i^*)K\langle x_i, x\rangle + b \qquad (31)$$

where

x_i : the training sample with the target value y_i

w : the weight vector

b : the constant

c : penalty parameter

a_i : Lagrange multipliers

ξ_i : the slack variable

Kernel functions are given by the formula:

-Linear :

$K(x_i, x_j) = (x_i * x_j)^d$, when $d = 1$, it becomes the linear kernel. (32)

-Polynominal (inhomogeneous) :

$K(x_i, x_j) = (x_i * x_j + r)^d$ (33)

-Gaussian radial basis function :

$K(x_i, x_j) = exp(-\gamma||x_i - x_j||^2)$ for $\gamma > 0$, sometimes parametrized using $\gamma = 1/2\sigma^2$ (34)

-Sigmoid function (hyperbolic tangent) :

$K(x_i, x_j) = tanh(\kappa x_i * x_j + c)$ for some $\kappa > 0$ and $c < 0$ (35)

3.2.4. Sequential Minimal Optimization for Regression

SMOReg is machine learning algorithm that Shevade et al. (2000) improved and extended the sequential minimal optimization (SMO) algorithm of Smola and Scholkopf (1998) to solve regression problems using SVM. Smola and Scholkopf's (1998) SMO algorithm is applied for training the SVR model. It is implemented by replacing all missing values and converting their nominal attributes to binary ones. Additionally, it normalizes all attributes by default. Thus, the coefficients in the output are based on the normalized data, not the original data.

SMOReg algorithm based on the Pearson VII function-based universal kernel function (PUK) is given by the formula:

$$K\left(x_i, x_j\right) = \frac{1}{[1+(\frac{2\sqrt{||x_i-x_j||^2}\sqrt{2^{(1/\omega)}-1}}{\sigma})^2]^{\omega}} \tag{36}$$

where

x_i, x_j : n-dimensional vectors

σ : Gaussian parameter

ω : the parameter that controls the shape of curve

CHAPTER 4: EVALUATION OF THE METHODS APPLIED

In this thesis, the demand forecasting of a factory operating in the automotive sector has been studied. Line/machine production plans and stocks are studied based on weekly sales values. For this reason, the weekly demands of the products were taken as a basis in the study. The most demanded products were selected as the data set. Due to the continuous renewal of the models (products), the life of the products is usually a maximum of 5 years, and mass production is not fully started in the first 6 months. For this reason, since the data between the first 6-12 months are not healthy, the study was conducted on the 4-year weekly demands of the products for training phase. Four-year data covers customer demands between 2019-2022 on a weekly basis. In order to test the model, the first 9-week actual demands in 2023 were taken. In the data analysis, it has been determined that there is no missing data since the data is stored regularly in the factory. In addition, whether there is an outlier data was investigated with the help of interquartile range method (IQR). IQR method is introduced by Tukey (1977) for construction of a boxplot to detect outliers by carried out quartiles instead of mean and standard deviation which can be applied to both symmetric and skewed data. It is less sensitive to extreme observations. IQR method of outlier detection is as follows.

$$[Q_1 - 1.5 * IQR, Q_3 + 1.5 * IQR\,] \qquad (37)$$

where

Q_1: first quartile

Q_3: third quartile

IQR: interquartile range which is the difference between inner fences and outer fences

The IQR method is applied for outlier detection. No data was excluded because there was no outlier data.

Since the training data is weekly basis covering 4 years, 208 data were studied as training data and 9-week actual data were used for testing. The study was carried out with methods that are frequently used in literature and compared with each other. When the success of statistical methods and machine learning methods were compared in the literature, statistical methods are successful in some studies, whereas machine learning methods were more successful than others. In addition, statistical and machine learning methods were also discussed and evaluated among other methods.

The models were applied and evaluated under two headings; statistical methods and machine learning methods, and to compare the performances of the models. For these, Minitab software was used for statistical methods, WEKA software was used for machine learning methods. Each method is trained by using the data consists of 208 weeks correspond to four-year demand data between 2019-2022. Once the training process is completed, each method was tested for nine weeks against the observed demand data at the beginning of 2023 and performances of the methods were compared based on MAE and RMSE criteria.

The screenshot of Minitab software used for seasonal decomposition method is shown Figure 2. The obtained data were first loaded into the Minitab software, then seasonal decomposition for additive and multiplicative methods were applied both the training data and test data, and MAE and RMSE are used for comparing the results.

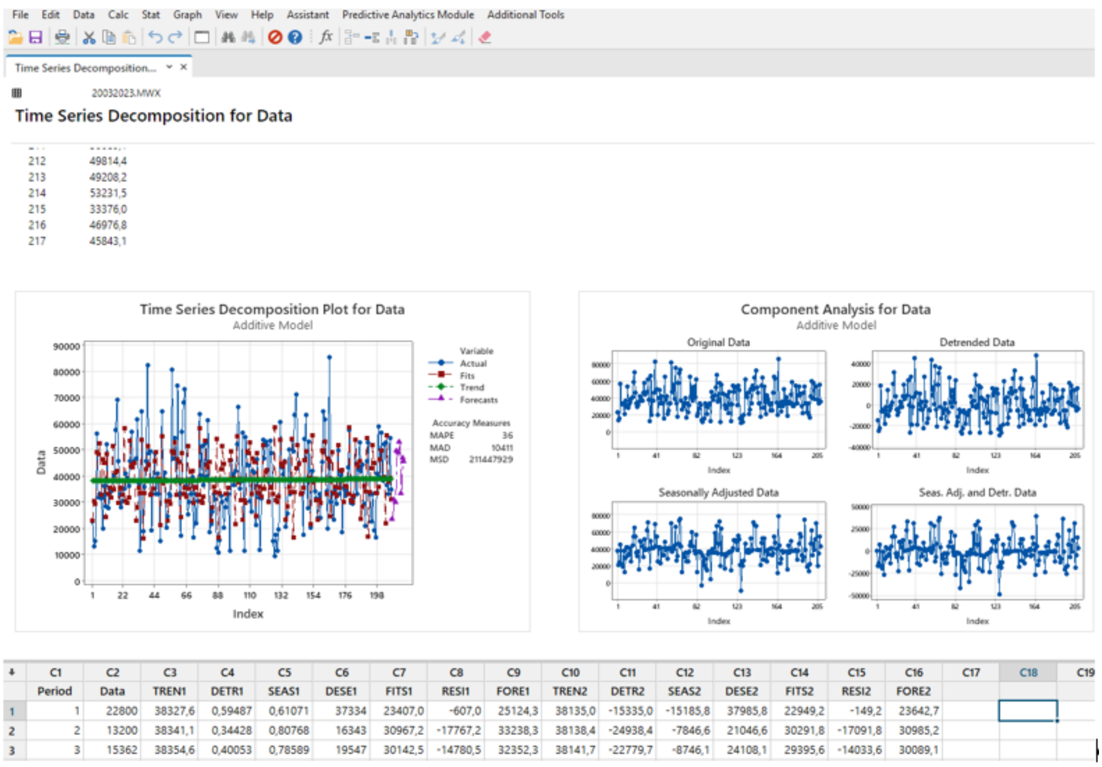


Figure 2. Screenshot of Minitab Software Used for Seasonal Decomposition

The screenshot of WEKA software used for machine learning method is shown Figure 3. The obtained data was loaded into the WEKA software. Machine learning methods were applied first on the training data and then on the test data, respectively. The results obtained were compared based on MAE and RMSE measures.

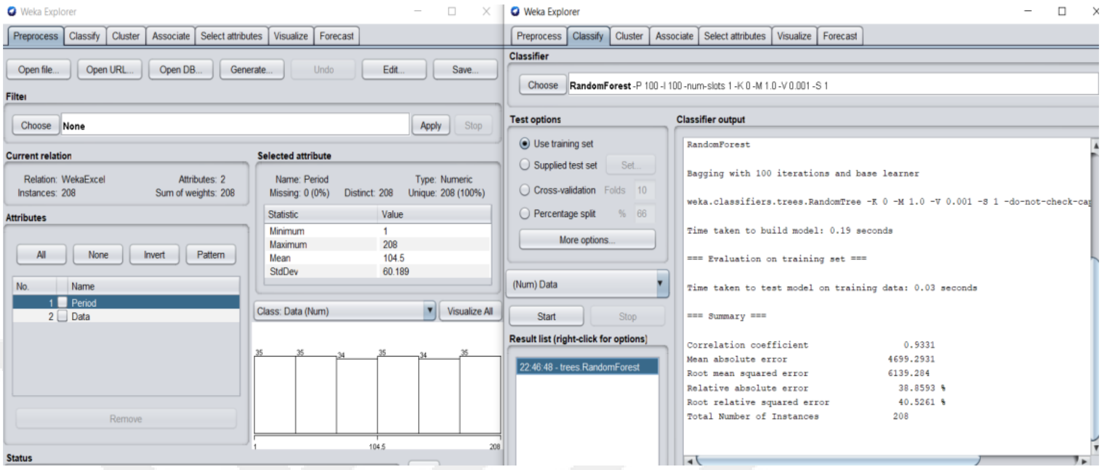


Figure 3. The screenshot of Weka Software

Since MAE and RMSE are common performance measures provided by the software used and these measures are mostly preferred in the literature, performances of the models were evaluated with these measures throughout the study. The measures of MAE and RMSE are computed as follows:

$$MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n} \qquad (38)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \qquad (39)$$

where

$\hat{y}_i$: predicted value

y_i: observed value

4.1. Results Obtained from Statistical Methods

Statistical methods related to demand forecasting introduced in Chapter 3 were applied by using 208 weekly demand data. The single exponential smoothing which is used in the analysis of stationary univariate time series without trend and seasonality, the double exponential smoothing, whose basic approach is the same as single exponential smoothing, additionally suitable for trending time series. On the other hand, triple exponential smoothing which is the most advanced method of exponential smoothing methods and suitable for time series with seasonality in addition to the trend. Seasonal decomposition additive model which is considered useful in the analysis of time series affected by periodically changing factors and where the addition method stands out, seasonal decomposition multiplicative model, in which the multiplicative method stands out, linear and non-stationary model ARIMA were used. For the ARIMA method, combination of [0-2] were tried of p (AR), d (I), q (MA) values, respectively; and the two models that provided the best results regarding both compatible with the ACF and PACF evaluations and minimum MAE and RMSE errors were included in the study.

The results obtained from statistical methods in terms of MAE and RMSE are given in Table 1 for training, in Table 2 for testing, respectively. According to the results obtained, the seasonal decomposition (additive) model has the best performance for the training data, while the ARIMA (2,1,1) model has the best performance over the test data.

Table 1. Results of Statistical Methods for Training

Model	**MAE**	**RMSE**
Single Exponential Smoothing	12,369.17	15,717.16
Double Exponential Smoothing	12,378.59	15,722.33
Triple Exponential Smoothing	13,781.74	17,306.76
Seasonal Decomposition Additive	10,410.96	14,541.25
Seasonal Decomposition Multiplicative	10,776.50	14,715.83
ARIMA (1,1,1)	11,895.70	15,064.00
ARIMA (2,1,1)	11,718.70	14,972.60

Table 2. Results of Statistical Methods for Testing

Model	MAE	RMSE
Single Exponential Smoothing	52,771.03	60,341.72
Double Exponential Smoothing	10,612.60	11,252.80
Triple Exponential Smoothing	9,360.25	10,929.30
Seasonal Decomposition (Additive)	15,450.71	17,900.45
Seasonal Decomposition (Multiplicative)	16,264.28	19,051.98
ARIMA (1,1,1)	9,392.07	10,691.13
ARIMA (2,1,1)	8,998.71	10,604.73

4.2. Results Obtained from Machine Learning Methods

As discussed in Chapter 3, four different machine learning methods namely, RF, ANN, SVR, and SMOReg were applied using WEKA software.

Parameters for the methods and the number of iterations set by WEKA software were not changed during the trials to ensure an impartial comparison, except the number of hidden neurons for ANN was varied between 2 and 8. Each method was trained with 208 weekly demand data and results including performance measures and predictions were recorded to make comparisons among methods. The performance measures of the methods for training are given in Table 3.

Table 3. Results of Machine Learning Methods for Training

Model	**MAE**	**RMSE**
Random Forest	4,699.29	6,139.28
Artificial Neural Network (1-4-1)	12,800.20	15,630.63
Support Vector Regression (Linear)	11,968.60	15,298.15
Sequential Minimal Optimization Regression (PUK)	11,486.06	15,021.02

As seen in Table 3, RF has outperformed the other machine learning methods regarding with MAE and RMSE measures. The best performance was obtained from ANN with (1-4-1) structure. For SVR, four kernels (linear, polynomial, radial basis, and sigmoid) were run and the best performance was provided by linear kernel. The performance of SMOReg with PUK (Pearson VII function-based Universal) kernel was better than SVR, but worst than ANN. To sum up briefly, the RF was the best method to forecast demand for such products within the company.

After the training has been completed, each method was tested with observed 9-week actual demand data and performance measures were calculated. The performance measures of the methods for testing are given in Table 4.

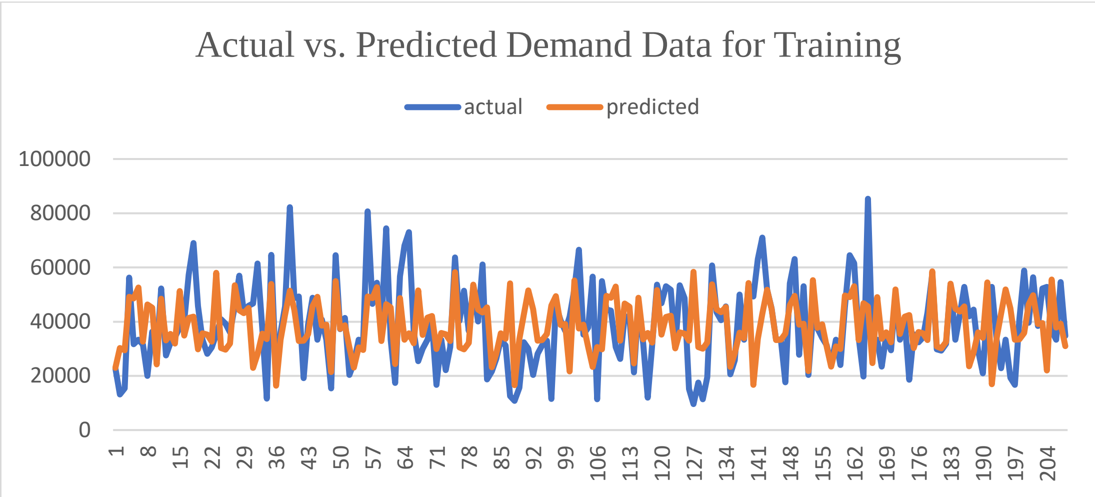
Table 4. Results of Machine Learning Methods for Testing

Model	**MAE**	**RMSE**
Random Forest	9,366.24	10,599.38
Artificial Neural Network (1-4-1)	10,693.49	11,335.66
Support Vector Regression (Linear)	8,706.04	10,778.88
Sequential Minimal Optimization Regression (PUK)	10,334.98	11,077.88

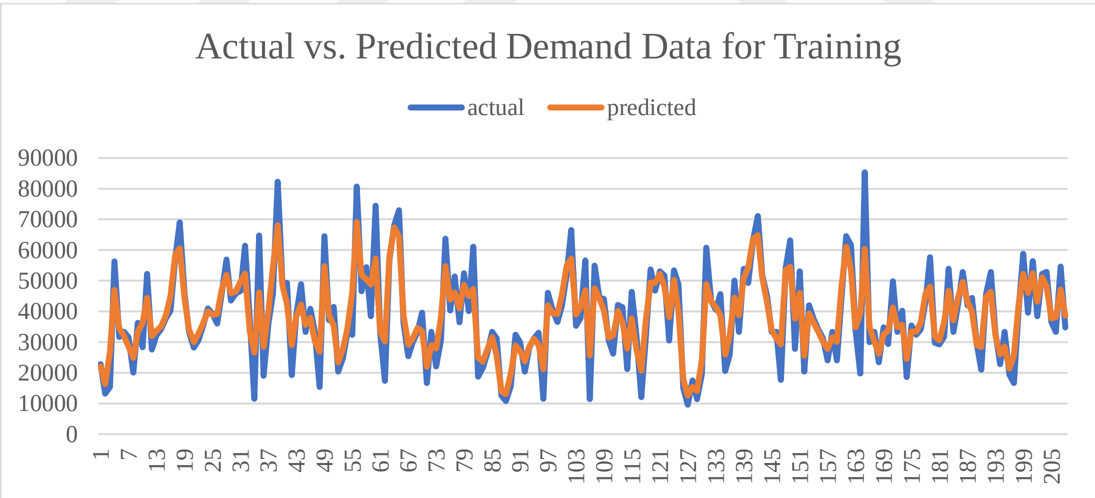
Based on Table 4, it can be said that the performance of RF method was better than other methods according to RMSE. However, as far as MAE measure is concerned, SVR with linear kernel has provided slightly higher performance than RF. In addition, SMOReg with PUK kernel was also better than ANN.

CHAPTER 5: RESULTS AND DISCUSSION

The RF method was investigated in predicted the demand for the following weeks. The Seasonal Decomposition (additive model) are compared with RF. For comparison purpose, the results obtained from SD using additive model and RF are given in Figure 4.



(a) Seasonal Decomposition (Additive Model)



(b) Random Forest

Figure 4. Compatibility of Forecasting Methods with Actual Demand Data

In Figure 4, the comparison graph of the predicted values of the actual demand data with the applied models is given. When the figure is examined, it is seen that both SD and RF methods produce good results in general. However, it can be said that the predictions made against peak points and sudden changes for these two methods are a bit more unsuccessful than the rest of the predictions. In addition, it should be noted that the applied models adapt to the demand data. Finally, when comparing the methods used, it was observed that the results produced by the RF method had a better compromise with the original data.

Figure 5 shows the scatter diagram of the actual demand data and the predicted data using the RF method. The correlation coefficient of 0.9331 shows that there is a very strong relationship between the actual and predicted demand data of the model.

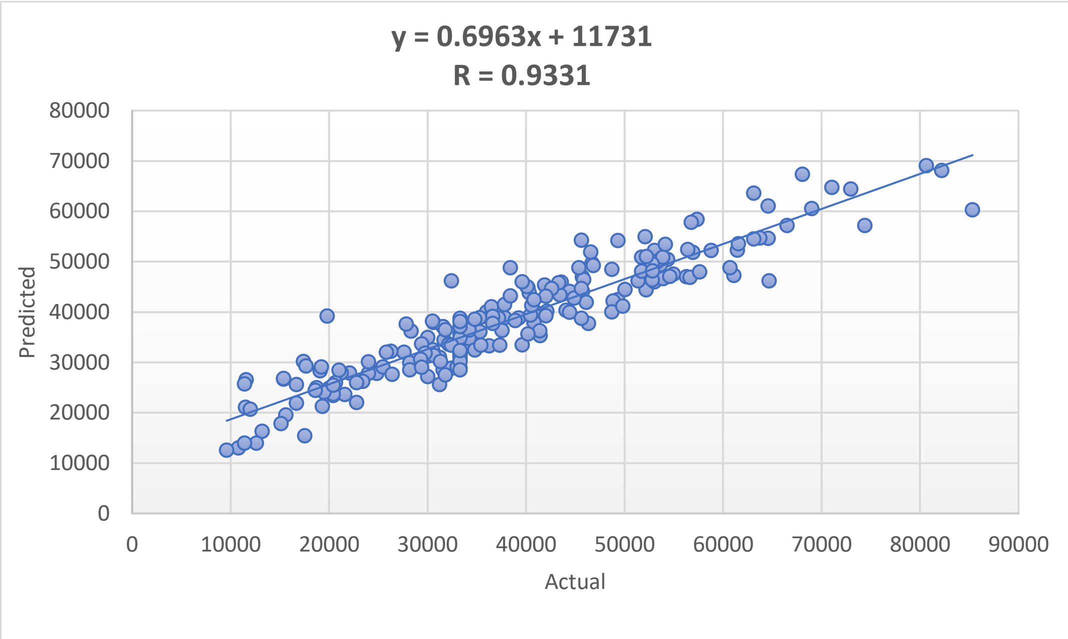


Figure 5. Scatter Diagram Between Actual and Predicted Demand Data

CHAPTER 6: CONCLUSION AND FUTURE WORK

In this thesis, statistical and machine learning methods were considered to predict demand for a runner product. The methods were applied in both the training and the testing phase, and their performances was interpreted according to the results obtained. The time series data used in the study was taken from a company in the automotive industry. The data were obtained from SAP system of the company. Since there is no outlier or missing data, no correction was needed to make on the data. Due to the continuous renewal of the models (products), and mass production is not fully started in the first 6-12 months. The first year, until the data is healthy, is not included because the data is provided from the automotive sector. ES, SD, ARIMA as statistical methods, RF, ANN, SVR, SMOReg as machine learning methods were applied.

Statistical and machine learning methods were evaluated and compared among themselves, then the best results of the two methods were compared. The comparison was evaluated with error values such as MAE and RMSE. Even though the global conjecture may be changed due to the firm's expectations, according to performance measures, the RF method with the lowest prediction errors was proposed to the company for predicting demand of a runner product for the next weeks. The firm was agreed to apply this proposal for demand forecasting for future production planning.

For further works, as the new actual demand data appear, the methods should be run for better performance. The new arrivals of demand for such products may be considered for reliable and updated predictions. On the other hand, different forecasting methods may be applied for reliable predictions depending on the different data (i.e. different independent variables other than time frame, such as currency exchange, global indicators).

REFERENCES

Akdağ, T. (2019). *Application of sales forecasting techniques for a company in the home appliances*, ProQuest Dissertations & Theses Global [Online]. Available at https://www.proquest.com/dissertations-theses/application-sales-forecasting-techniques-company/docview/2476859454/se-2 (Accessed 10 Sep 2022).

Akgök, L. (2022). *A Comparative study to forecast vehicle prices in automotive industry using machine learning techniques* [Online]. Available at: https://tez.yok.gov.tr/UlusalTezMerkezi/752791 (Accessed 15 Dec 2022).

Altınçöp, H. and Oktay, A.B. (2018). *Air pollution forecasting with random forest time series analysis*, 2018 International Conference on Artificial Intelligence and Data Processing (IDAP) [Online]. Available at: https://ieeexplore.ieee.org/document/8620768 (Accessed 20 Dec 2022).

Ballı, S., (2021). *Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods*, Chaos, Solitons and Fractals, Vol. 142 [Online]. Available at https://www.sciencedirect.com/science/article/pii/S0960077920309048 (Accessed 1 Dec 2022).

Breiman, L. (2001). *Random forests*, Machine Learning, Vol. 45, pp. 5-32.

Brown, R.G. (1959). *Statistical forecasting for inventory control*, New York: McGraw-Hill; 1959.

Budiman K. and Ifriza, Y.N. (2021). *Analysis of earthquake forecasting using random forest*, [Online]. Available at: https://shmpublisher.com/index.php/joscex (Accessed 5 Dec 2022).

Büyükşahin, Ü.Ç. and Ertekin, Ş., (2019). *Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition,* Neurocomputing, Vol. 361, pp. 151–163 [Online]. Available at https://www.sciencedirect.com/science/article/pii/S0925231219309178 (Accessed 10 Jan 2023).

Chen, A., Leung, M.T. and Hazem, D. (2003). *Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index*, Computers and Operations Research, Vol. 30, pp. 901-923.

Chen, J., Li, M. and Wang, W. (2012). *Statistical uncertainty estimation using random forests and its application to drought forecast*, Hindawi Publishing Corporation

Mathematical Problems in Engineering, Vol. 2012, pp 12.

Çiftçi, S. and Batur Sir, G.D. (2022). *An application for forecasting the number of applications to the emergency department with time series analysis and machine learning methods*, Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi [Online]. Available at: https://jag.journalagent.com/pajes/pdfs/PAJES-18488-RESEARCH_ARTICLE-CIFTCI.pdf (Accessed 5 Jan 2023).

Dedeoğlu, T. (2019). *The health sector demand forecasting*, Trakya Üniversitesi İktisadi ve İdari Bilimler Fakültesi e-Dergi, Vol. 10(1), pp. 25-38.

Demirezen, S. (2021). *Forecasting of gold prices with ARIMA and random forest regression and evaluating of their prediction performance*, Social Sciences Research Journal, Vol. 10 (3), pp. 747-758.

Demirezen, S. and Çetin, M. (2021). *Forecasting of market clearing price with random forest regression and support vector regression,* Journal of Quantitative Sciences, Vol. 3, pp. 1-15.

Dinçoğlu, P. (2022). *Sales forecasting in the retail sector with data mining techniques*. [Online]. Available at https://hdl.handle.net/20.500.12415/9305 (Accessed 5 Dec 2022).

Doğan G. (2022). *Estimation of torsional moment of reinforced concrete beams with machine learning algorithms*, El-Cezerî Journal of Science and Engineering, Vol. 9, pp. 912-924.

Drisya, G.V., Valsaraj, P., Asokanb, K. and Kumara, K.S. (2017). *Wind speed forecast using random forest learning method,* International Journal on Computer Science and Engineering, Vol. 9.

Duygu Çelik, B. (2022). *Air pollution prediction with machine learning methods: Manisa and Zonguldak example* [Online]. Available at: https://tez.yok.gov.tr/UlusalTezMerkezi (Accessed 15 Dec 2022).

Gardner, E. S. and McKenzie, E. (1985). *Forecasting trends in time series,* Management Science, Vol. 31(10), pp. 1237–1246.

Gothai, E., Thamilselvan, R., Rajalaxmi, R.R., Sadana, R.M., Ragavi, A., Sakthivel R. (2021). *Prediction of COVID-19 growth and trend using machine learning approach*, https://www.sciencedirect.com/science/article/pii/S2214785321029059 (Accessed 5 Dec 2022).

Gültepe Y. (2019). *A comparative assessment on air pollution estimation by machine learning algorithms*, European Journal of Science and Technology, Vol. 16, pp. 8-1.

Herrera, M., Torgo, L., Izquierdo, J. and Pérez-Garcı, R. (2010). *Predictive models for forecasting hourly urban water demand,* Journal of Hydrology, Vol. 387, pp. 141–150 [Online]. Available at https://www.sciencedirect.com/science/article/pii/S0022169410001861 (Accessed 5 Dec 2022).

Holt, C. C. (1957). *Forecasting trends and seasonal by exponentially weighted averages*, ONR Memorandum No. 52, Carnegie Institute of Technology, Pittsburgh, USA, International Journal of Forecasting 2004, Vol. 20, pp. 5–13.

Hong, W.H., Yap, J.H., Selvachandran, G., Thong, P.H. and Son, L.H. (2021). *Forecasting mortality rates using hybrid Lee–Carter model, artificial neural network and random forest*, Complex & Intelligent Systems, Vol. 7, pp. 163–189.

Hyndman, R., Koehler, A. B., Ord, J. K. and Snyder, R.D. (2008). *Forecasting with exponential smoothing: the state space approach*, Springer Science and Business Media.

Javeri, I. (2021). *Comparative analysis of statistical and machine learning models for time-series forecasting,* Dissertations & Theses Global [Online]. Available at https://www.proquest.com/dissertations-theses/comparative-analysis-statistical-machine-learning/docview/2553829782/se-2 (Accessed 20 Sep 2022).

Kane M.J., Price, N., Scotch, M. and Rabinowitz, P. (2014). *Comparison of ARIMA and random forest time series models for prediction of avian influenza H5N1 outbreaks*, BMC Bioinformatics, Vol. 15, pp. 276 [Online]. Available at: http://www.biomedcentral.com/1471-2105/15/276 (Accessed 8 Dec 2022).

Khandelwal, I., Adhikari, R. and Verma, G., (2015). *ANN models based on DWT decomposition*, Procedia Computer Science, Vol. 48, pp. 173 – 179 [Online]. Available at https://www.sciencedirect.com/science/article/pii/S1877050915006766#section-cited-by (Accessed 5 Jan 2023).

Lahouar A. and Ben Hadj Slama, J. (2015). *Day-ahead load forecast using random forest and expert input selection*, Energy Conversion and Management, Vol. 103, pp. 1040–1051 [Online]. Available at https://www.sciencedirect.com/science/article/pii/S0196890415006925 (Accessed 1 Dec 2022).

Li, B., Yang, G., Wan, R., Dai, X. and Zhang, Y. (2016). *Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China*, Hydrology Research, Vol. 47, pp. 69–83.

Matei, O., Rusu, T., Petrovan, A. and Mihut, G. (2017). *A Data Mining System for Real Time Soil Moisture Prediction*, Procedia Computer Science, Vol. 181, pp. 837- 844 [Online]. Available at https://www.sciencedirect.com/science/article/pii/S1877705817310603 (Accessed 5 Dec 2022).

Mei, J., He, D., Harley, R., Habetler, T. and Qu., Gç (2014), *A random forest method for real-time price forecasting in New York Electricity Market*, 2014 IEEE PES General Meeting | Conference & Exposition [Online]. Available at: https://ieeexplore.ieee.org/document/6939932 (Accessed 17 Dec 2022).

Nacar, E.N. (2021). *Sales forecasting via machine learning algorithms*, Journal of Industrial Engineering, Vol. 32(2), pp. 307-320.

Örnek, Z (2018). *Demand Forecasting in Mobile Phone Industry* [Online]. Available at: https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=Fha46f910qKfkmjED-_65g&no=bLhsvAK4pU45TGRNtQe2AQ (Accessed 8 Dec 2022).

Öztürk, S. (2020). *Forecasting of sales quantites by machine learning methods in fast fashion sector* [Online]. Available at: https://polen.itu.edu.tr:8443/server/api/core/bitstreams/b75035bd-ef33-4d45-ae26-6b98532c81e6/content (Accessed 15 Dec 2022).

Rahman, M. A. A. (2008). *Stochastic demand forecast and inventory management of a seasonal product supply chain system,* ProQuest Dissertations & Theses Global [Online]. Available at https://www.proquest.com/dissertations-theses/stochastic-demand-forecast-inventory-management/docview/2675226915/se-2 (Accessed 5 Sep 2022).

Şahinarslan, F.V. (2019). *Machine learning algorithms to forecast population: Turkey example* [Online]. Available at: https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=Wme4mi9GzH-8YEWAKPIX5Q&no=V1WzSdSAB3uLLihNQ2BAlw (Accessed 25 Dec 2022).

Sarı, M. (2016). *Yapay sinir ağları ve bir otomotiv firmasinda satış talep tahmini uygulaması*, ProQuest Dissertations & Theses Global [Online]. Available at https://www.proquest.com/dissertations-theses/yapay-sinir-ağlari-ve-bir-otomotiv/docview/2700373813 (Accessed 6 Sep 2022).

Şengül, Z. (2022). *Bitcoin price presiction using machine learning algorithms* [Online]. Available at: https://dspace.trakya.edu.tr/xmlui/handle/trakya/8165 (Accessed 8 Dec 2022).

Seong B. and Lee K. (2021). *Intervention analysis based on exponential smoothing methods: applications to 9/11 and COVID-19 effects*, Economic Modelling, Vol. 98 (2021), pp. 290-301 [Online]. Available at https://www.sciencedirect.com/science/article/pii/S0264999320312505 (Accessed 5 Sep 2022).

Shakri, I.H. (2021). *Time series prediction using machine learning: a case of Bitcoin returns*, Studies in Economics and Finance, Vol. 39, pp. 458-470 [Online]. Available at https://www.emerald.com/insight/1086-7376.htm (Accessed 5 Dec 2022).

Shevade, S.K., Keerthi, S.S., Bhattacharyya, C. and Murthy, K.R.K. (2000). *Improvements to SMO Algorithm for SVM Regression*, IEEE Transactions on Neural Networks, Vol. 11, pp. 1188-1193.

Smola, A.J. and Scholkopf, B. (1998). *A tutorial on support vector regression*, Statistics and Computing, Vol. 14, pp. 199-222.

Solanki, C. (2019). *Sales forecasting using a hybrid approach of time series analysis and artificial neural networks*, ProQuest Dissertations & Theses Global [Online]. Available at https://www.proquest.com/dissertations-theses/sales-forecasting-using-hybrid-approach-time/docview/2297413542/se-2 (Accessed 8 Sep 2022).

Taghiyeh, S. (2020). *Time series forecasting using machine learning: development and extensions*, ProQuest Dissertations & Theses Global [Online]. Available at https://www.proquest.com/dissertations-theses/time-series-forecasting-using-machine-learning/docview/2432924272/se-2 (Accessed 15 Sep 2022).

Tarmanini, C., Sarma, N., Gezegin, C. and Ozgonenel, O., (2023). *Short term load forecasting based on ARIMA and ANN approaches,* Energy Reports, Vol. 9, pp. 550-557 [Online]. Available at https://www.sciencedirect.com/science/article/pii/S2352484723000653 (Accessed 10 Jan 2023).

Taskaya Temizel, T. and Casey, M.C. (2005). *A comparative study of autoregressive neural network hybrids*, Neural Networks, Vol. 18, pp. 781–789.

Taskaya-Temizel, T., Casey, M.-C. and Ahmad, K. (2005). *Pre-processing inputs for optimally-configured time-delay neural networks*, IEE Electronic Letters, Vol. 41, pp. 198–200.

Tavukcu, A.S. and Sennaroglu B. (2021). *Applying forecasting methods to reduce the cost of spare parts inventory in a company*, Journal of Industrial Engineering, Vol. 32(3), pp. 396-413.

Temizel., T.T. and Casey, M.C., (2005). *A comparative study of autoregressive neural network hybrids,* Neural Networks, Vol. 18, pp. 781–789 [Online]. Available at https://www.sciencedirect.com/science/article/pii/S0893608005001401 (Accessed 10 Jan 2023).

Wang, H. (2011). *Using kernel partial least squares (K-PLS) regression for time-series forecasting*, ProQuest Dissertations & Theses Global [Online]. Available at https://www.proquest.com/dissertations-theses/using-kernel-partial-least-squares-k-pls/docview/921615485/se-2 (Accessed 5 Sep 2022).

Wang, P., Xu., K., Ding, Z., Du, Y., Liu, W., Sun, B., Zhu, Z. and Tang, H. (2022). *An online electricity market price forecasting method via random forest*, IEEE Transactions on Industry Applications, Vol. 58(6) [Online]. Available at: https://ieeexplore.ieee.org/document/9855865 (Accessed 20 Dec 2022).

Winters, P.T. (1960). *Forecasting sales by exponentially weighted moving averages*, Management Science, Vol. 6 (3), pp. 324-342.

Yang, Z.R. and Yang, Z. (2014). *Comprehensive biomedical physics,* Karolinska Institute, Stockholm, Sweden: Elsevier.

Yegen, N. (2020). *Sales forecasting in retail sector with data mining* [Online]. Available at: https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=Z5VEeCQbgf7T896Ya0Lw-g&no=_7Ke_VztAjNgEf3pZbC3Rw (Accessed 5 Jan 2023).

Yen, C.F., Hsieh, H.Y., Su, K.W., Yu, M.C. and Leu, J.S. (2018). *Solar power prediction via support vector machine and random forest,* E3S Web of Conferences 69, 01004.

Yu P.S., Yang T.C., Chen S.Y., Kuo C.M. and Tseng H.W. (2017). *Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting*, Journal of Hydrology, Vol. 552, pp. 92-104 [Online]. Available at https://www.sciencedirect.com/science/article/pii/S0022169417304225 (Accessed 1 Dec 2022).

Zhang, G., Patuwo, B.E. and Hu, M.Y. (1998). *Forecasting with artificial neural networks: the state of the art*, International Journal of Forecasting, Vol. 14, pp. 35-62.

Zhang, G.P., (2003). *Time series forecasting using a hybrid ARIMA and neural network model,* Neurocomputing, Vol. 50, pp. 159 – 175 [Online]. Available at https://www.sciencedirect.com/science/article/pii/S0925231201007020 (Accessed 5 Jan 2023).

Zhang, G.P. and Qi, G.M.(2005). *Neural network forecasting for seasonal and trend time series*, European Journal of Operational Research, Vol. 160, pp. 501-514.

Zhou, R. (2020). *Time series algorithms in machine learning - a graph approach to multivariate forecasting,* ProQuest Dissertations & Theses Global [Online]. Available at https://www.proquest.com/dissertations-theses/time-series-algorithms-machine-learning-graph/docview/2451176711/se-2 (Accessed 15 Sep 2022).

Otexts. (2016). *Time series components* [Online]. Available at: https://otexts.com/fpp2/components.html (Accessed 3 January 2023).