THE REPUBLIC OF TURKEY
BAHÇEŞEHİR UNIVERSITY

SALES PREDICTION IN THE FAST-FOOD SECTOR USING TIME SERIES DATA

Master Thesis

KORAY YILMAZ

İSTANBUL, 2020

THE REPUBLIC OF TURKEY

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

COMPUTER ENGINEERING

SALES PREDICTION IN THE FAST-FOOD SECTOR USING TIME SERIES DATA

Master's Thesis

KORAY YILMAZ

Thesis Supervisor: ASSOC. PROF. DR. SÜREYYA AKYÜZ

İSTANBUL, 2020

THE REPUBLIC OF TURKEY

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

COMPUTER ENGINEERING

Name of the thesis	: Sales Prediction In The Fast-Food Sector Using Time Series Data
Name/Last Name of the Student	: Koray YILMAZ
Date of the Defense of Thesis	: 07/01/2020

The thesis has been approved by the Graduate School of Natural and Applied Sciences.

Assist. Prof. Dr. Yücel Batu SALMAN
Graduate School Director

I certify that this thesis meets all the requirements as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Tarkan AYDIN
Program Coordinator

This is to certify that we have read this thesis and we find it fully adequate in scope, quality and content, as a thesis for the degree of Master of Science.

Examining Comitee Members	Signature
Thesis Supervisor Assoc. Prof. Dr. Süreyya AKYÜZ	--------------------------
Member Assoc. Prof. Dr. Birsen Eygi ERDOĞAN	--------------------------
Member Prof. Dr. Alper TUNGA	--------------------------

ACKNOWLEDGEMENTS

I want to express my gratitude to my thesis advisor Assoc. Prof. Dr. Süreyya AKYÜZ For brilliantly planning, helping and giving important knowledge in the process.

I'm thankful for Assist. Prof. Dr. Tarkan AYDIN who gave his precious contributions and to professors of the institute who shared their knowledge in all masters study.

I'm always grateful to my family who have supported and encouraged me all the time.

İstanbul, 2020 Koray YILMAZ

ABSTRACT

SALES PREDICTION IN THE FAST-FOOD SECTOR USING TIME SERIES DATA

Koray YILMAZ

Computer Engineering

Thesis Supervisor: Assoc. Prof. Dr. Süreyya AKYÜZ

January 2020, 44 Pages

What is prediction? This question has always been a quest for the future since the beginning of history. People want to predict future events to find solutions from today in order to make plans. Corporations attach enormous importance to sales predictions. Also, the fast-food sector raised its revenue in remarkable amounts and developed a need for sales prediction. All the sales companies, as well as the fast-food firms, also want to predict their future sales. Because of this, the methods that identify sales predictions with the best accuracy are searched. In this thesis, to have adequate results in sales prediction, it was foreseen to use traditional time-series models. Thus, the present study aims to find the most successful models. In sales prediction, the accuracy rates of models were tried to be identified. The models were suggested according to the success rates for this dataset. For this identification, as a method, the Correlation and Partial Correlation graphs, Akaike Information Criterion, Bayesian Information Criterion, Root Mean Square Errors, and Standard Deviation criteria were used. In this research, different results for the model that predicts the best for the fast-food data were found. The Autoregression and Vector Autoregression models were identified as giving the best results among the examined models for these data.

Keywords: Sales Prediction, Autoregression (AR), Vector Autoregression (VAR), Fast-Food.

ÖZET

ZAMAN SERİSİ VERİLERİ KULLANILARAK HAZIR YEMEK SEKTÖRÜNDE SATIŞ TAHMİNİ

Koray YILMAZ

Bilgisayar Mühendisliği

Tez Danışmanı: Doç. Dr. Süreyya AKYÜZ

Ocak 2020, 44 Sayfa

Tahmin etmek ne demektir? Tarihin başlangıcından beri geleceği arayış olmuştur. Gelecek olayları tahmin etmek istenir öyle ki bugünden çözüm bulmak ve planlar yapmak amaçlanır. Firmalar, satış tahminlerine çok önem vermektedir. Hazır yemek sektörü de gelirini dikkate değer miktarlarda arttırmış, satış tahminlerine ihtiyaç duyar hale gelmiştir. Tüm satış firmaları gibi, hazır yemek firmaları da gelecek satışlarını tahmin etmek istemektedir. Bunun için, satış tahminlerini en başarılı yakınlıkla tespit eden yöntemler aranmaktadır. Bu tez çalışmasında, satış tahminlerinde yeterli sonuç elde etmek için, geleneksel zaman serileri modellerinin kullanılması öngörülerek en başarılı modelin bulunması amaçlanmıştır. Satış tahminlerinde, modellerin başarı oranları tespit edilmeye çalışılmıştır. Başarı oranlarına göre, bu veri için kullanılabilecek modeller önerilmiştir. Bu tespit için, yöntem olarak Korelasyon ve Kısmi Korelasyon grafikleri, Akaike Bilgi Kriteri, Bayesian Bilgi Kriteri, Ortalama Hata Kareleri Kökü ve Standart Sapma ölçüm kriterleri kullanılmıştır. Bu araştırmada, hangi modelin hazır yemek verisi ile en iyi tahmin yaptığı konusunda değişik sonuçlar bulunmuştur. Yöntemde kullandığımız ölçüm kriterlerine göre, Özbağlanım ve Vektör Özbağlanım modellerinin, bu veri için incelenen modeller içinde, en iyi sonuçları verdiği tespit edilmiştir.

Anahtar Kelimeler: Satış Tahmini, Özbağlanım (ÖB), Vektör Özbağlanım (VÖB), Hazır Yemek.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

ACF	:	Autocorrelation Function
AIC	:	Akaike Information Criterion
ANN	:	Artificial Neural Network
AR	:	Autoregression
ARIMA	:	Autoregressive Integrated Moving Average
ARMA	:	Autoregressive Moving Average
BIC	:	Bayesian Information Criterion
BPNN	:	Backpropagation Neural Network
GARCH	:	Generalized Autoregressive Conditional Heteroscedasticity
HWES	:	Holt Winter’s Exponential Smoothing
MA	:	Moving Average
PACF	:	Partial Autocorrelation Function
RMSE	:	Root Mean Square Error
SARIMAX	:	Seasonal Autoregressive Integrated Moving Average
SES	:	Simple Exponential Smoothing
Std. Dev.	:	Standard Deviation
SVM	:	Support Vector Machine
VAR	:	Vector Autoregression
VARMA	:	Vector Autoregression Moving Average

SYMBOLS

A selected constant in the equation of time-series models	:	c
A weight value which affects the pattern of time-series models	:	ϕ_p
Backshift time operator	:	B
Error term	:	ε_t
First weight value in time series equations	:	θ_1
Past value of observation at time t-p	:	y_{t-p}
Residual sums of square	:	RSS
Sum of n terms	:	$\sum_{i=1}^{n}$
The level of smoothing at time t	:	ℓ_t
The logarithm function	:	log()
The likelihood of the model that was proposed	:	$L(\hat{\theta})$
The number of parameters that the model estimates	:	k
The smoothing parameter	:	α
Value of observation value of time series at time t	:	y_t
Value of prediction at time t	:	$\hat{y}t$

1. INTRODUCTION

From the first days of agriculture, producing sufficient food and nutrients was a goal so that active life could thrive. The food production adapted to a changing demographic structure, consumer preferences, socio-economic conditions, environmental concerns, science, and technology developments. This food sales system has several effects on many areas of life, with many processes, beyond nutritious foods. The food production process evolved in time and changed in a complex way (Nesheim et al. 2015, p. 21). As a result of these developments, food companies need more sales predictions to know their future sales.

The Gross Domestic Product (GDP) in Turkey was worth 766.51 billion US dollars in 2018. In 2011, the consumer food sector was 15 billion US dollars, and in 2016, it was 17.5 billion US dollars; fast-food restaurants have a share of 15% of these sales in Turkey (Türkiye-Avrupa Eğitim ve Bilimsel Araştırmalar Vakfı (TAVAK) 2018). The sales sector requires predictions to produce more realistic designs. The fast-food sector is a developing field in sales and needs future predictions. The fast-food sector has been increasing quickly in recent years because of societal-economic reasons. When the data from the USDA's Economic Research Service are examined, however, sales at fast food establishments increased the most in the mid-1980s, while institutional food sales decreased (Nesheim et al. 2015, p. 197).

In the food retail industry, accurate sales forecasting plays a significant role, given that the stock management benefits from these predictions. The more accurate sales prediction means an increase in sales and customer satisfaction, and less wasted foods.

The time series models can be applied in the food retail industry, forecasting at different time intervals. In fast-food or fresh fruit sectors, for example, short-term forecasting is preferable, as they have a short shelf life. In these cases, daily forecasting is more advantageous. Agreeing to this, as a prediction method, time intervals were utilized on an hourly, daily, and weekly basis.

Overstocking and understocking errors are two major mistakes if wrong forecasting is built. In the food sector, overstocking may bring various problems, such as food waste, shrinking, thinning out the prices, or not enough shelves. On the other hand, low stocking brings on other problems, including losing the confidence of customers, damage in market image, or sales losses (Arunraj and Ahrens 2015, p. 322).

Finding the best classical time-series model in the prediction of fast-food sales is very important in order to overcome the problems reported above. The most classical solution to this problem is to utilize univariate and multivariate regression models where the best solutions are obtained according to Root Mean Square Error (RMSE), Akaike information criterion (AIC), Bayesian information criterion (BIC), and Standard Deviation criteria.

Time series analysis and modeling are developing science area for researchers for decades. Time series analysis, modeling aims to collect data that have been formed by past observations in order to develop a suitable model that describes the structure of the series (Adhikari and Agrawal 2013, p. 9).

Prediction is the main aim in forming a model. The model is used and by understanding the previous structure and predict the future. As it is mentioned in papers, the time-series predictions are used for determining the future data, which is based on historical values (Raicharoen, Lursinsap, and Sanguanbhoki 2003, p. 741). The time series prediction, or forecasting, is widely applied in many areas, such as engineering, science, finance, and business, which relate to the revenue prediction (Adhikari and Agrawal 2013, p. 9). In this study, the suitable model for the food sales dataset obtained from a Turkish company was found by comparing some classical regression models, such as Autoregressive, Moving Average, Autoregressive Moving Average, Autoregressive Integrated Moving Average, Seasonal Autoregressive Integrated Moving Average, Vector Autoregression, Vector Autoregression Moving Average, Simple Exponential Smoothing, and Holt Winter's Exponential Smoothing.

The Autoregressive Integrated Moving Average (ARIMA) model is the most popular and classical time series analysis method in the literature (Alsudani and Liu 2017, p. 667). In this model, it was assumed that the data are linear time series and have a known statistical distribution, like a normal distribution (Adhikari and Agrawal 2013, p. 9).

“Autoregressive (AR), Moving Average (MA), and Autoregressive Moving Average (ARMA) models” (Natrella et al. 2019), which are defined as subclasses of ARIMA models, are also a part of this research. The “Seasonal Autoregressive Integrated Moving Average (SARIMAX) model” (Hyndman and Athanasopoulos 2018, p. 252), which is proposed by Box Jenkins, a variation of the ARIMA model, and a seasonal approach to ARIMA were also tested.

The popularity of ARIMA models comes from their success in representing several varieties of time series with simplicity. Using Box-Jenkins methodology can result in finding the optimal model for the time series prediction process (Zhang 2007). However, ARIMA models have some shortcomings. For example, they are used for linear forms of time series data. Various approaches were used and were proposed for nonlinear models in the literature, such as Neural Network Models (Zhang 2003).

Another method in the time series forecasting is the development of Vapnik's Support Vector Machine (SVM) conceptual. In 1995, the method was proposed to be used in the time series forecasting the difference between SMV and Artificial Neural Networks or traditional models is that SMV proposes a better, optimal solution than the others (Kim 2003, p. 308).

Vector autoregression (VAR) models and Vector Autoregression Moving Average (VARMA) models were used in this thesis since these two models became the most popular and classical methods for multivariate time series analysis. The VAR model is a natural extension of univariate models to multivariate autoregression models in the time-series analysis (Zivot and Wang 2002).

Other methods that were used in this study are the Exponential Smoothing models: Simple Exponential Smoothing (SES) and Holt Winter's Exponential Smoothing (HWES). In 1956, Robert Goodell Brown suggested Exponential smoothing without referencing past works in the statistical literature, and then his work was improved upon by Charles C. Holt in 1957 (Hyndman and Athanasopoulos 2018, p. 183). These methods were implemented with our dataset in this study.

1.1 TIME SERIES IN STATISTICAL LEARNING

There are several definitions of time series in the literature. One of the definitions is as follows: “an ordered sequence of values of a variable at equally spaced time intervals” (Natrella et al. 2019). Another detailed definition is “a series of observations, x_t, observed over a period of time” (Rao 2018, p. 10).

Time Series Applications: The usage of time series often can be divided into two branches;

a) Understanding the effects, causes, and structure of the past data

b) Fitting a model and continuing to forecast and monitior (Natrella et al. 2019).

Time series analysis is defined as the systematic way of answering statistical and mathematical questions about correlations in time (Shumway and Stoffer 2011, p. 1).

The time-domain approach was explained by Shumway and Stoffer (2011) as follows. The time-series approach is motivated from the explanation of the correlation between adjacent observations. These observations are past values and current ones. The aim is to find which model explains this relation better. Then, the model and the past values are used to find future values.

Another explanation of time series analysis can be explained briefly, as follows. Fitting a time series into a model is called a Time Series Analysis. The parameters of the proposed model are estimated by the given historical data (Hipel and McLeod 1994, p. 65). In a time-series prediction, the past observations are used to produce a suitable mathematical model that provides information (Zhang 2007). The future outcomes are then predicted using the model, which is sometimes called *forecasting*. Time-series forecasting has important applications in diverse disciplines. Often, valuable strategic decisions and important critical measures are taken based on the forecast results. Thus, making a good forecast, which involves fitting an adequate model to a time series, is significant. Over the past several decades, many attempts have been realized by researchers for the evolution and improvement of suitable time-series forecasting models. In the literature review section, these endeavors were covered in particular (Adhikari and Agrawal 2013).

If it is desired to make an example and a landmark in these efforts, it can be the ARIMA model work of Box and Jenkins (1970; see also Box et al. 1994). The developed systematic model called ARIMA is the most used classical approach to cover time-correlated modeling and prediction.

Agreeing to the definitions noted above, the dataset in the fast-food sector may include observations that form a time series. The dataset was collected into three different time intervals. By hourly, each hour's revenue of the fast food firm branch's sales was probed, and daily basis and weekly basis time series were considered.

In this research, the best model, according to the statistical values AIC, BIC, RMSE, and Standard Deviation, was chosen. Afterward, predictions were made to forecast future values using the best model. Predictions were visually illustrated.

The present study aims to make sales predictions using generic real data supplied by a fast-food society. Also, it was discussed which of the regression models, such as Autoregressive, Vector Autoregressive, or Holt Winter’s Exponential Smoothing, are the best for the determinations. In this study, the price features of the data were the inputs, and recent prices were used to predict future ones.

Many more models were constructed to have more accurate results for the time series in recent years. There are measurement criteria, which are the best model for the time series test data. One of the measurement criteria is the Akeike Information Criteria, and after that, the Bayesian Information Criteria is proposed; also, we can measure the accuracy with Root Mean Square Errors. The best model was chosen according to the RMSE, AIC, and BIC values. There are nine methods, AR, MA, ARMA, ARIMA, SARIMAX, VAR, VARMA, SES, and HWES, implemented in this thesis for food sales prediction. The prediction performance of several of these methods were compared.

2. LITERATURE REVIEW

Prediction techniques have been continuously developed in the last few decades. These techniques vary from regression models to neural networks or support vector machines. Some machine-learning algorithms and regression models have been proposed in forecasting problems. Forecasting is also widely used in the food sales sector.

Although Neural Network experiments have not been carried out in this research, the work of Kong and Martin (1995) can be discussed in this study. They applied a backpropagation neural network (BPNN) to forecast future sales volumes of a food product for a large Victorian food wholesaler. They suggest that wrong parameter selection in the BPNN model leads to slow convergence and/or wrong output.

To deal with the uncertainty in seasonality, Chang (1997) showed a fuzzy forecasting technique in food sales. He played on both seasonal and trend fuzziness in his study. Doganis et al. (2006) presented a nonlinear time series sales forecasting model for combining neural networks and a generic algorithm for sales of a big manufacturing company's milk sales department. Their work provided fewer errors and proved that adaptive neural networks are more accurate than other time series models. Taylor (2007) developed an exponentially weighted quantile regression method that generates interval forecasts from quantile predictions. It was put forward that his method gave better results than traditional methods. In the last decades, some models have been proposed, which was stated by Arunraj and Ahrens (2014, p. 321) as follows: Chen and Qu (2009) managed to make a model that works for perishable foods, which reduces the prediction error more than other statistical methods.

ANN has an important place in sales predictions. Hasin et al. (2011) proposed fuzzy ANN in forecasting the sales of selected merchandise in a retail chain in Bangladesh. In their study, they proved that the forecasting values of fuzzy ANN are better than Holt Winters Exponential Smoothing (HWES) model in terms of the mean absolute percentage error.

According to Adhikari et al. (2013), the following two works are important in food sector sales prediction. Lee et al. (2012) used the sales prediction of the Backpropagation Neural

Network (BPNN) model, which compared these results in logistic regression and moving average. Also, Shukla and Jharkharia (2013) used ARIMA in forecasting the wholesale of vegetables in Indian grocery stores.

2.1 TIME SERIES MODELS IN PRICE PREDICTION

Prediction is always going to be a major area for the researchers who want to improve the accuracy of the existing models. The prediction area is important for institutions and companies. It is valuable to have future predictions based on a model that would lead them to plan and develop effective strategies.

In this literature review, the most popular prediction models were examined according to Zhang, and the time series prediction models are discussed, especially the price prediction models (Zhang 2003). It is purported to trace the historical overview of the models. Some models were not used in this research, since linear methods were used and the classical approach and ARIMA models have been more popular than even Artificial Neural Networks in short-term predictions until today.

The variants of the ARIMA model were analyzed in this literature review since they are the leading models in traditional regression analysis in time series. Exponential smoothing models (SES, HWES) were a part of this study and were widely used. Vector autoregression models (VAR, VARMA) were multivariate models that were observed in the literature review. Likewise, some brief information about the history of SVM, generalized autoregressive conditional heteroskedasticity (GARCH), and ANN were presented in this section.

Stochasticity in the time-series was first introduced by Yule in 1927. In his work, he described that time series could be regarded as the realization of a stochastic process. Before this study, in the 19th century, the time series were defined according to the deterministic world. Planted on the idea of Yule researchers, such as Slutsky, Walker, Yaglom including him, developed AR and MA models (Gooijer and Hyndman 2006).

A decade after Yule's introduction to stochasticity, Wold published his theorem, called Wold's theorem: "Every weakly stationary, purely nondeterministic, stochastic process can be written as a linear filter (linear combination) of a sequence of uncorrelated random variables" (Wold, 1938).

Wold's theorem leads to the solution and formulas of Kolmogorov's (1941) linear forecasting problem's solution. After that is solved, there are papers and researches about parameter estimation, model checking, forecasting. One of the researchers who made these works is Newbold (Newbold 1983).

Forecasting and control as a book was a breakthrough that combines the existing literature written by Box and Jenkins in 1970. Also, the Box–Jenkins approach was developed by these authors and was used for identification, estimation, and verification of time series in a three-stage iterative cycle. The book had a huge impact on the modern approach to time series analysis, which was led to be used by the developing of computer science, used in many branches of science. ARIMA models and their variations introduced here became very popular in a large number of science areas (Zhang 2003).

The VAR and VARMA models are used for multivariate regression in the inquiry. These models allow data to be used more effectively, rather than using one field. For the VARMA model, a literature explanation was made by Athanasopoulos and Vahid in the 2008 article "VARMA models depend on Wold decomposition theorem" (Wold 1938). These are the models that depend on finite order, stationary time series.

George Athanasopoulos and Farshid Vahid continued to conduct some important studies in the area of the VARMA model (Quenouille 1957; Hannan 1969; Tunnicliffe-Wilson 1973; Hillmer and Tiao 1979; Tiao and Box 1981; Tiao and Tsay 1989; Tsay 1991; Poskitt 1992; Liitkepohl 1993; Liitkepohl and Poskitt 1996; Reinsel 1997; Thea 2001). Athanasopoulos and Vahid continued their studies in the area of the VARMA model with their work VARMA versus VAR article 2008.

Different from the VARMA model, the VAR model was developed. In the VAR model, the figuring out period is covered as follows: "After the publishment of Christopher Sims (Sims 1980), the finite-order VAR model has become the most important modeling of macro-econometric literature" (Athanasopoulos and Vahid 2008). Other model types used were the Exponential Smoothing models that are SES and HWES.

There are studies about Exponential Smoothing of Brown, Holt and Winters that this subject is firstly proposed: “Exponential smoothing (Brown, 1959; Holt, 1957; Winters, 1960), become the most powerful models. Predictions are weighted averages of observed value as past values that have weights decaying exponentially as time pass” (Hyndman and Athanasopoulos 2018, p. 183).

3. DATA AND THE METHOD

In this thesis, various time series analysis algorithms were performed, along with the data set on fast-food sales predictions provided by a software technology firm, which is a part of a holding. The models of regression were examined on this data at certain time intervals. These time intervals were arranged hourly, daily, and weekly. Two years of a total dataset was handled. The predictions on this dataset helped to determine which model best fits the dataset.

Equally, it was explicated in the literature review that there exist dozens of examples that explain the data of time series. Forecasting has been a huge area of research for many years. Its goal is to find which traditional models best fit the data and to achieve a high accuracy of sales prediction for the time to come.

3.1 DATASET

The data were obtained from a fast-food firm that is working with a tech firm. The dataset and its details were elucidated in this section.

The data that were used in this study were covering about two years of sales transactions of sales at the fast-food firm. The dataset consists of 8 columns. In the first column, an identity number is assigned to the label FKItemId. This shows which item is sold. Detailed time fields of when the transaction took place were provided. Moreover, which menu's ordered given in a categorical form. Some other field that is offered is the prices of the material sold. The other field is the number of menus sold in the same transaction. Finally, the DiscPrice, which is the final value of the corresponding sale transactions total sale sum is given.

The dataset explicated in detail in the following judgment of convictions. There is no nominal field in them. No missing data were present in the transactions. What it was done about the time fields was combining them and to form a date value in a newly formed field for that transaction. After that, the hourly, daily, and weekly sums of these individual transactions of sales were calculated and got the DiscPrice field the sum for that time intervals. However, foremost, the dataset fields are presented in detail in Table 3.1 below, with the first ten transactions, which formed a total of 523,546 lines.

Table 3.1: First ten tuples of the dataset

FKItemId	DateOfBusiness	FKOrderModeId	Hour	Minute	Price (TL)	Quantity	DiscPrice(TL)
10001	9.01.2015	4	0	17	186.0975	1	186.0975
40015	9.01.2015	2	0	37	93.0488	2	186.0975
20007	9.01.2015	4	0	11	106.8338	1	106.8338
30005	9.01.2015	4	0	9	72.3713	2	144.7425
50006	9.01.2015	4	0	9	71.6820	2	143.3640
10018	9.01.2015	2	0	42	268.8075	1	268.8075
10018	9.01.2015	2	0	43	268.8075	1	268.8075
10018	9.01.2015	4	0	17	227.4525	1	227.4525
10018	9.01.2015	4	0	43	227.4525	1	227.4525

If it has closely looked at the dataset, the date fields can be seen. For example, all ten of these transactions occurred in the first hour on 09.01.2015; when we combine them hourly, we will get the sum of all the revenue by adding the DiscPrice's field. In these experiments, a time-series approach in the traditional model variations of ARIMA, Exponential smoothing was used as a univariate manner. This means that we tried to form a model from the DiscPrice field and time field by combining the Date, Hour, and Minute fields. The code *see Appendix A.1 Code-1* presents how the revenue for the desired time interval was combined. On the other hand, multivariate methods, such as Vector autoregression and VARMA, were also used, which are also classical models. The models were used in such a way that the VAR model gave better results, as can be seen in this study. The fields that were used can be seen in the following code in detail “see. Appendix A.1 Code-2”:

As can be seen in this code, Price, Quantity, and DiscPrice were used in a multivariate model in time-series.

3.2 AUTOREGRESSION - AR

The first model that will be considered is the autoregressive model. It has a formula like the following (Hyndman and Athanasopoulos 2018):

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{3.1}$$

This is the order p autoregressive model. The model is called *regressive* since the output depends on linearly on previous values with coefficients in a stochastic term. If the formula is closely investigated, an AR(p) model is represented here, p is the order of the model, which shows how many lags will be used to predict the output. c is constant, and ε_t is white noise, which means it is not very important as compared to the first lagged data. By changing the parameters $\phi 1$, ϕp results in different time-series patterns. “The variance of the error term ε_t will only change the scale of the series, not the patterns” (Hyndman and Athanasopoulos 2018). For the ϕ parameters of the model, the following values are reached:

For an AR(1) model: $-1<\phi 1<1$ (3.2a)

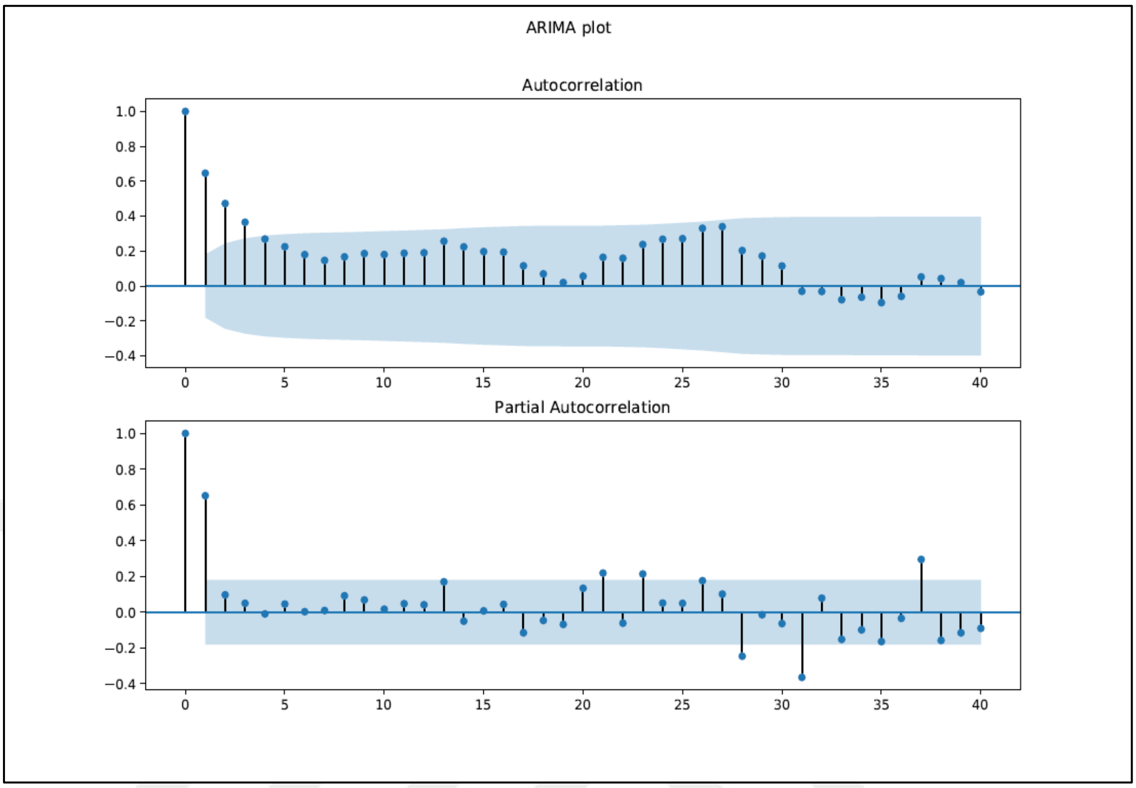
For an AR(2) model: $-1 < \phi 2 < 1, \phi 1 + \phi 2 < 1, \phi 2 - \phi 1 < 1$ (3.2b)

(Hyndman and Athanasopoulos 2018)

For a higher order of p, this value is adjusted by the library of the programming language, which is much more complex.

The fast-food data have a price field of transactions in time. Then, when the autocorrelation and partial autocorrelation graph were obtained by these data, the following consequences will be:

Figure 3.1: Autocorrelation and Partial Autocorrelation Graphs



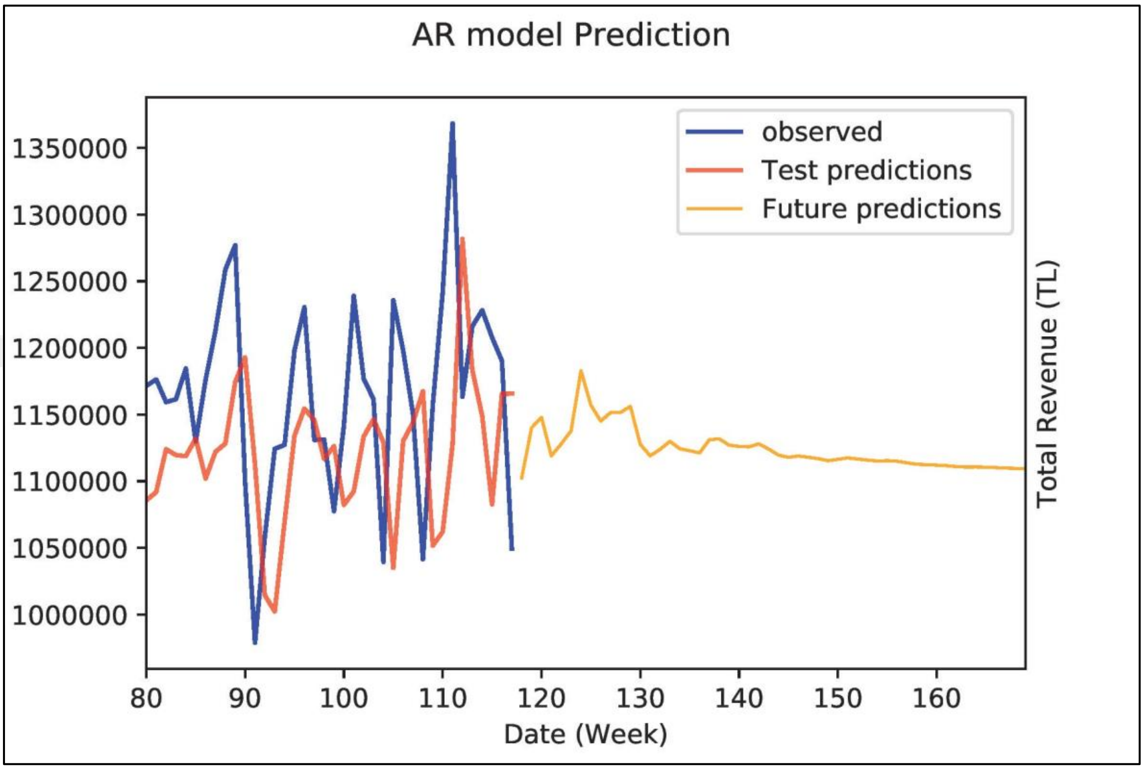
Equally, it can be seen from Figure 3.1 that, in the autocorrelation part, the lags were diminishing in order and the more important one was the partial autocorrelation, where the first two lags were above the white noise, which is represented by a blue interval area. This suggests that the first two coefficients and past values were significant, which was an AR (2) model.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \quad (3.3)$$

In python, there is a library called statsmodels, which automatically chooses the best fitting lag value by making statistical tests and preparing a linear regression model. This model was used and fit the data for making a prediction. It can be explained in the code of implementation “see. Appendix A.1 Code-3”, where we used a procedure in which the prediction was made.

DiscPrice was obtained and added to each transaction and the calculated prices for that instant time. In other words, hourly, daily, and weekly sales were calculated for the fast-food dataset. It can be seen what the AR model's sales predictions were getting when the program was run as follows:

Figure 3.2: Weekly Sales Prediction Graph with the AR model



The graph of Date–Price is shown for the AR model weekly basis in Figure 3.2. In this study, 30% of the dataset was used as a test set and the rest was used for training the model. As can be seen in Figure 3.2, the test date started from the 80th week and continued until the 120th week. After training with 70% of the data, which corresponded to 80 weeks, the data were tested with the model. The red curves show the test results, whereas the blue ones were actual observed values. Then, a new model was trained and got future predictions that are depicted by yellow lines. The AR model was trained in daily and hourly time intervals. The answers are presented in Figures 3.3 and 3.4, respectively.

Figure 3.3: Daily Sales Prediction Graph with the AR model

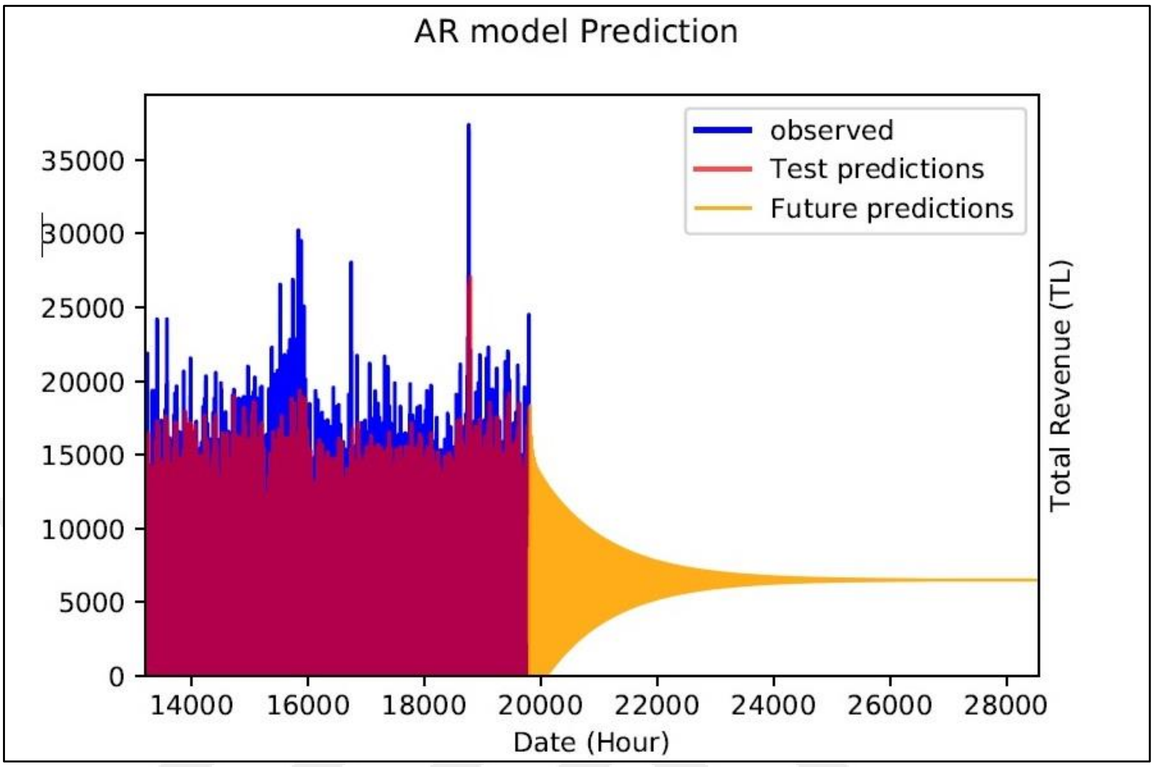
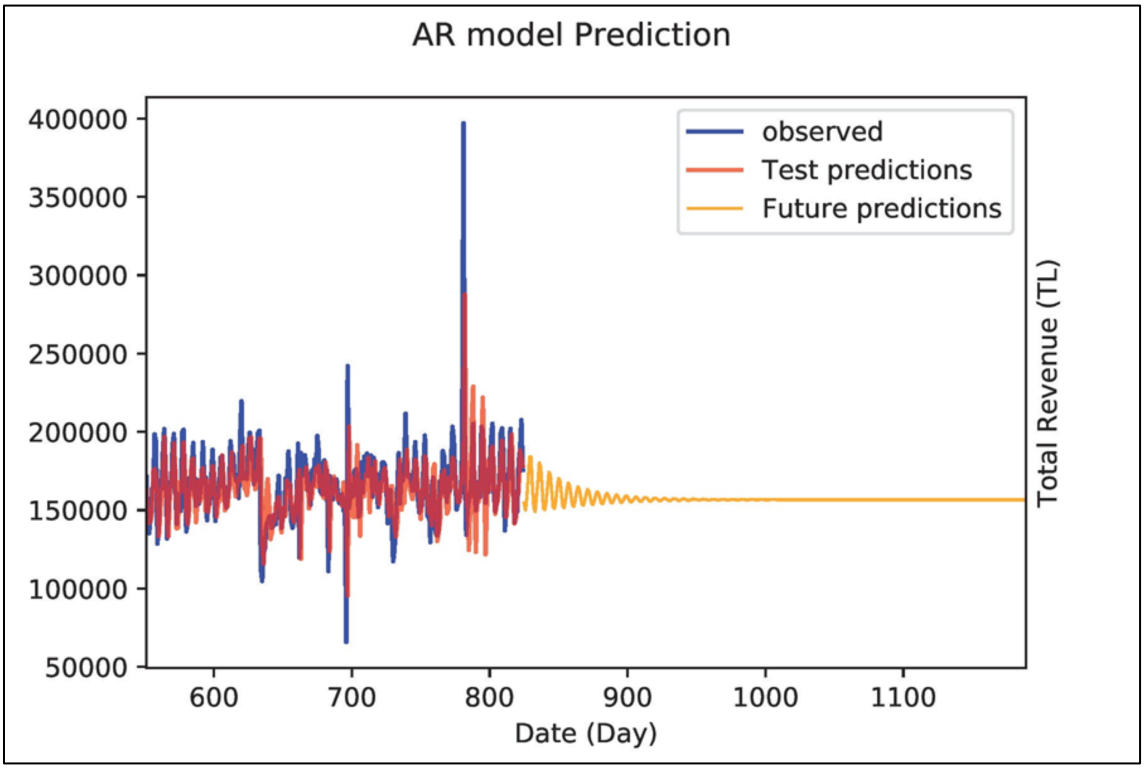


Figure 3.4: Hourly Sales Prediction Graph with the AR model



3.3 MOVING AVERAGE - MA

In the autoregressive model, a linear combination of past values of the desired variable was used, and predictions were made. In other words, past values of a forecast variable in the regression were used. However, in a moving average model, past forecast errors in a model were used that were like a regressive one.

The following formula can help one to understand the moving average models. It was taken from Hyndman and Athanasopoulos (2018):

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{3.4}$$

In the formula, ε_t is white noise, and the model is called MA (q), where q is the order of the MA model. It does not observe ε_t values, as it is not a usual regression model. The y_t values can be predicted, as it can be thought of as a weighted moving average of past errors multiplied with constants. The MA model must not be confused with moving average smoothing, since the smoothing is used to estimate the trend-cycle of past values, while the moving average is used for forecasting future values.

Modifying the parameters θ1,…, θq results in different time-series patterns. As it was the same for the AR models, the variance of the error term ε_t does not change the patterns, only the scales (Hyndman and Athanasopoulos 2018, p.231).

An MA (1) first-order MA model is a linear combination of the first lag of forecasts.

Here, it was thought that $-1 < \theta < 1$, so that the past forecasts had less weight than the present. Hence, the process was invertible, and the effects of past values of the series decreased with time. However, if $|\theta| \geq 1$, it isn’t the desired situation in which the effect of past observations increases with the time. The following θ values are worthy of the MA (1) and MA (2) models. The orders above these are calculated by the programming language libraries.

The invertibility constraints for other models are similar to stationary constraints.

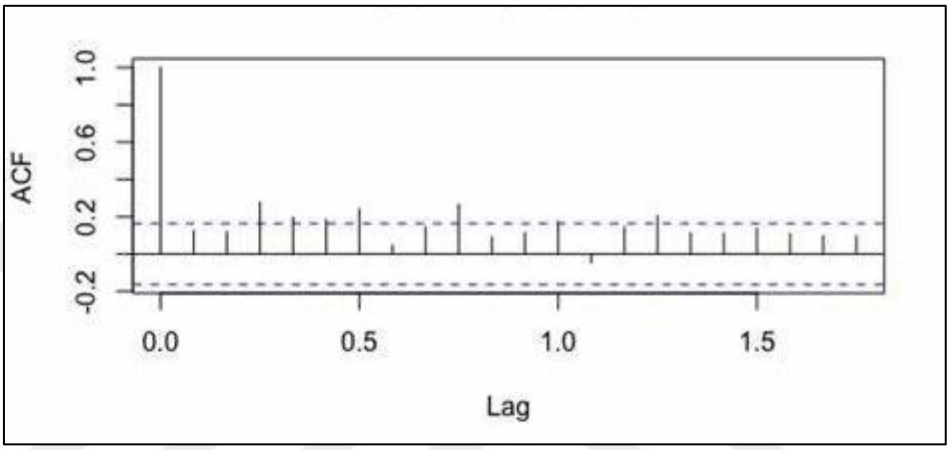
For an MA (1) model: $-1 < \theta 1 < 1.$ (3.5a)

For an MA (2) model: $-1 < \theta 2 < 1,\ \theta 2 + \theta 1 > -1,\ \theta 1 - \theta 2 < 1$ (3.5b)

(Hyndman and Athanasopoulos 2018)

The Autocorrelation Function (ACF) graph and lags of the MA (1) model can be expressed in the graph as follows:

Figure 3.5: Example ACF graph

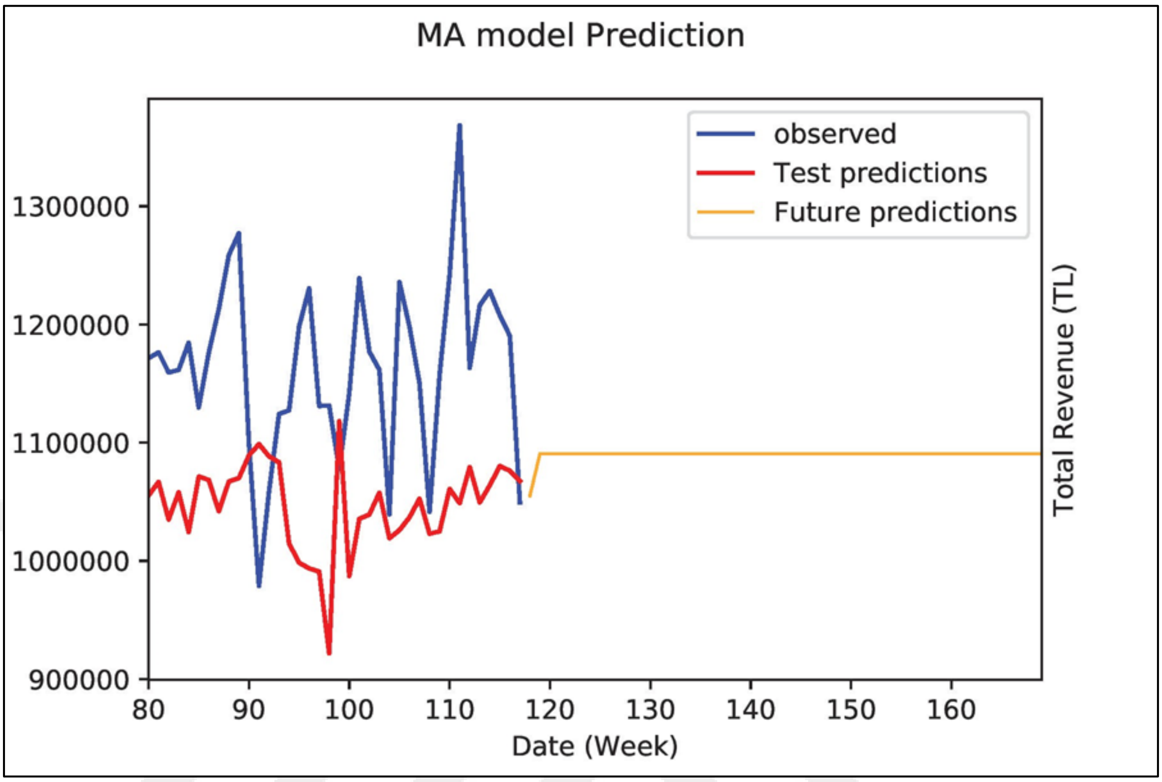


It can be observed that, in the data ACF and PACF diagrams, the situation was different. In MA (1), only one lags ACF value was above white noise, merely in the graphs, as it can be remembered (Figure 3.1). The desired model was not an MA function, as there were many more lags above the white noise in the ACF graph. However, the MA (1) model was examined for the future tests to ascertain which model was the best in sales prediction according to the fast-food dataset.

For prediction purposes in the MA (1) model, a code was used. Equally, it can be seen in pseudo-code “see. Appendix A.1 Code-4” that the ARMA function was used, where p part was 0 and q part was 1, which means the MA (q=1) model and perform the prediction.

It can be seen that the MA model’s sales predictions were getting when the program was run, as follows.

Figure 3.6: Weekly Sales Prediction Graph with MA model



Equally, it can be understood in the graph; this model didn't give a full performance in forecasting. This was stated in explaining the ACF and PACF diagrams for the dataset. The data best fit for the AR model according to these diagrams. However, for calculating the difference and research purposes MA, models results were also received.

In the experiments and results, the result of RMSE, AIC, and BIC were given, according to each model. The daily and hourly prediction graphs for the MA model are as follows.

Figure 3.7: Daily Sales Prediction Graph with MA model

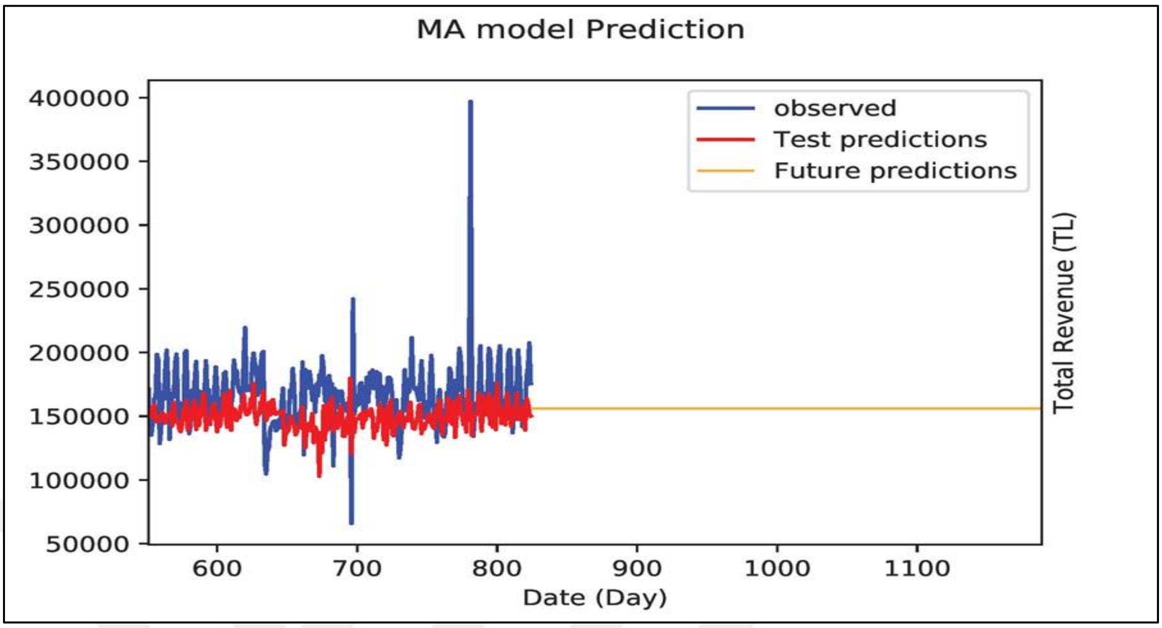
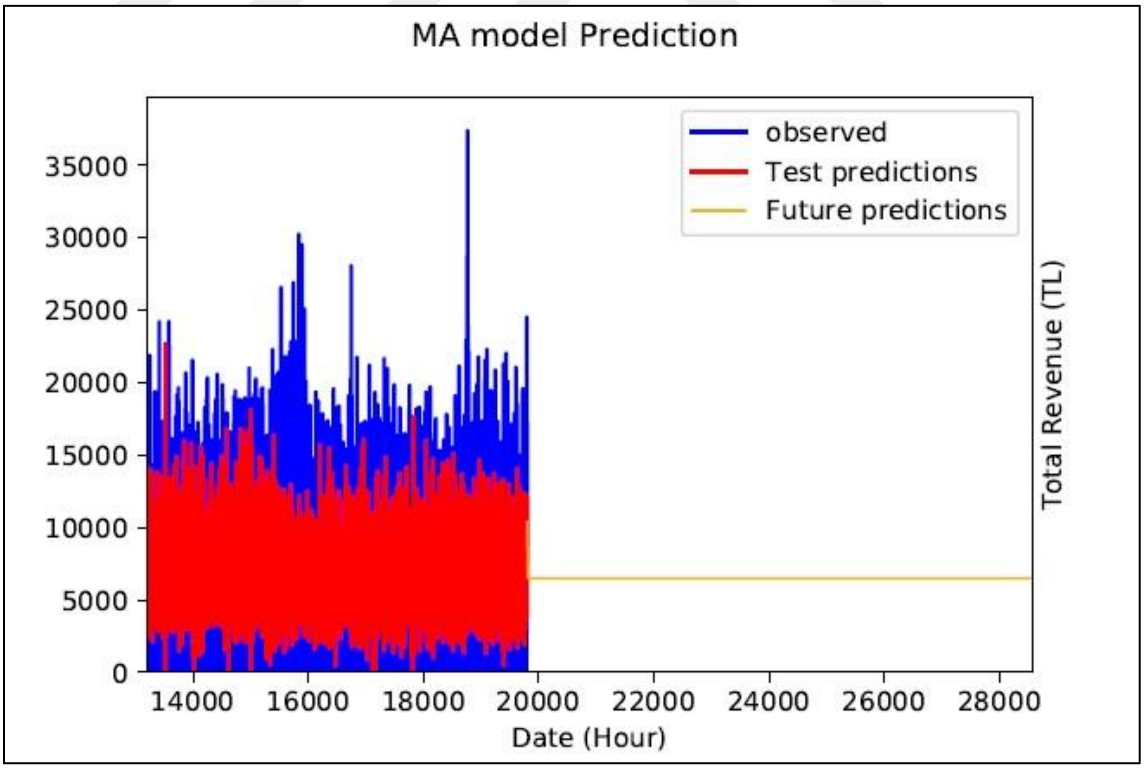


Figure 3.8: Hourly Sales Prediction Graph with MA model



3.4 AUTOREGRESSIVE MOVING AVERAGE – ARMA

When AR (p) and MA (q) models are combined into a model called autoregressive (AR), Moving Average (MA) is obtained, which is the ARMA (p, q) model for the time-series. When it is recalled that AR(p) has the formula as "Equation 3.1":

While MA (q) throws the formula as "Equation 3.4",

it was obtained by summing these two; the ARMA (p, q) model

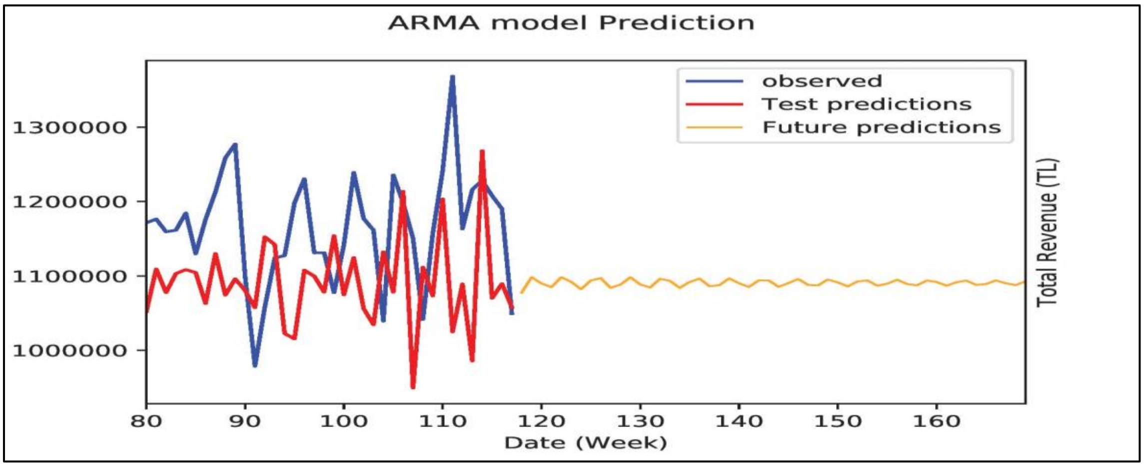
$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (3.6)$$

In time, t the MA part of the model is used as a moving average of q terms over q past values. On the other hand, in the AR part, p terms of recent weighted values are expressing the y term (J de Smith 2018).

A code was used for the prediction purposes in the ARMA model "see. Appendix A.1 Code-5". Here, the orders were selected according to a search for the best AIC values, and that fit the sales price time graphs.

It can be seen that the ARMA model's sales predictions were getting when the program was run, as follows:

Figure 3.9: Weekly Sales Prediction Graph with ARMA model



The AIC value for the model is given in the results section. According to this graph, the model was not the best one since it didn't fit the test results. Daily and hourly graphs are as follows:

Figure 3.10:Daily Sales Prediction Graph with ARMA model

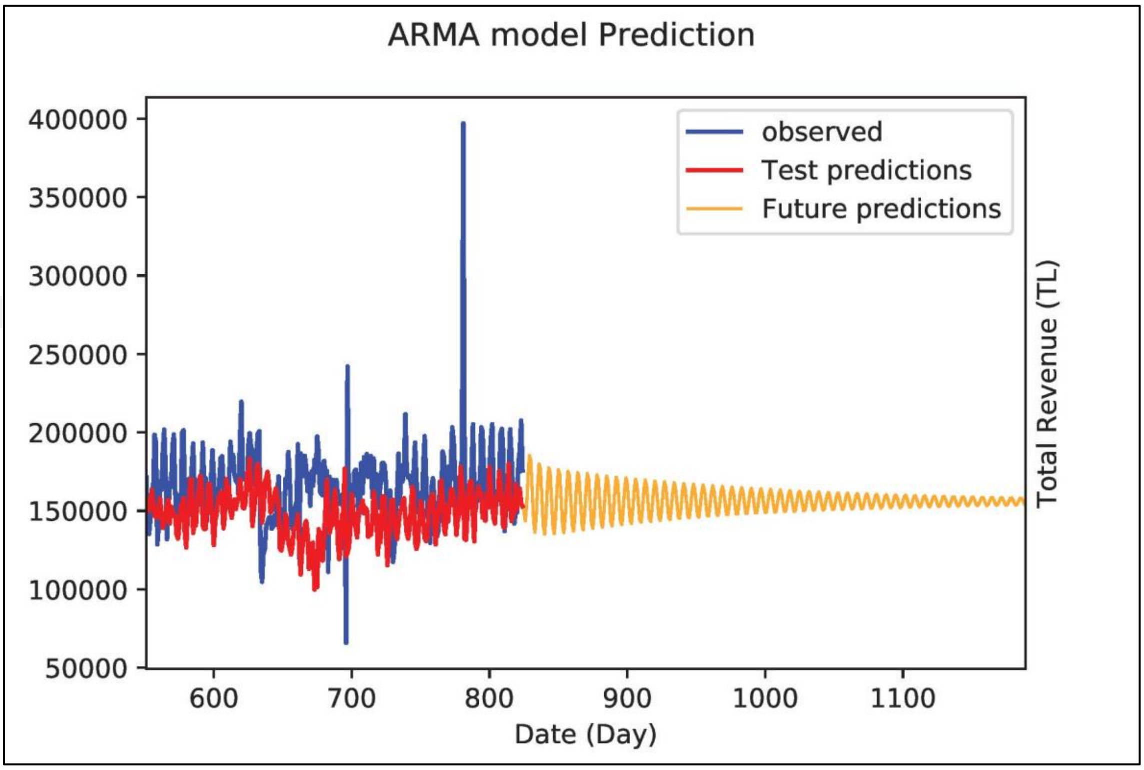
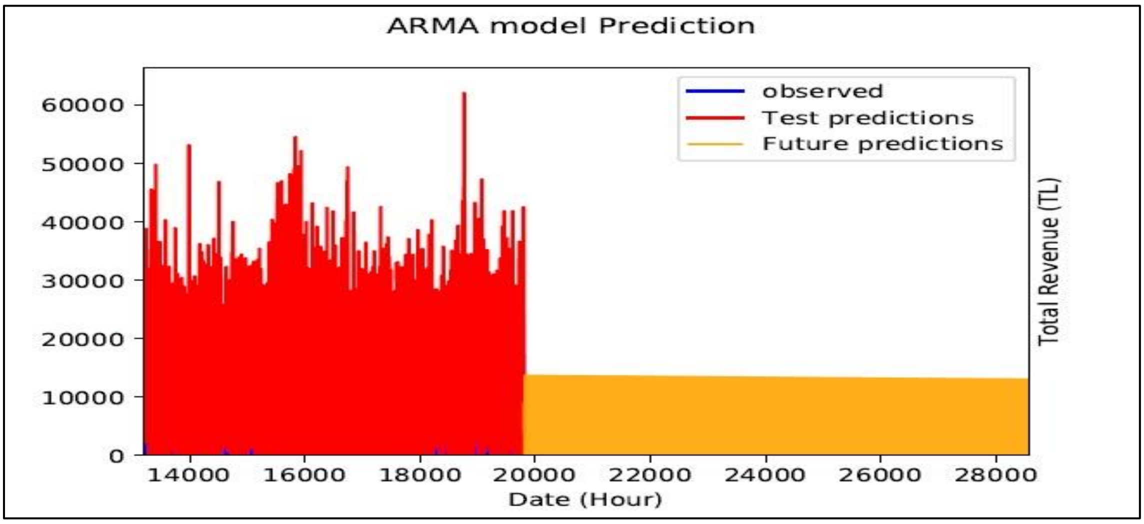


Figure 3.11: Hourly Sales Prediction Graph with ARMA model



3.5 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE - ARIMA

The term ARIMA comes from Autoregressive (AR) Integrated (I) Moving Average (MA). Thus, ARIMA models can be viewed as a combination of AR and MA models. The integrated part is the difference. The primary difference of ARIMA models from ARMA models is that ARIMA models involve differencing the original time-series data applied for prediction (Makridakis and Hibon 1998, p.148).

The notation for ARMA is ARMA (p,q), whereas the notation for ARIMA is ARIMA (p,d,q). The *p* stands for the number of autoregressive terms, *d* is the number of differencing needed, and *q* is the number of lagged forecast errors in the prediction equation. The only case in which an ARIMA model can be expressed as an ARMA model is when there is no differencing needed to make the time series stationary.

Differencing the time-series means forming a new time series by subtracting observation 1 from time 2, observation 2 from observation 3, and so on. The point of this is to remove certain trends, such as seasonality, downward/upward trends, or inconsistent variance in time series data.

The formula for ARIMA models forecasting is as follows:

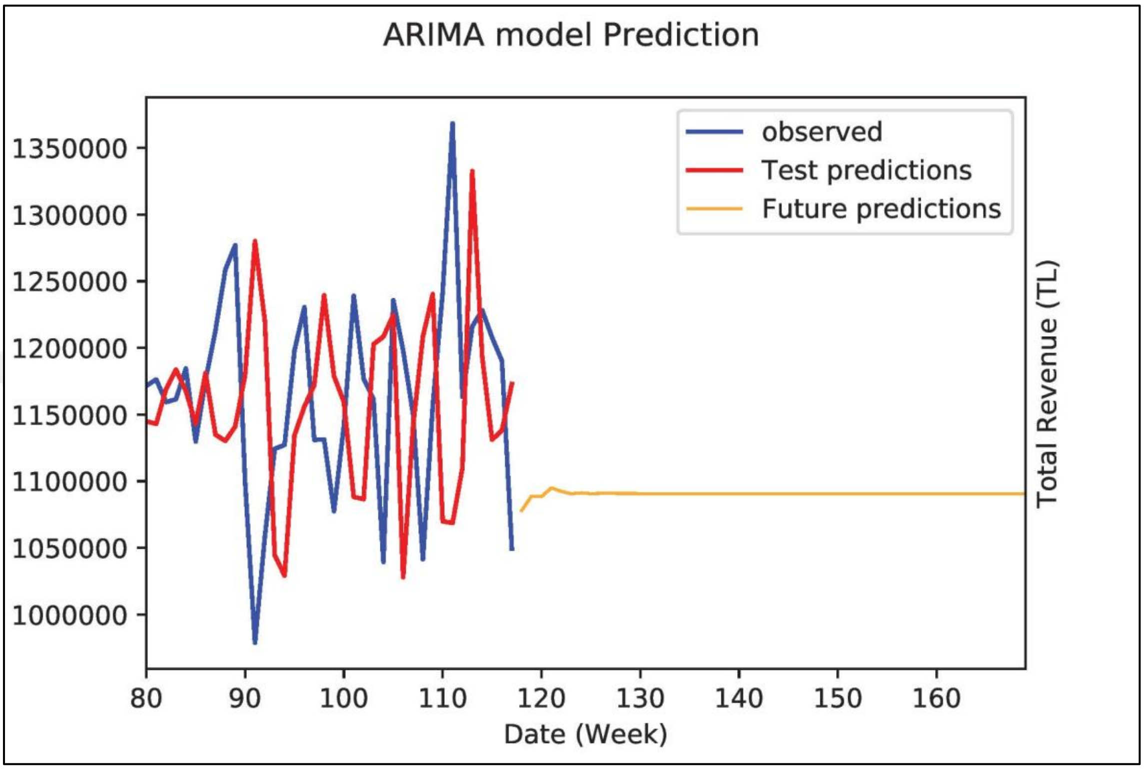
$$y'_t = c + \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon t \tag{3.7}$$

In the formula of Equation 3.7, y′t is the differenced series (it may have been differenced more than once). The predictors on the right-hand side include both lagged values of yt and lagged errors. This is called an ARIMA (p,d,q) model" (Hyndman and Athanasopoulos 2018).

A code was utilized for the prediction purposes in the ARMA model "See Appendix A.1 Code-6". The orders were selected according to a search for the best AIC values and that fit the sales price time graphs.

With respect to the weekly, daily, and hourly sales prediction graphs, the following can be noted.

Figure 3.12:Weekly Sales Prediction Graph with ARIMA model



These graphs present the start of the test set and future predictions, respectively. The daily graph and hourly were also like the following:

Figure 3.13:Daily Sales Prediction Graph with ARIMA model

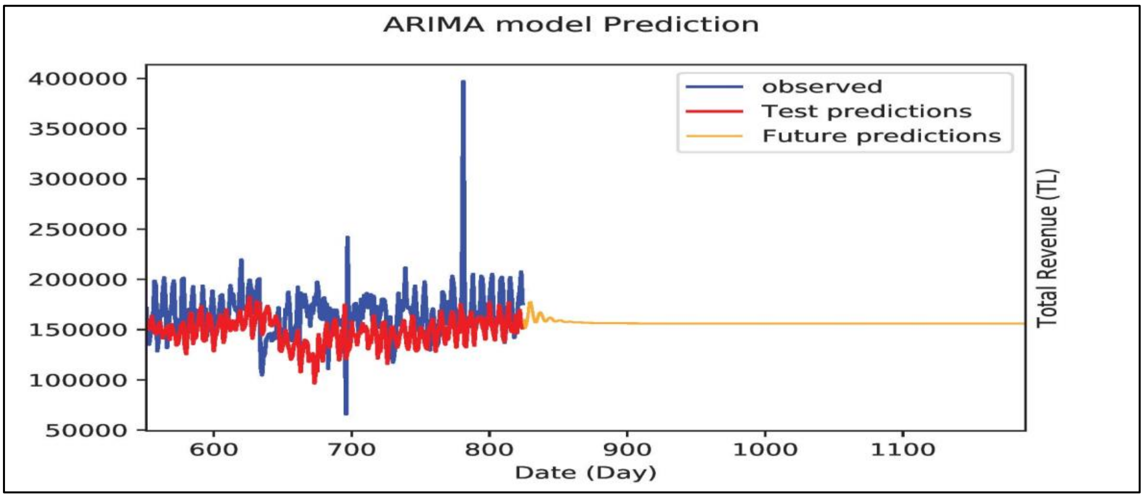
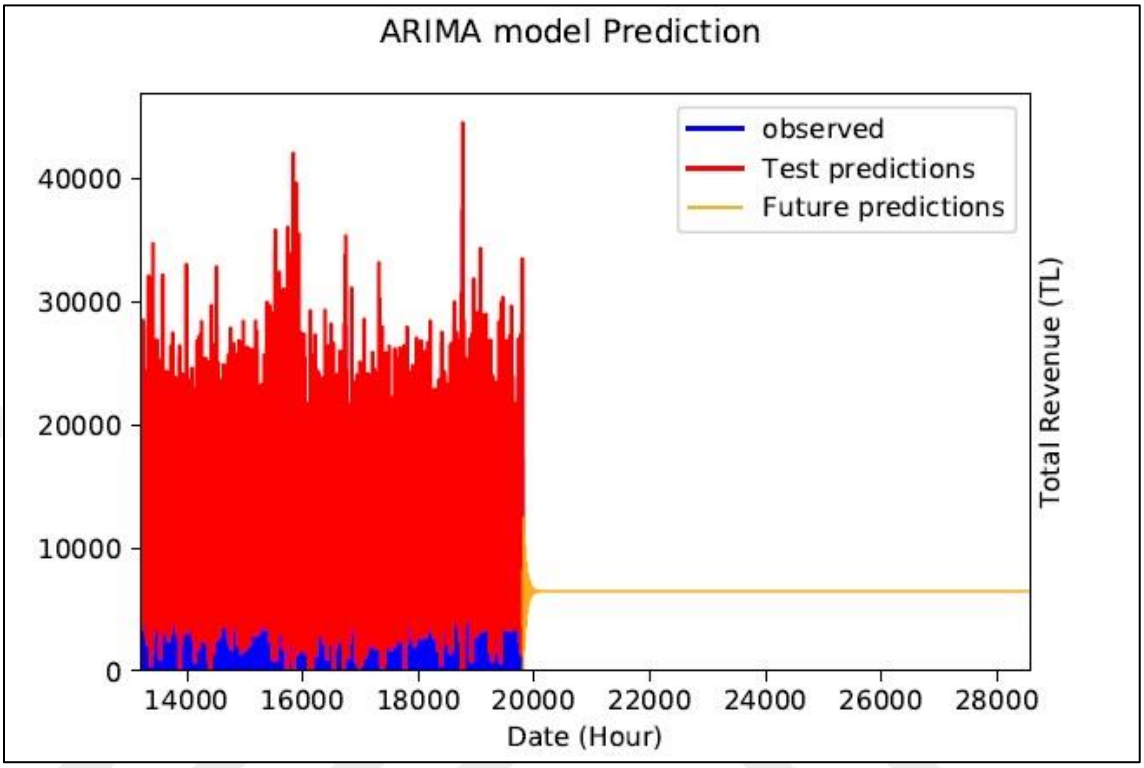


Figure 3.14:Hourly Sales Prediction Graph with ARIMA model



3.6 SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE - SARIMAX

The time-series models that have been observed so far focus on nonseasonal data. Another ARIMA model is presented that is capable of passing results for seasonal time-series data. The formula is for that is as follows:

$$ARIMA\ (p, d, q)\ \ (P, D, Q)m \tag{3.8}$$

"m is the observations per year in Seasonal-ARIMA or SARIMAX model, where small letters (p,d,q) are non-seasonal, [and] capital letters (P,D,Q) are the seasonal part in the Equation 3.8" (Hyndman and Athanasopoulos 2018, p.253).

The SARIMA parts are formed of back shifts of the seasonal parts of the model. ARIMA $(1,1,1)(1,1,1)_4$ model that is for quarterly data (m=4) and may be shown as

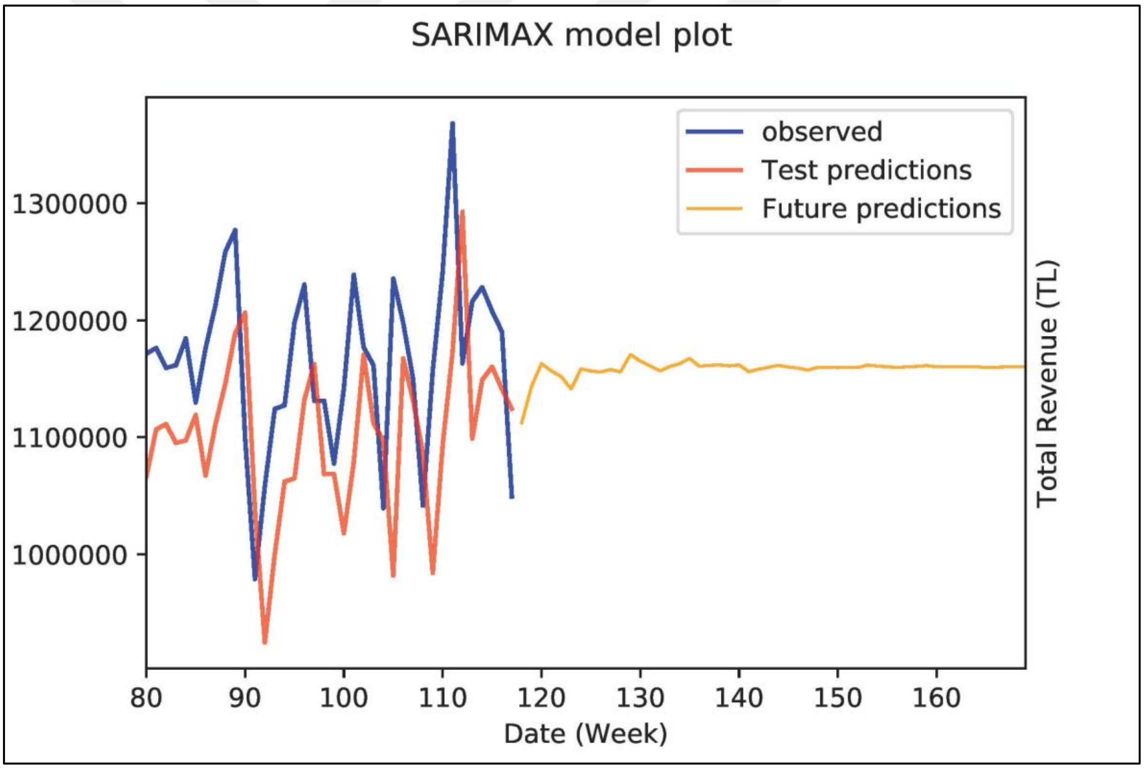
$$(1-\phi_1 B)\,(1-\Phi_1 B^4)(1-B)(1-B^4)y_t = (1+\theta_1 B)\,(1+\Theta_1 B^4)\varepsilon_t \qquad (3.9)$$

The results are achieved by multiplying seasonal terms with nonseasonal terms.

The modeling structure was the same as that of ARIMA, but the m value must be given and used as a seasonal model. The best fit was used by looking at the AIC values. The parameters were chosen as order=(1, 1, 1) and seasonal order=(1,0,1, 12), which can be seen in the code “see. Appendix A.1 Code-7”.

When the weekly, daily, and hourly graphs for the SARIMAX models were plotted, the following figures were obtained, respectively:

Figure 3.15:Weekly Sales Prediction Graph with SARIMAX model



It can be understood that this test-data better fit in the SARIMAX model than the MA, ARMA, and ARIMA models. The same success could be ascertained from the daily graph, too. However, the future predictions were on the same level, which had a little difference in magnitude.

Figure 3.16:Daily Sales Prediction Graph with SARIMAX model

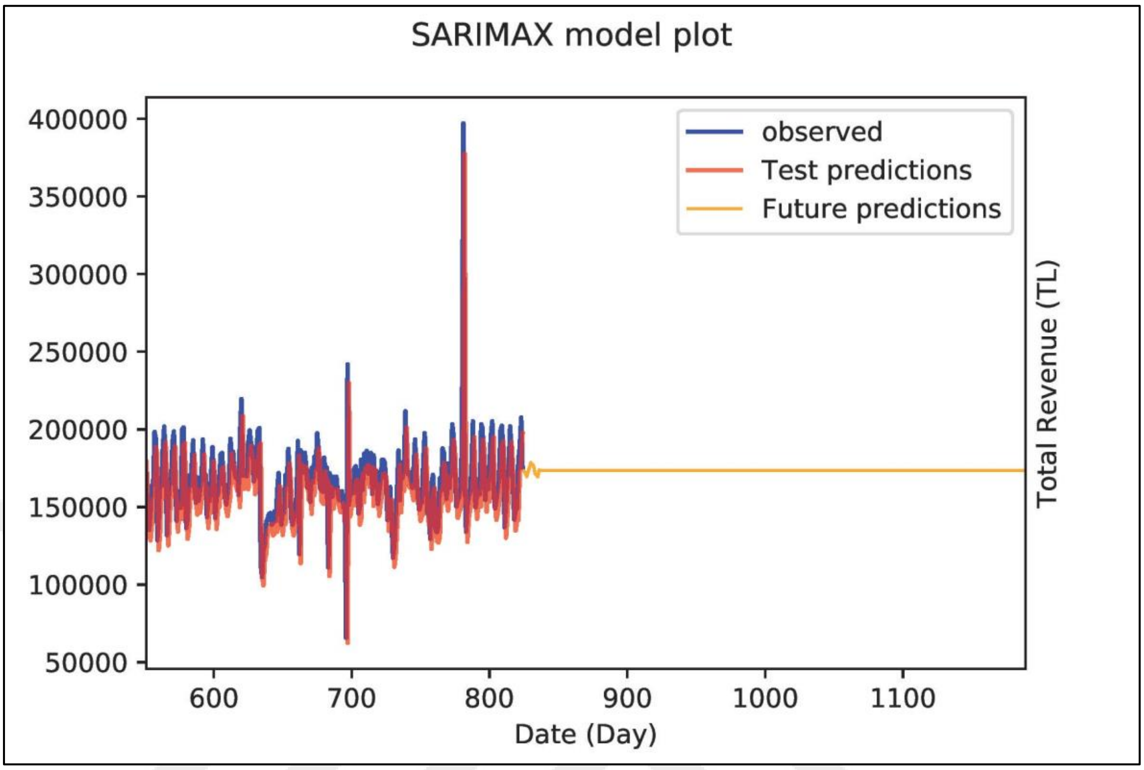
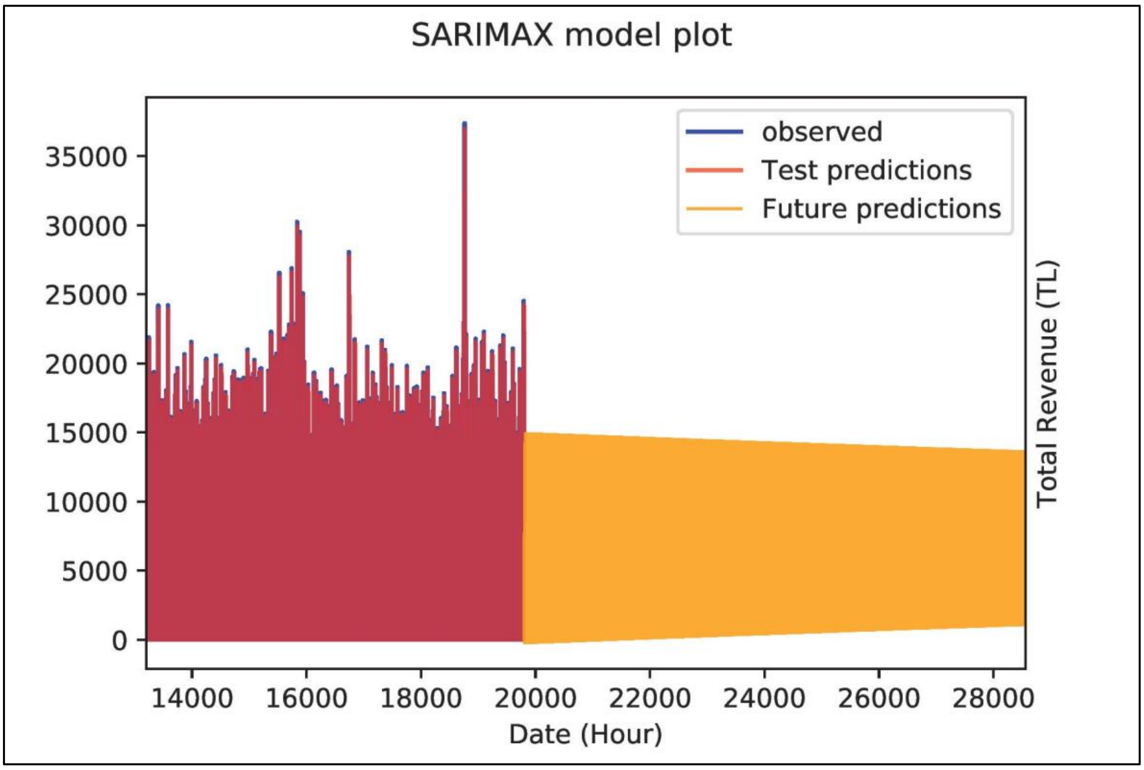


Figure 3.17:Hourly Sales Prediction Graph with SARIMAX model



3.7 VECTOR AUTOREGRESSION - VAR

The prediction models that have been reported so far are working on a univariate basis. When it is desired to predict, according to more variables in the data, a multivariate model must be applied. The Vector Autoregression models provide better results in prediction, as they use multiple variables from the dataset. The data are stationary, which means the data have no trends, and there is more than one variable that may affect the results.

Before giving the graphs for the model and pseudo-codes, some formulas of the model, in general, were slightly examined. The model was easy to use and result oriented with the multivariate property of it. Here, a couple of formulas were given again, as was done in previous models. Each variable had its equation; in other words, there were formulas for each of the variables separately.

If it is wanted to take a closer look, two variables are given with a constant and lags of time. A VAR (1) with two variables—in other words, two dimensions—was given as an example. To be reliable and clear, the same notation was practiced as it was done in other models where the formulas from (Hyndman and Athanasopoulos 2018, p.330):

$$y_{1,t} = c_1 + \phi_{11,1} y_{1,t-1} + \phi_{12,1} y_{2,t-1} + \varepsilon_{1,t} \tag{3.10}$$

$$y_{2,t} = c_2 + \phi_{21,1} y_{1,t-1} + \phi_{22,1} y_{2,t-1} + \varepsilon_{2,t} \tag{3.11}$$

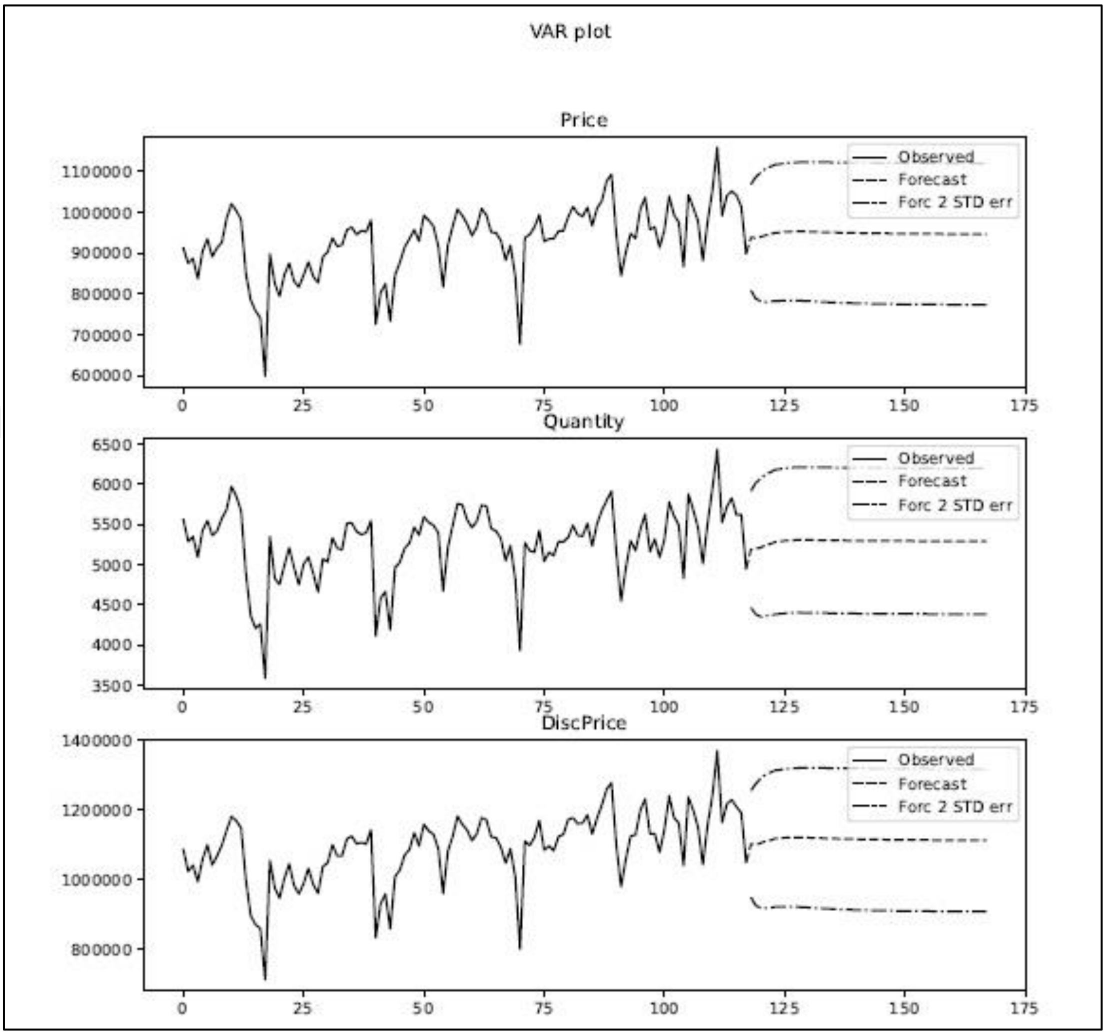
where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are white noise. The forecast was performed for each variable as shown in equations (3.10) and (3.11)

It was passed on the pseudo-code for the VAR model that was used in this research and showing which variables were applied in the model, Price, Quantity,DiscPrice “see. Appendix A.1 Code-8”.

Foremost, the data were organized as shown in seriVAR3; we used *Price*, *Quantity*, and *Disc Price* as three variables in date1 intervals. Date1 was formed as weekly, daily, or hourly intervals. The data are organized (seriVAR3) to seriVAR2 in these time intervals, and the data were used as seriVAR.

The graphs for the VAR model are as follows.

Figure 3.18:Weekly Sales Prediction Graph with VAR model

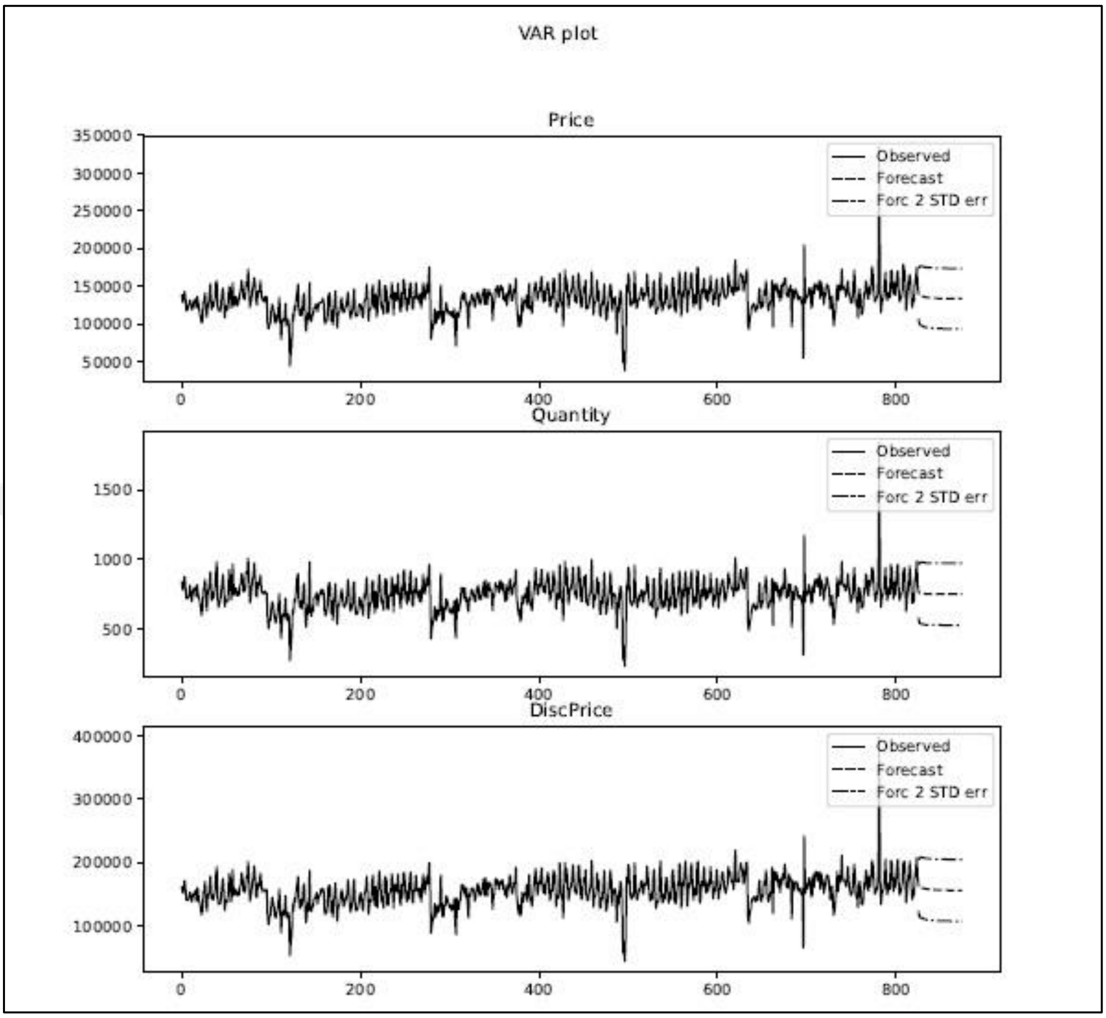


As can be viewed in Figure 3.18, for all fields that were included in the model, the predictions held. The predictions are shown with dash lines, and predictions with dash and dots show the predictions with the standard deviation.

It was discussed in the results section, but it is important to say in here that this model gave better performance, according to the AIC and BIC values, than most of the univariate models, but it was approximately the same level as the AR model.

The daily graphs were also as follows.

Figure 3.19:Daily Sales Prediction Graph with VAR model



3.8 VECTOR AUTOREGRESSION MOVING AVERAGE - VARMA

A finite order p—in other words, VAR (p)—and a finite order q, as an MA (q) combination, form a Vector ARMA or VARMA (p, q) model. This model is also multivariate, as it was in the Vector autoregression model. It has superiority over the ARIMA models that are univariate. However, in this case, the AR model is still the best answer, as it can be investigated ACF and PACF graphs in Figure 3.1

A two field (variable) with one lag of VARMA (1,1) equations can be as follows:

$$y_{1,t} = c_1 + \phi_{11,1} y_{1,t-1} + \phi_{12,1} y_{2,t-1} + \varepsilon_{1,t} + \theta_{11,1}\varepsilon_{1,t-1} + \theta_{12,1}\varepsilon_{2,t-2} \quad (3.12)$$

$$y_{2,t} = c_2 + \phi_{21,1}y_{1,t-1} + \phi_{22,1}y_{2,t-1} + \varepsilon_{2,t} + \theta_{21,1}\varepsilon_{1,t-1} + \theta_{22,1}\varepsilon_{2,t-2} \tag{3.13}$$

where c is the constant, the ϕy parts are a VAR(1) model's past values, ε_t is white noise, and the $\theta\varepsilon$ parts are model MA (1). The two equations are organized as lag 1 VARMA (1,1) model.

If it is desired to predict with more lags, ϕ and θ variables that are calculated by the least-square means are needed to calculate for a further approach.

The model showed better performance than the AR and VAR models concerning RMSE, but not regarding the AIC and BIC criteria.

When the pseudo-code for VARMA was examined, it can be determined that a multivariate data frame was used "see. Appendix A.1 Code-9".

3.9 SIMPLE EXPONENTIAL SMOOTHING - SES

The next time points were observed as an exponentially weighted linear function of previous time points in the simple Exponential Smoothing or short SES model. There was no seasonal or trend in the data to use it directly.

The level of smoothing at the time t is represented with ℓ_t, and the equation formulas are given below.

Equation of prediction:

$$\hat{y}_{t+h|t} = \ell_t \tag{3.14}$$

Equation of smoothing:

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1} \tag{3.15}$$

While setting h=1 means the fitted values will be shown if t=T, then the predictions beyond the training data will be found.

If it is replaced ℓt and ℓt−1 with the value comes from Equation 3.14 then the Equation 3.15 will be found in a form that will show the weighted average form of SES:

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\, \hat{y}_{t|t-1} \tag{3.16}$$

Another equation that explains smoothing at time T+1 can be given as follows:

$$y_{T+1|T} = \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots \tag{3.17}$$

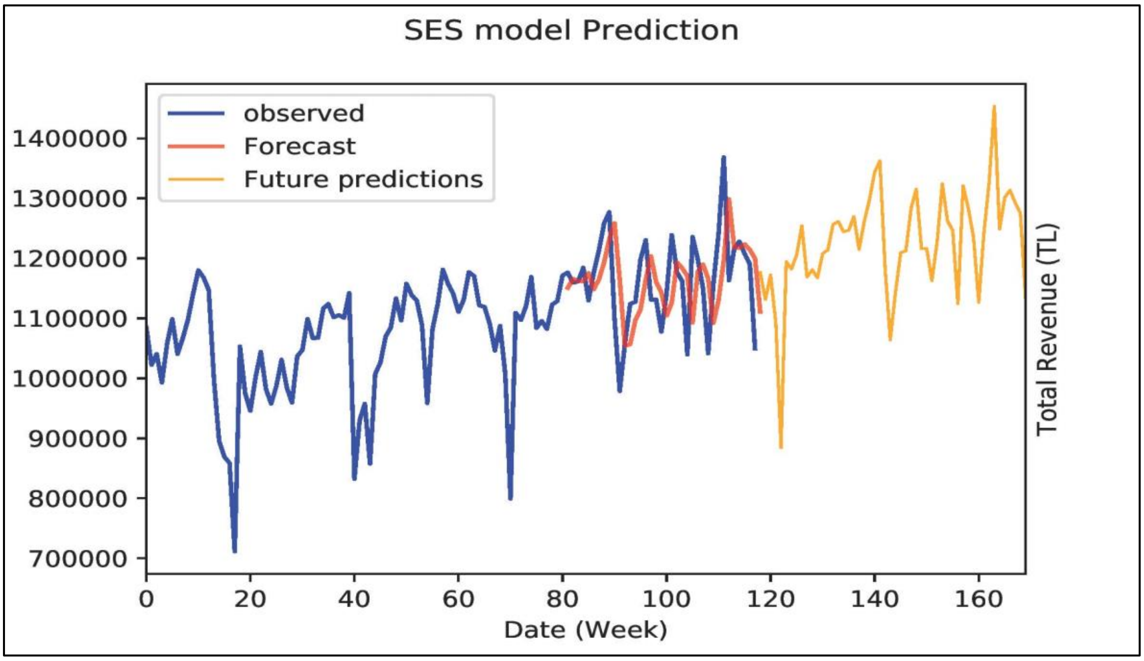
where $0 \leq \alpha \leq 1$ is the smoothing parameter.

This model of SES performance, according to other models that were discussed, is not as good. However, better than VARMA in the name of AIC and BIC and only better than the MA model in the name of RMSE criteria.

The pseudo-code for the model prediction was written “see. Appendix A.1 Code-10”.

Next, graphs for the fast-food sales dataset and price prediction according to three different time intervals were examined. The first interval was weekly, and the other intervals were daily and hourly.

Figure 3.20:Weekly Sales Prediction Graph with SES model



The red lines in Figure 3.20 show that the test data fit well. The yellow lines are forecasting the future, while the bluish ones are the original dataset prices for a weekly basis.

Figure 3.21:Daily Sales Prediction Graph with SES model

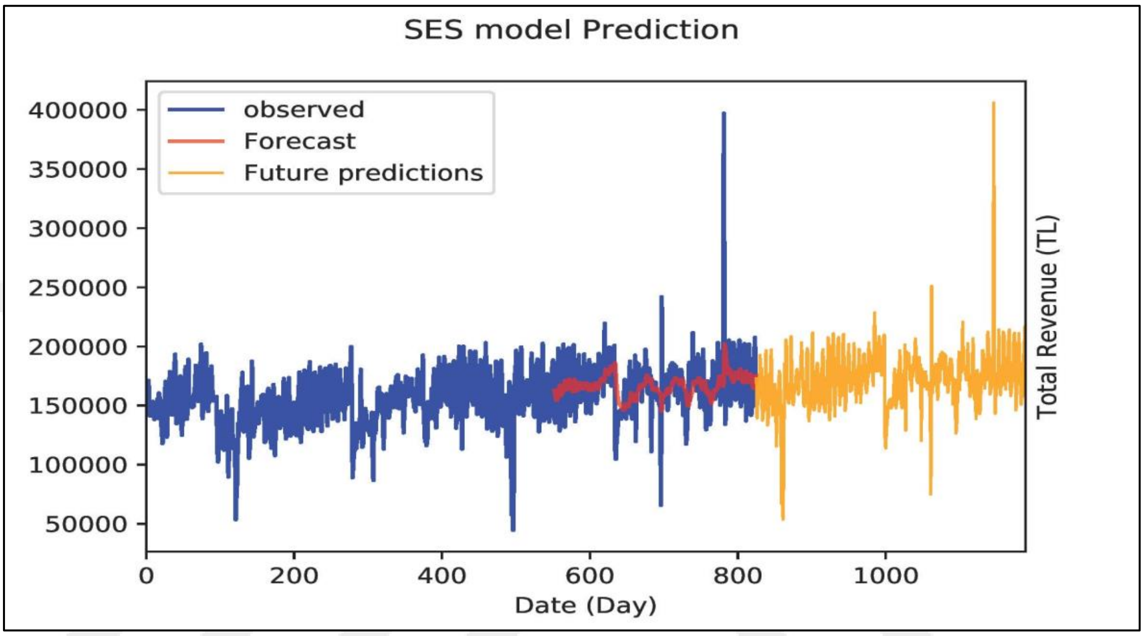
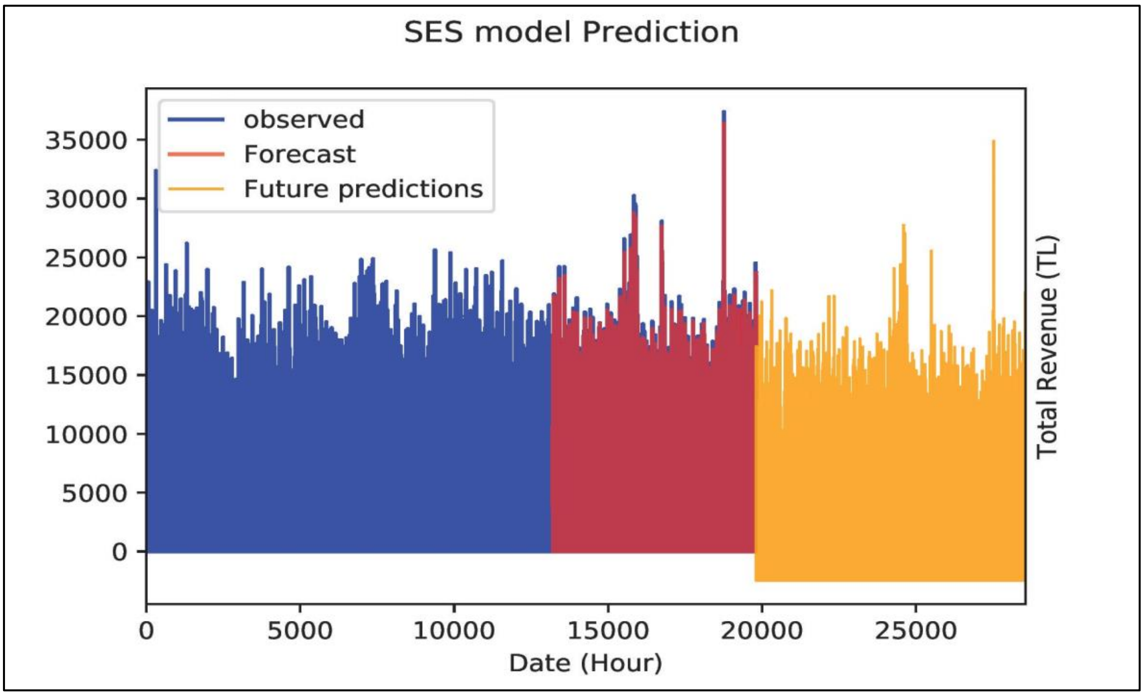


Figure 3.22:Hourly Sales Prediction Graph with SES model



3.10 HOLT WINTER'S EXPONENTIAL SMOOTHING - HWES

The seasonal trend version of SES is proposed by Holt Winter. The other name of the model is a third-level smoothing. As the data show no trends, the results obtained were the same as those for Simple Exponential Smoothing.

The pseudo-code that was utilized for the HWES model is given at appendices "see. Appendix A.1 Code-11".

If the graphs for the Holt-Winters were examined along the dataset, they were the same as the results in the SES model. Besides, it can be seen from the RMSE, AIC, and BIC criteria values. They were exactly the same.

3.11 RMSE, AIC, BIC, AND STANDARD DEVIATION

RMSE can be interpreted as follows. At a given x-axis value, there exist y values of observation, and when the regression line is delineated for the predictions at certain x values, there are ŷ values. A path to assess the accuracy of the predictions must be set up. The ideal form is that all the ŷ prediction values are equal to y actual observation values, which means having zero error. However, in most cases, there exists an error. To measure this error in a percentage, first, it must be found the residuals which are $\hat{y}_i$-y_i after subtraction must be taken the square of it and all the points which were desired to find errors will be added. Then, it is divided by n, which is the number of observations. The square root of the found MSE value is Root-MSE. If it is written in the formula:

$$RootMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \tag{3.18}$$

In the code, this value was calculated with the functions of the programming language, which is python.

This criterion was used as a measurement of the success of the models, but alone, this was not enough. Another criterion must be practiced as easily as this one. The criteria is explained in the following lines briefly and show the results of experiments in graphs and tables and discuss which model best fits the purpose of predicting the fast-food sales with time-series analysis.

AIC and BIC values as model selection criteria:

The best model is the one that mirrors reality best, one which explains the real values in the smallest error and enables the researcher to make predictions more accurately than the other models. Thus far, the root-mean-square error is explained; this is a criterion in regression analysis. It is referred to in future posts. Merely, in time-series, there are different criteria proposed in late years. If they are examined closely, it can be seen that AIC, which is the Akaike information criterion, is proposed and developed earlier than BIC, which is the Bayesian information criterion.

The Akaike information criterion was first described by the name giver Akaike. Kullback and Leibler found a measure in 1951 when trying to account for model reality, which diminishes the loss of data. Two decades later, Akaike published an association between the likelihood estimation and Kullback-Leibler measures (Fabozzi et al. 2014). This relationship was explained in a formula by Akaike, as follows:

$$AIC = N * log(RSS/N) + 2k \tag{3.19}$$

with N being the number of observations and RSS the mentioned residual sums of squares.

In regression problems, the smaller the value of AIC, the better the model is said to be found. Thus, the aim was to find the Akaike information criterion (AIC) values of the methods and prepare them and choose the proper model for the predictions.

Another touchstone for the concern is the Bayesian information criterion, which is shortly called BIC, proposed by Schwarz in 1978, also gave the name to it as a Schwarz information criterion that is used for model selection. There are differences in AIC and BIC that is the greater penalty imposed for the number of parameters by the BIC than AIC. Burnham and Anderson provided some theoretical arguments for the benefits of the AIC. Yang (2005) elucidated why AIC is better than BIC in model selection through the example of multivariate regression analysis.

$$BIC = k\ log(n) - \ 2log(L(\theta)) \qquad (3.20)$$

Here, in Equation 3.20, n is the size of the dataset, the number of observations, or the number of data points being worked on. k is the number of parameters that the model estimates, and θ are the set of all parameters. L(θ) stands for the likelihood of the model that was proposed.

Again, the smaller the value of the criteria, which is BIC, the better performance in explaining the observed values with the model; in other words, if the BIC value is smaller, the errors that are seen in estimation are smaller, and the likelihood increases. Thus, if the BIC value is smaller, it is said that the model is performing better and better fits the purposes.

A definition was consecrated, and all these answers are explained in the graphics and tables related to the fast-food data in particular.

Standard Deviation: Standard deviation is a measure of the amount of variation of a set of values (Wan et al. 2014). A low standard deviation indicates that the values are close to the mean that is the expected value of the dataset, while a high standard deviation shows us the values are scattered over a wider range. In regression, the smaller the value of RMSE compared to the standard deviation, the better the outcomes. Thus, it was the other criteria for us to choose the best models.

3.12 FLOW CHART

A flowchart that was used in this research is presented in Figure 3.23 and Figure 3.24

If it is wanted to tell how the program we implement for fast-food sales prediction and finding the best model that fits this data problem, this flow chart can be used. In the first step, the data were read from the .csv comma-delimited file and read which interval that will be used for time-series. For a weekly interval, 3 was entered, 2 was entered for a daily interval, and 1 was entered for an hourly interval.

In the next step, the data were organized. The DiscPrice field was used for the univariate models. On the other hand, for multivariate VAR and VARMA models, more fields were applied. They were organized according to intervals that were taken as input.

Next, the data were split into training and test sets. 1/3 for the test set and 2/3 for training set were the proportions in this split operation.

Subsequently, three DataFrames were prepared that would hold the results, the models for RMSE and standard deviation, the models AIC for the AIC values that the corresponding model gives, and the model BIC for the BIC values of the nine models.

There existed a procedure called Calculate to which the data were passed and prediction results were received in return. The multivariate models used another procedure called Calculate2. The graphs were plotted for actual values, test data prediction and future prediction for the time-series and fill the three DataFrames (models, models AIC and models BIC) with the results for each model were wanted to compare.

The other object of the program was to record future predictions. Each of the models was used and recorded their future sales predictions in fast-food using time-series. These nine models were found in Python libraries and were implemented, organize the data, train the models with the training data, test with test data, find the accuracy of the models and also made the future predictions depending on these models.

Figure 3.23: Flowchart - Main

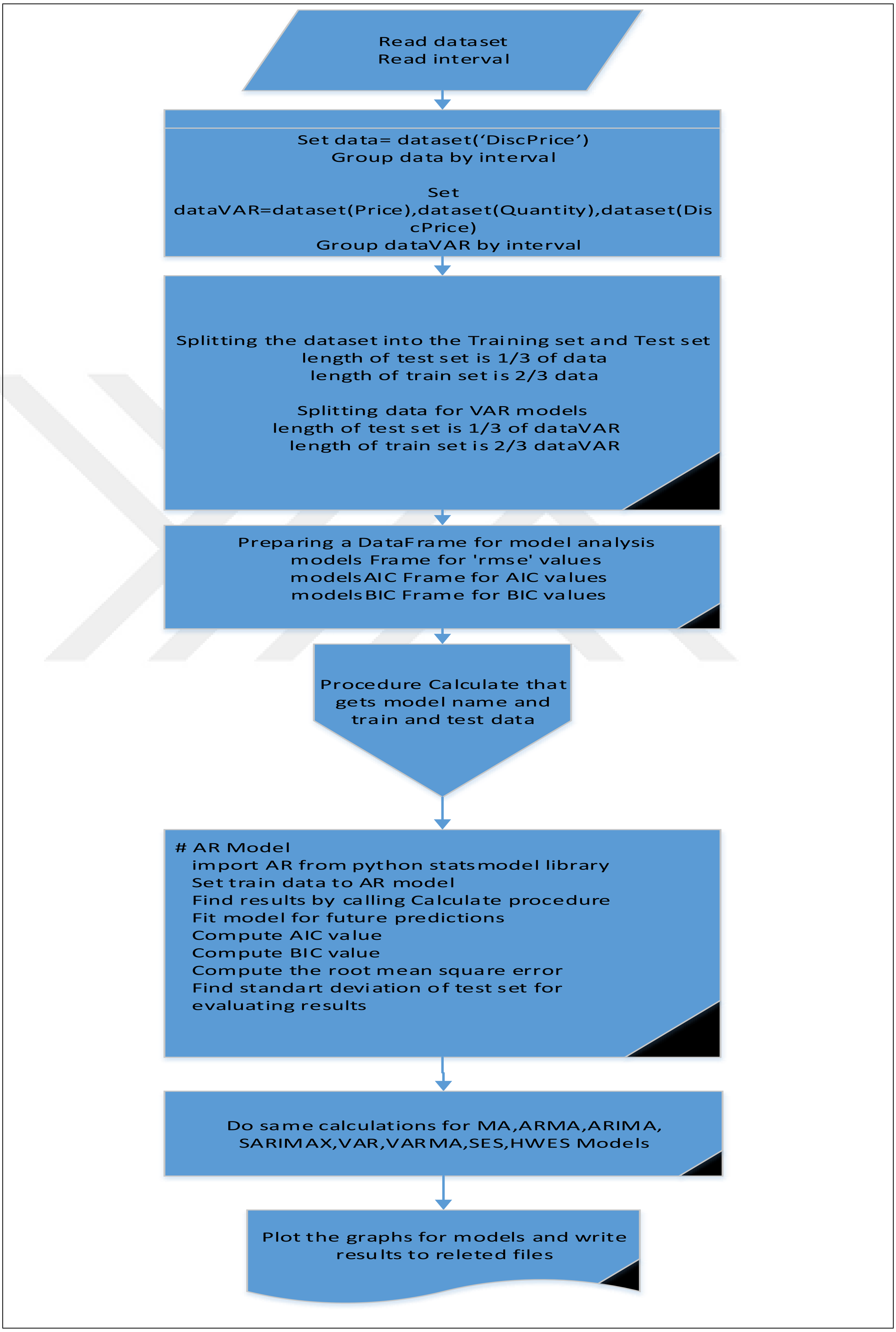
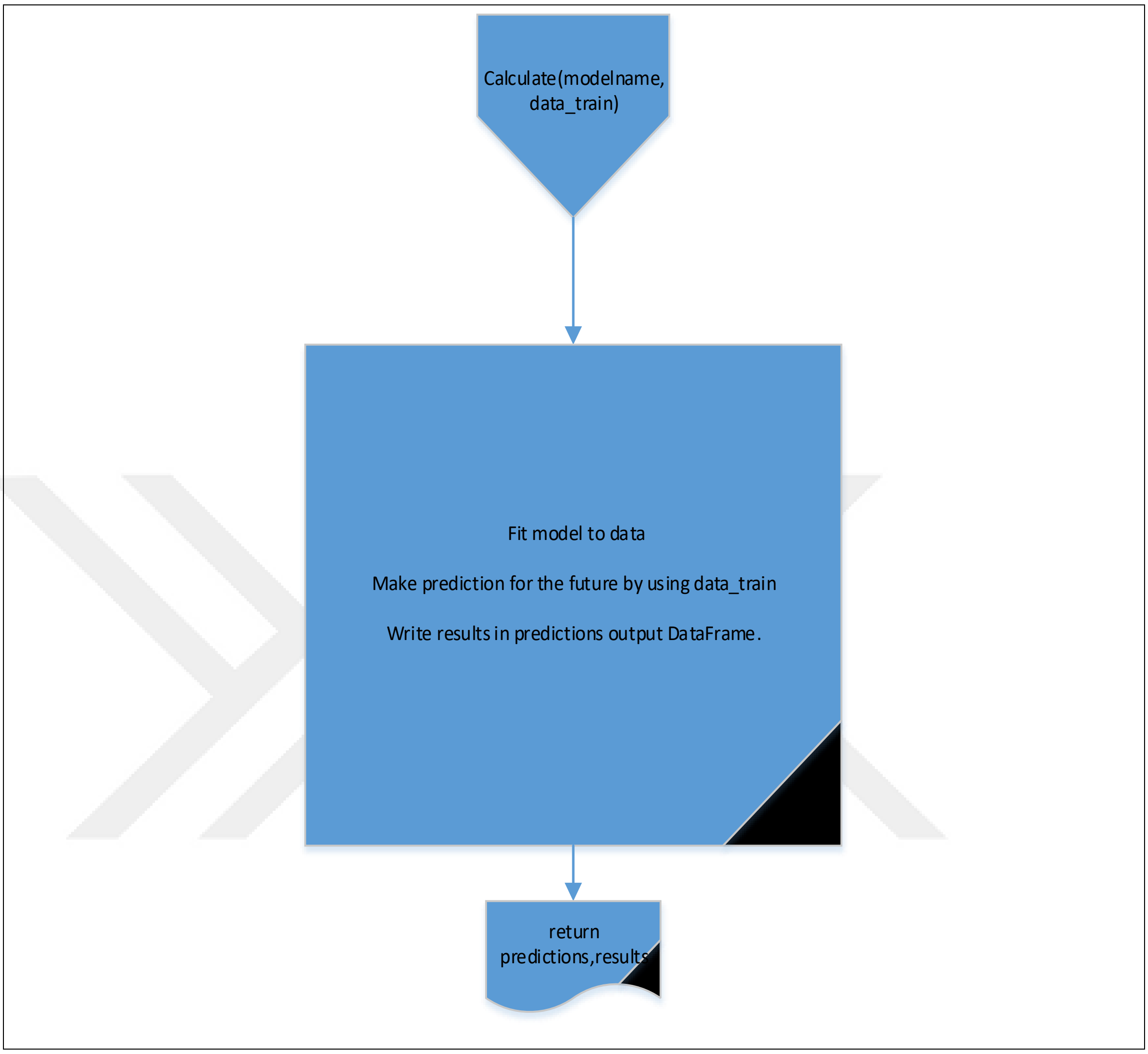


Figure 3.24: Flowchart - Procedure



4. EXPERIMENTS AND RESULTS

It was discussed in the discussions and conclusions section which model was better in sales prediction of the fast-food data concerning these four criteria. However, foremost, it is given in the Tables below for the experiments as results:

Table 4.1: RMSE values for models and Standard deviation in weekly intervals

	Standard deviation	AR	MA	ARMA	ARIMA	SARIMAX	VAR	VARMA	SES	HWES
rmse	0.189657	0.25361	0.339468	0.308575	0.295161	0.266589	0.048371	0.014907	0.305263	0.305263

As can be seen from Table 4.1, the smallest values were VARMA, with 0.014907 and VAR 0.048371. Also, AR was small, with 0.25361. The standard deviation was 0.189657, which was only greater than the VARMA and VAR models. According to this table, in weekly intervals, VARMA and VAR showed best, and AR was following model, but the AIC and BIC values were slightly different.

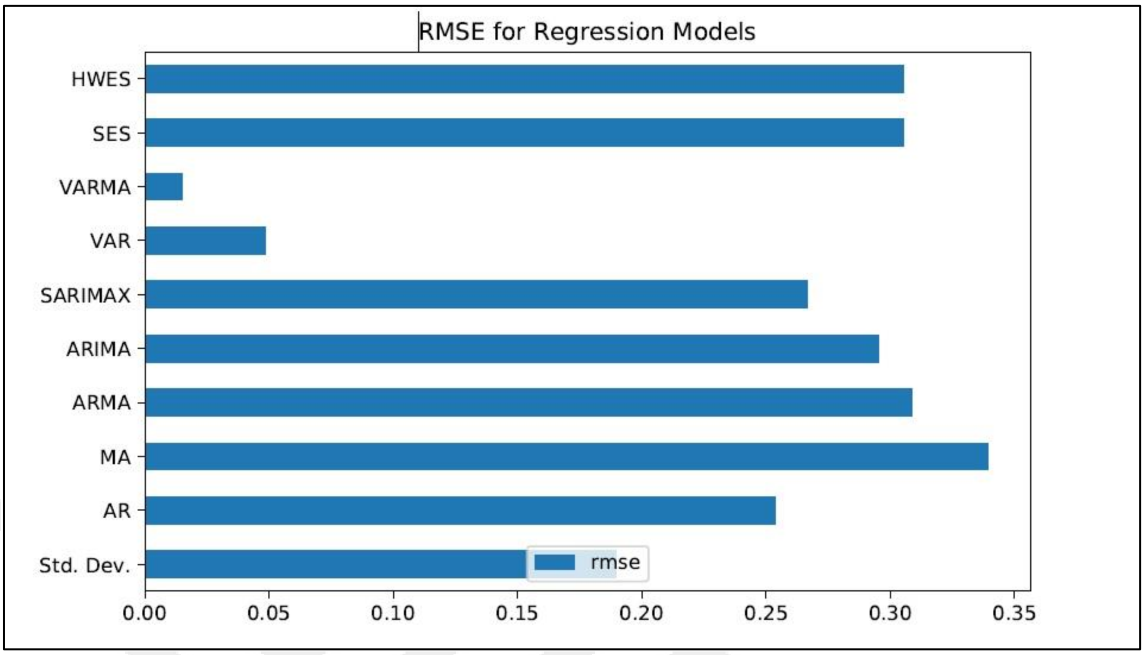
Table 4.2: AIC and BIC values for models at weekly intervals

	AR	MA	ARMA	ARIMA	SARIMAX	VAR	VARMA	SES	HWES
AIC	22.86172	2041.82	2041.82	2038.278	1659.571	48.81982	6784.68	2672.414	2672.414
BIC	23.28264	2048.966	2063.356	2059.716	1670.443	49.53589	6834.553	2677.955	2677.955

In Table 4.2, it can be seen that the best resolutions in the name of AIC and BIC criteria were derived by the AR and VAR models in a manner whereby the lowest values were received these models on a weekly basis.

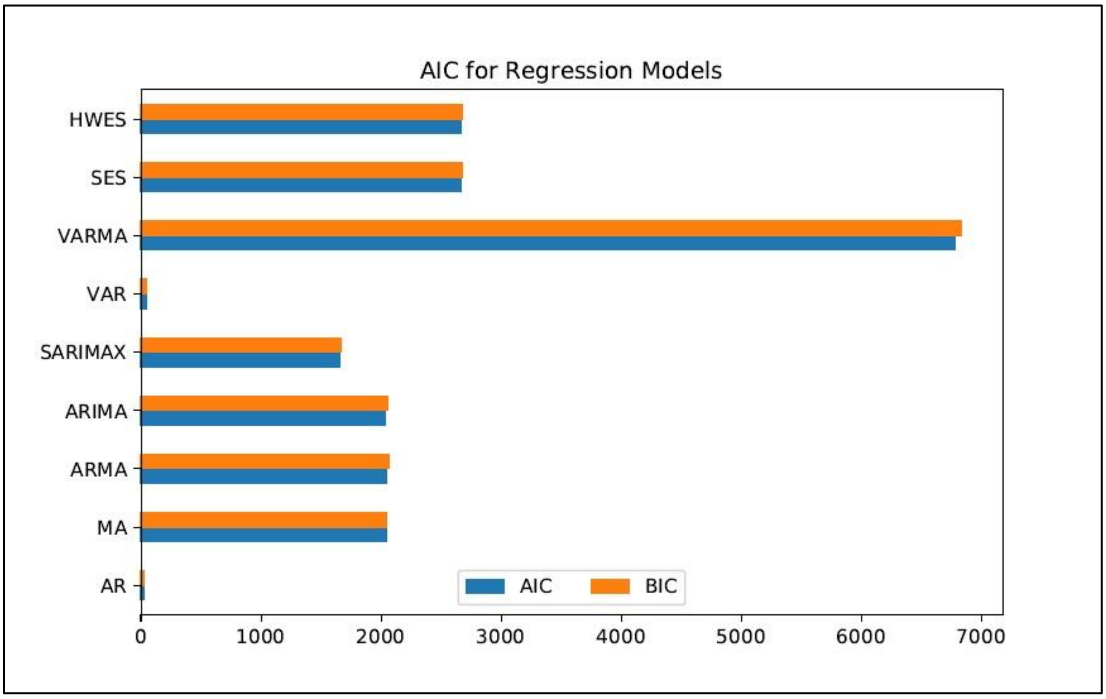
There are two graphs for these criteria, comparisons as follows:

Figure 4.1: RMSE and standard deviation for the models on a weekly basis



VAR and VARMA performed with good results below the standard deviation, but AR was also beneficial.

Figure 4.2: AIC and BIC values for the models on a weekly basis



According to these criteria, the AR model and VAR model were the best models. Contrary to RMSE performance, the VARMA model was the worst model according to these criteria.

In daily intervals of the time-series, the following effects were analyzed. The first examined table is RMSE and standard deviation.

Table 4.3: RMSE values for models and standard deviation in daily intervals

	Standard deviation	AR	MA	ARMA	ARIMA	SARIMAX	VAR	VARMA	SES	HWES
rmse	0.076309	0.074391	0.090661	0.357275	0.379133	0.093241	0.070886	0.034115	0.098899	0.098899

The best result was again VARMA, with 0.034115. The second result was VAR, with 0.070886. In addition, AR showed better performance, below the level of standard deviation.

Table 4.4: AIC and BIC values for models in daily intervals

	AR	MA	ARMA	ARIMA	SARIMAX	VAR	VARMA	SES	HWES
AIC	19.40673	12394.93	12394.93	12286.23	12055.51	41.25716	41246.13	16528.04	16528.04
BIC	19.56728	12407.87	12304.89	12325.03	12076.93	41.42912	41331.01	16537.47	16537.47

As it was in weekly intervals again, the AR model and VAR models were the best according to Table 4.4.

The graphs for the daily interval are given as follows.

Figure 4.3: RMSE and standard deviation for the models on a daily basis

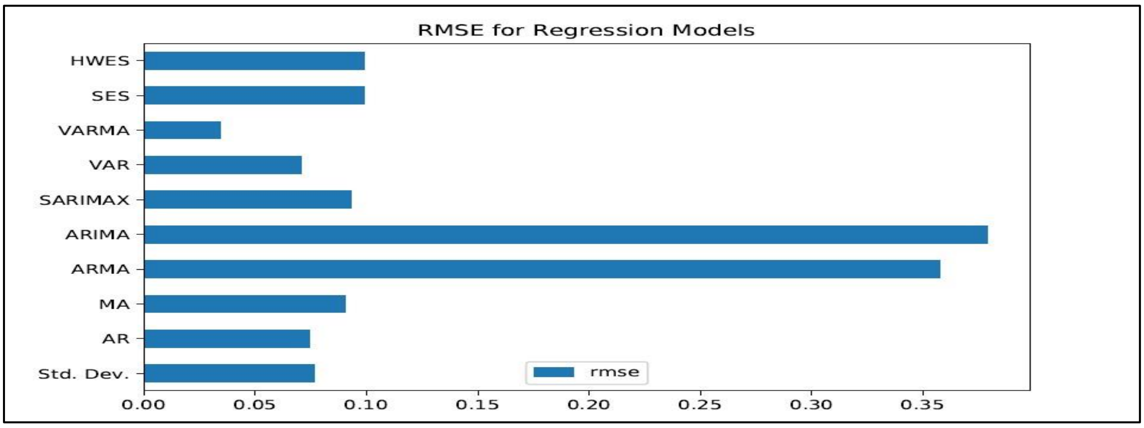
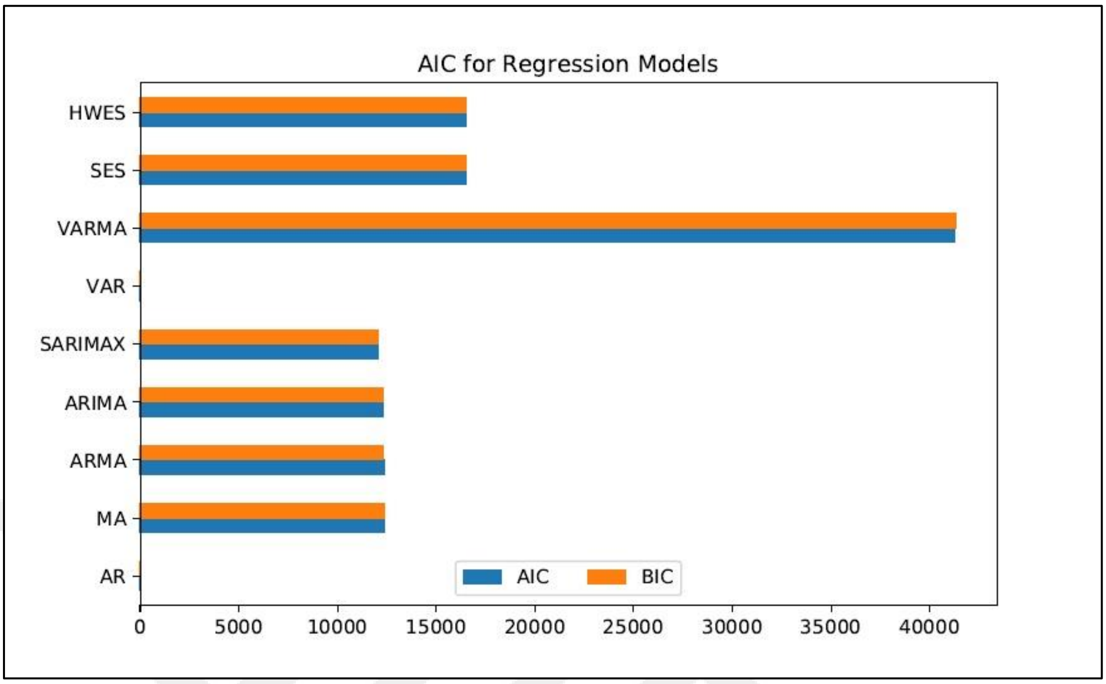


Figure 4.4: AIC and BIC values for the models on a daily basis



At hourly intervals of the time-series, the following answers were received. The first examined table is RMSE and standard deviation:

Table 4.5: RMSE values for models and standard deviation in hourly intervals

	Standard deviation	AR	MA	ARMA	ARIMA	SARIMAX	VAR	VARMA	SES	HWES
rmse	0.15492	0.071067	0.155991	0.244931	0.177035	0.098205	0.136815	0.088398	0.137158	0.137158

This time, as can be seen from Table 4.5, the best model was changing. If recalled from Table 4.1 and Table 4.3, the AR model was the third in the order. However, this time, it can be seen in Table 4.5 that the AR model was the best model, which was followed by the VARMA and SARIMAX models, which were below the level of standard deviation.

If it has deeply looked at the following criteria, it can be said which is the best model:

Table 4.6: AIC and BIC values for models in hourly intervals

	AR	MA	ARMA	ARIMA	SARIMAX	VAR	VARMA	SES	HWES
AIC	15.491132	258,003.77	258,003.77	247,422.2	245,658.47	30.71445	777,906.36	323,177.53	323,177.53
BIC	15.515589	258,026.23	244,883.68	247,489.59	245,695.9	30.726411	778,048.45	323,193.31	323,193.31

Again, the AR model had the best value with respect to AIC and BIC criteria. The VAR model followed as second best. The differences can be seen visually from the graphs below.

Figure 4.5: RMSE and standard deviation for the models on an hourly basis

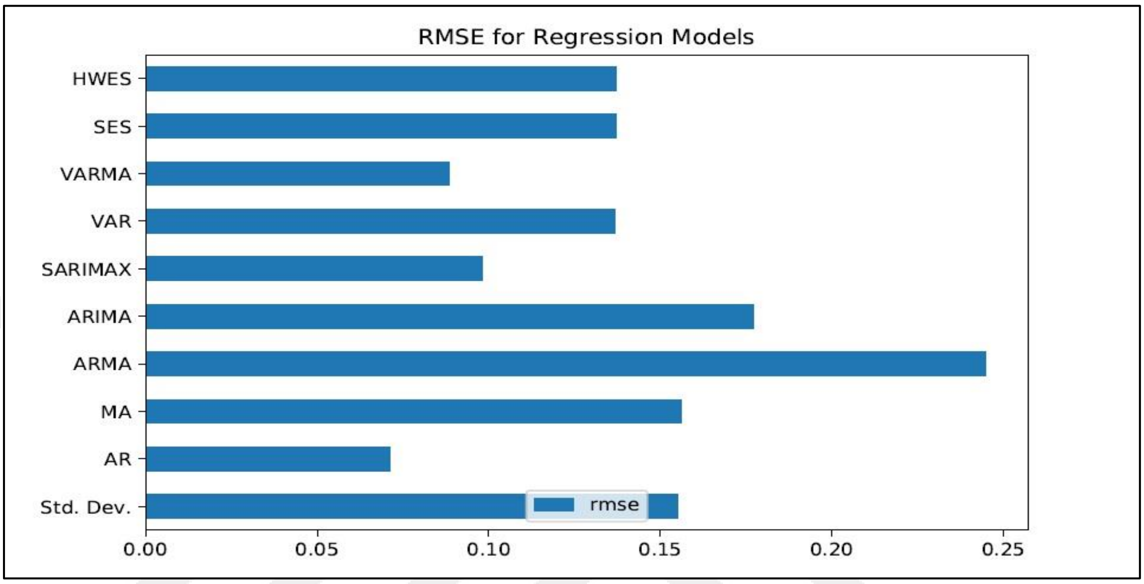
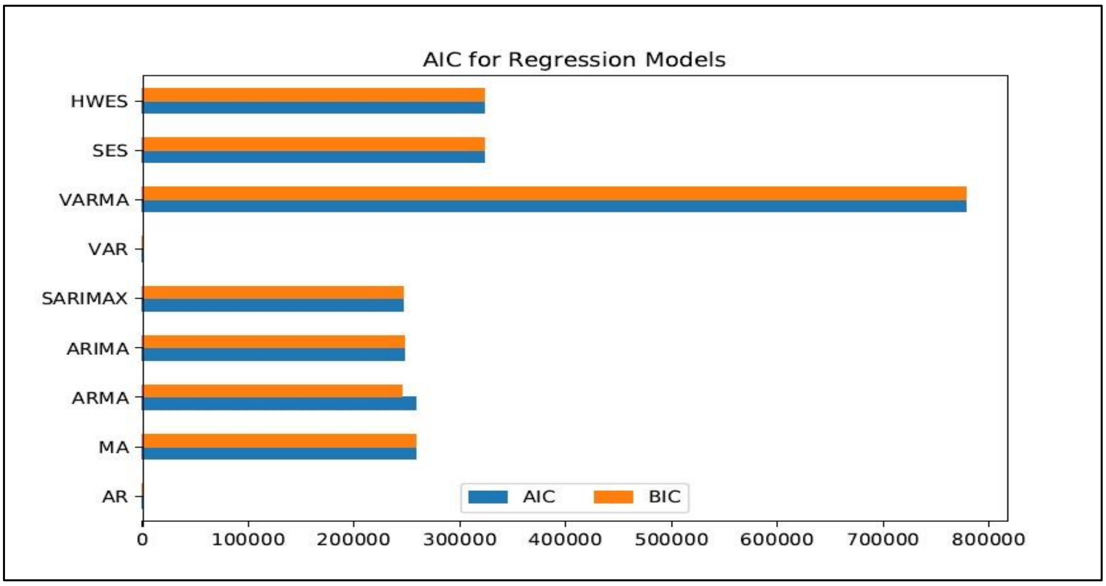


Figure 4.6: AIC and BIC values for the models on an hourly basis



5. DISCUSSION AND CONCLUSION

This study aimed to find the best regression model that explains the sales prediction for the time-series dataset of a fast-food firm. Nine different traditional models for the time-series were used. Univariate and multivariate approaches, according to these models, were examined.

The comparison between the nine models was accomplished by testing the data according to important criteria in the time series models' problems. These were the AIC, BIC, and RMSE values.

There are time intervals for the sales data that the time series were organized. In most of the results, it can be seen that AR model was the best fit for prediction purposes, and the VAR model followed it. According to the criteria above, it can be seen from the Autocorrelation and Partial Autocorrelation graphs that the data fit best for an AR model.

There exist other models, such as SVM and ANN, but this study aimed to obtain a resolution in which the traditional regression model best explains this sales prediction problem. Also, a program was developed for making predictions for future sales on a weekly, daily, and hourly interval basis. The dataset was used as a training and test set for model selection, and the models made future predictions that were recorded.

In conclusion, it is not a surprise that the AR model was the best prediction model, since the ACF and PACF graphs showed us that it is this kind of data was at hand. In the fast-food sector, it is important to make predictions, and the best regression models are the AR and VAR models, according to this research.

REFERENCES

Books

Adhikari, R. & Agrawal, R. K., 2013. *An Introductory Study on Time Series Modeling and Forecasting*, Germany: LAP Lambert Academic Publishing.

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B. G., 2014. *The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications*., John Wiley & Sons, Inc., Appendix E.

Hipel, K.W., & McLeod, A.I., 1994. *Time Series Modelling of Water Resources and Environmental Systems*, Elsevier.

Hyndman, R.J., & Athanasopoulos, G., 2018. *Forecasting: principles and practice,* 2nd print edition, Published by Otexts.

J de Smith, M., 2018. *Statistical Analysis Handbook*, Edinburg: The Winchelsea Press, Drumlin Security Ltd, Section:17.3 ARMA and ARIMA (Box- Jenkins) models.

Nesheim M.C., Oria M., & Yih P.T., 2015. A Framework for Assessing Effects of the Food System, Washington (DC): National Academies Press (US), p. 21,197.

Shumway, R. S. & Stoffer, D. S., 2011. *Time Series Analysis and Its Applications With R Examples,* Third Edition New York: Springer.

Wold, H., 1938. *A study in the analysis of stationary time series.* Second Edition. Stockholm: Almqvist & Wiksell

Zivot E., & Wang, J., 2002. *Modelling Financial Time Series with S-PLUS.* Insightful Corporation (formerly MathSoft, Inc) Multivariate time series and VAR models. Seatle, WA: Insightful Corporation.

Periodicals

Alsudani, A., Sabeeh, R. & Liu, J., 2017. The Use of Some of the Information Criterion in Determining the Best Model for Forecasting of Thalassemia Cases Depending on Iraqi Patient Data Using ARIMA Model. *Journal of Applied Mathematics and Physics*. **5** (3), pp. 667-679.

Arunraj, N.S. & Ahrens D., 2015. A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting, *Int. J. Production Economics*. **170**, pp. 321-335.

Athanasopoulos G., & Vahid F., 2008. VARMA versus VAR for Macroeconomic Forecasting, *Journal of Business & Economic Statistics*. **26** (2), pp. 237-252.

Chen, F.L. & Ou, T.Y., 2009. Gray relation analysis and multilayer functional link network sales forecasting model for perishable food in convenience store. *Expert Syst. Appl*. **36** (3), pp. 7054–7063.

Doganis, P., Alexandridis, A., Patrinos, P., & Sarimveis, H., 2006. Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *Journal of Food Engineering*. **75** (2), pp. 196–204.

Gooijer, J. G., & Hyndman, R.J., 2006. 25 years of time series forecasting, *International Journal of Forecasting*. **22** (3), pp. 443–473.

Hasin, M., Ghosh, S., & Shareef, M., 2011, An ANN Approach to Demand Forecasting in Retail Trade in Bangladesh. *International Journal of Trade, Economics, and Finance*. **2** (2), pp. 154-160.

Kim, K.J. 2003. Financial time series forecasting using support vector machines, *Journal of Neurocomputing*. **55** (1-2), p.308

Lee, W., Chen, C., & Chen, K., 2012. A comparative study on the forecast of fresh food sales using logistic regression, moving average and BPNN methods, *Journal of Marine Science and Technology*. **20** (2), pp.142–152.

Makridakis, S., & Hibon M., 1998. ARMA Models and the Box–Jenkins Methodology, *Journal of Forecasting.* **16** (3), pp. 147-207.

Newbold, P., 1983. ARIMA model building and the time series analysis approach to forecasting. *Journal of Forecasting.* **2** (1), pp. 23–35.

Shukla, M. and Jharkharia, S., 2013. Applicability of ARIMA models in the wholesale vegetable market: an investigation. *Int. J. Inf. Syst. Supply Chain Manag.* **6** (3), pp. 105–119.

Wan, X., Wang, W., Liu, J. et al., 2014. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol.* **14** (1), p. 135.

Zhang, G. P., 2007. A neural network ensemble method with jittered training data for the time series forecasting, *Information Sciences.* **177** (23), pp. 5329–5346.

Zhang, G. P., 2003. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model, *Neurocomputing.* **50**, pp. 159-175

Other Publications

Natrella, M., Croarkin, C., And Guthrie, W., 2019. Section Box-Jenkins Models, Section 6.4.4.5., Section Definitions, Applications, and Techniques, Section 6.4.1. [online] http://www.itl.nist.gov/div898/handbook, *e-Handbook of Statistical Methods*, [accessed on 7 December 2019].

Kong, J., & Martin, G., 1995. A backpropagation neural network for sales forecasting. *In: Proceedings of the IEEE International Conference on Neural Networks*, 2 pp. 1007–1011.

Raicharoen, T., Lursinsap, C., & Sanguanbhoki, P., 2003. Application of Critical Support Vector Machine to Time-Series Prediction. *Proceedings of the International Symposium on Circuits and Systems, Bangkok*, pp. 741-744.

Rao S. S., 2018. Texas A&M University, Dept. of Statistics, *A course in Time Series Analysis*, [online] http://www.stat.tamu.edu/~suhasini/teaching673/time_series.pdf [accessed on 7 December 2019].

Türkiye-Avrupa Eğitim ve Bilimsel Araştırmalar Vakfı (TAVAK), 2018. *Türkiye'de Yeme İçme Sektörünün Yapısı Gastroekonomisinin Boyutları Nasıl Gelişiyor?* Report.

APPENDICES

Appendix A.1 Code parts from implementation

Code-1:

```
inter="1H"  #hourly
inter="1d"  #daily
inter="7d" # weekly
seri3=pd.DataFrame({'DiscPrice':dataset['DiscPrice'], 'date1':dataset.index})
seri2 = seri3.groupby(pd.Grouper(key='date1', freq=inter)).sum()
```

Code-2:

```
seriVAR3=pd.DataFrame({'Price':dataset['Price'],'Quantity':dataset['Quantity'],
'DiscPrice':dataset['DiscPrice'], 'date1':dataset.index})
seriVAR2 = seriVAR3.groupby(pd.Grouper(key='date1', freq=inter)).sum()
```

Code-3:

```
model = AR(seri_train)
predictions_AR,results_AR=Calculate(model,"AR",seri_train,seri_test)
```

Code-4:

```
model = ARMA(seri_train, order=(0,1))
predictions_MA,results_MA=Calculate(model,"MA",seri_train,seri_test)
model = ARMA(seri,order=(0,1))
results3_MA=model.fit()
```

```
forecast= results3_MA.predict(start=ntrain+nsplits, end=ntrain+nsplits+step)
```

Code-5:

```
model = ARMA(seri_train, order=(6,2))

predictions_ARMA,resultsARMA=Calculate(model,"ARMA",seri_train,seri_test)

model = ARMA(seri,order=(6,2))

results3_ARMA=model.fit()

forecast= results3_ARMA.predict(start=ntrain+nsplits, end=ntrain+nsplits+step)
```

Code-6:

```
model = ARIMA(seri_train, order=(6,0,1))

predictionsARIMA,resultsARIMA=Calculate(model,"ARIMA",seri_train,seri_test)

model = ARIMA(seri,order=(6,0,1))

results3_ARIMA=model.fit()

forecast= results3_ARIMA.predict(start=ntrain+nsplits, end=ntrain+nsplits+step)
```

Code-7:

```
model =SARIMAX(seri_train, order=(1, 1, 1 ),seasonal_order=(1,0,1, 12))
predictions_SARIMAX,results_SARIMAX=Calculate2(model,"SARIMAX",seri_train,
seri_test)

mod2 = SARIMAX(seri, order=(1, 1, 1), seasonal_order=(1,0,1, 12))

results2 = mod2.fit()

forecast=results2.predict(start=ntrain+nsplits, end=ntrain+nsplits+step)
```

Code-8:

```
seriVAR3=pd.DataFrame({'Price':dataset['Price'],'Quantity':dataset['Quantity'],
'DiscPrice':dataset['DiscPrice'], 'date1':dataset.index})

seriVAR2 = seriVAR3.groupby(pd.Grouper(key='date1', freq=inter)).sum()

seriVAR=pd.DataFrame(seriVAR2)

model = VAR(seriVAR)

results_VAR = model.fit(3)

print(results_VAR.summary())

# make prediction

ypred_VAR = results_VAR.forecast(results_VAR.y, steps=len(seri))
```

Code-9:

```
model = VARMAX(seriVAR,Order=(2,1))

results_VARMA = model.fit()

# make prediction

ypred_VARMA = results_VARMA.predict(ntrain+1, len(seri)+100,typ='levels')
```

Code-10:

```
model = SimpleExpSmoothing(seri)

results2_SES = model.fit()

# make prediction

ypred2_SES = results2_SES.predict(ntrain+1,end=len(seri))
```

Code-11:

```
model = ExponentialSmoothing(seri)

results2_HWES = model.fit()

# make prediction

ypred2_HWES = results2_HWES.predict(ntrain+1,end=len(seri))
```