ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF ARTS AND SOCIAL SCIENCES

SALES FORECASTING IN FASHION RETAIL INDUSTRY WITH CLASSICAL AND MACHINE LEARNING METHODS

M.Sc. THESIS

Hanife IŞIK

Department of Economics

Economics M.Sc. Programme

JUNE 2020

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF ARTS AND SOCIAL SCIENCES

SALES FORECASTING IN FASHION RETAIL INDUSTRY WITH CLASSICAL AND MACHINE LEARNING METHODS

M.Sc. THESIS

Hanife IŞIK
(412171006)

Department of Economics

Economics M.Sc. Programme

Thesis Advisor: Assoc. Prof. Dr. Tolga YURET

JUNE 2020

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ SOSYAL BİLİMLER ENSTİTÜSÜ

MODA PERAKENDESİ SEKTÖRÜNDE KLASİK VE MAKİNE ÖĞRENMESİ METODLARI İLE SATIŞ TAHMİNİ

YÜKSEK LİSANS TEZİ

Hanife IŞIK
(412171006)

İktisat Anabilim Dalı

İktisat Yüksek Lisans Programı

Tez Danışmanı: Doç. Dr. Tolga YURET

HAZİRAN 2020

Hanife IŞIK, a M.Sc. student of ITU Graduate School of Arts and Social Sciences student ID 412171006, successfully defended the thesis/dissertation entitled – Sales Forecasting In Fashion Retail Industry With Classical And Machine Learning Methods, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assoc. Prof. Dr. Tolga YURET** Istanbul Technical University	.....
Jury Members :	**Prof. Dr. Bülent GÜLOĞLU** Istanbul Technical University	.....
	Asst. Prof. Dr. Murat GÜVEN Sakarya University	.....

Date of Submission : 15 June 2020
Date of Defense : 21 July 2020

To my family,

FOREWORD

This thesis is very valuable for me as it gives me the opportunity to work together in two of my favorite fields which are economics and fashion retail sector. Moreover, this study has been an important development tool for learning and developing myself in different programmings. I also think that this study has an important place due to drawing attention to the fact that classical methods remain important in our lives and showing how new developing technologies can be integrated into our lives as added value. Therefore, I am very pleased with the contribution I have made both to the literature and to myself with this study.

I would like to thank my advisor Assoc. Prof. Dr. Tolga YURET for his valuable supports and assistance during my M.Sc. Thesis study. Also, I really appreciate to Prof. Dr. Bülent GÜLOĞLU for his contributions and suggestions to my study.

I also would like thank my parents Mahinur and Enser and my brother Ufuk for their unlimited support and love in every moment of my life.

June 2020 Hanife IŞIK

TABLE OF CONTENTS

Page

FOREWORD ..... **ix**
TABLE OF CONTENTS ..... **xi**
ABBREVIATIONS ..... **xiii**
SYMBOLS ..... **xv**
LIST OF TABLES ..... **xvii**
LIST OF FIGURES ..... **xix**
SUMMARY ..... **xxi**
ÖZET ..... **xxiii**
1. INTRODUCTION **1**
2. LITERATURE REVIEW **5**
3. METHODOLOGY **13**
3.1 Autoregressive Moving Avarage Model 15
3.2 Bagging Regression Tree 19
3.3 Random Forest Algorithm 19
3.4 Gradient Boosting Algorithm 20
3.5 Adaptive Boosting Algorithm 20
3.6 Extreme Gradient Boosting Algorithm 21
3.7 Extremely Randomized Trees Algorithm 21
4. DATA DESCRIPTION **23**
5. RESULTS **25**
5.1 Forecasting Error Measures 25
5.2 Variable Importance 26
5.3 Model Comparison 28
6. CONCLUSIONS **33**
REFERENCES **35**
CURRICULUM VITAE **39**

ABBREVIATIONS

ACF **:** Autocorrelation Function
AdaBoost : Adaptive Boosting
ADF **:** Augmented Dickey-Fuller
AI **:** Artificial Intelligence
AIC **:** Akaike Information Criteria
ANN **:** Artificial Neural Network
AR **:** Autoregressive
ARIMA **:** Autoregressive Integrated Moving Average
ARMA **:** Autoregressive Moving Average
CART **:** Classification And Regression Trees
VAR **:** Vector Autoregressive
KPSS Test **:** Kwiatkowski – Phillips – Schmidt – Shin
LASSO **:** Least Absolute Shrinkage and Selection Operator
MA **:** Moving Average
MAE **:** Mean Absolute Error
MAPE **:** Mean Absolute Percentage Error
ME **:** Mean Error
ML **:** Machine Learning
MPE **:** Mean Percentage Error
OLS **:** Ordinay Least Squares
PACF **:** Partial Autocorrelation Function
PSE **:** Predicted Square Error
RMSE **:** Root Mean Squared Error
SKU **:** Stock Keeping Unit
XGBoost **:** Extreme Gradient Boosting

SYMBOLS

μ	**:** Mean
x_t	**:** Time series
$\emptyset_p$	**:** Constants at AR Process
φ_p	**:** Constants at MA Process
U_t	**:** Constants at MA Process
p	**:** Order of AR Process
d	**:** Order of Integrated Process
q	**:** Order of MA Process

LIST OF TABLES

Page

Table 3.1 : ADF Test result **16**
Table 3.2 : KPSS Test result **17**
Table 4.1 : Descriptive statics. **24**
Table 5.1 : Model performances without cross-validation. **28**
Table 5.2 : Model performances with cross-validation. **29**

LIST OF FIGURES

Page

Figure 2.1 : World textiles and clothing exports by country 2018**. 5**
Figure 2.2 : Commercial product Hierarchy in clothing industry (Thomassey, 2010)**. 7**
Figure 2.3 : Criteria for selection of forecasting methods (Armstrong & Yokuma, 1995)**. 9**
Figure 3.1 : Flow chart of applied methodology**. 14**
Figure 3.2 : Autocorrelation function of daily sales. **..... 18**
Figure 3.3 : Partial autocorrelation function of daily sales. **..... 18**
Figure 4.1 : Daily satore sales. **..... 23**
Figure 5.1 : Spearman correlation analysis. **..... 27**
Figure 5.2 : Variable importance from random forest algorithm**. 27**
Figure 5.3 : Comparison of models with cross-validation**. 29**

SALES FORECASTING IN FASHION RETAIL INDUSTRY WITH CLASSICAL AND MACHINE LEARNING METHODS

SUMMARY

This study includes the usage of classical econometric methods and machine learning models for short-term demand forecast of a fashion retail firm with hundreds of shops. Forecasting performances of different methods are copared. Retail clothing sector has an important place among all sectors both in Turkey as well as all over the world. According to the data of the global textile and clothing sector exports, Turkey has become one of the most important countries in this sector with 12 billion USD dollars in exports by 2018. This sector, which includes different functions from planning to production, logistics to merchandising, is a dynamic sector in which rapid changes occur due to impact of fashion and trend. Rapid adaptation to changes in this dynamic sector can contribute to the profitability of companies.

At this point, demand forecasting plays an important role for companies to organize their activities and take necessary measures in advance. Demand forecasting affects many points in this sector, from production planning to shop planning. Long-term demand forecast is an important input for many activities for the company, from labor investment to managing financial resources, opening new stores, and determining the collection mix. The most important activities influenced by the short-term demand forecast in this sector that are logistics and merchandising activities. While the unpredictable demand can cause a bottleneck or idle capacity in logistics warehouses. On the merchandising side, less inventory than demand case can cause loss of sales while inventory more than demand can causes to inventory cost in the store and disruption of merchandising operations due to physical space constraints.

In this study, demand estimation is performed for 2 different styles of 30 different stores at the daily level. This daily estimate is a requirement for the decision of how many products should be shipped to which store. This decision directly affects both logistics and merchandising activities and profitability. For this reason, there are teams working on long and short-term demand forecasting both in the company where this study is carried out and in different companies in this sector.

More accurate demand forecasting has become critical in this increasingly competitive sector. For this purpose, in this study, demand forecasting is performed by using different econometric and machine learning methods and the most accurate method for this sector is tried to be determined. While classical methods use linear algorithms when making predictions, machine learning methods use nonlinear algorithms. Therefore, using machine learning methods in data with high nonlinear effects is useful in order to increase accuracy and explain nonlinear effects in the model. Both methods aim to enhance the forecasting accuracy. The basis of classical econometric models is based on mathematics and statistics. Regression, Exponential Smoothing, Autoregressive Integrated Moving Average (ARIMA), Vector Autoregressive (VAR) and Vector Error Correction Model (VECM) methods are frequently used classical econometric methods. On the other hand, machine learning, which is a branch of

artificial intelligence, has important algorithms in the field of forecasting. Especially in the last 10 years, studies on machine learning and algorithms developed in this field have increased significantly. The successful results of these algorithms and the ease of application has increased the usage of these algorithms. However, these methods are highly dependent on computers and computer sciences can be considered as a disadvantage of these methods. In addition, black-box features in the functioning of algorithms can be interpreted as another disadvantage of these algorithms while classical econometric methods provide advantages with ease of explanation.

In this study, linear regression and ARIMA are used as forecasting methods for classical econometric methods. Machine learning methods that are Bagging Regression Tree, Extremely Randomized Trees, Random Forest Algorithm, Gradient Boosting, Adaptive Boosting, Extreme Gradient Boosting Algorithm are applied.

One of the widely used classical econometric method is ARIMA is used within this study. The series must be stationary to make predictions in the time series. Therefore, ADF test and KPSS test were used to test the stationarity of the series. ARMA model was used with the data set due to being stationary time series according to the performed test results. In addition, the parameters of the ARMA model were determined by analyzing the autocorrelation function and partial autocorrelation function of the data set. According to the evaluation of ARMA models, the most appropriate parameters were decided according to the Akaike Information Criteria (AIC).

Machine learning models differ from ARIMA in terms of establishing models by using independent variables that can affect the dependent variable. While preparing this study data set, all independent variables that can influence dependent variable are included in the data set. However, it does not mean that all included features are important. Therefore, feature engineering methods such as Spearman correlation analysis, variance analysis and random forest algorithm methods are applied in order to eliminate the unnecessary independent variables from models. Before applying the methods, the data set is divided into 3 parts as training data, validation data and test data. 20% of the data set was used as test data. Cross-validation is performed to avoid the over fitting problem of machine learning methods. Cross-validation, resampling procedure for evaluating machine learning methods, also helps to generalize predictive performance. For this purpose, the k-fold cross-validation method is applied k with 5.

Eight different methods where two of which are classical method and six of which are machine learning method are compared according to the means absolute error in test data and root mean square error. Adaptive Boosting Algorithm shows the highest performance with mean absolute error and root mean square error with 38.42 and 57.60, respectively. This method was followed by the Gradient Boosting Algorithm. Linear Regression method, which is one of the classical methods, has an average performance, while the ARIMA method has the most mediocre performance.

MODA PERAKENDESİ SEKTÖRÜNDE KLASİK VE MAKİNE ÖĞRENMESİ METODLARI İLE SATIŞ TAHMİNİ

ÖZET

Bu çalışma Türkiye'de ve dünyada moda perakendesi sektöründe yüzlerce mağazaya sahip olan bir firmanın, kısa dönemli talep tahmininin klasik ekonometrik yöntemler ve makine öğrenmesi modelleri ile yapılarak bu modellerin tahmin performanslarının karşılaştırılmasını kapsamaktadır. Perakende giyim sektörü tüm dünyada olduğu gibi Türkiye'de de sektörler arasında önemli bir yere sahiptir. 2018 yılı küresel tekstil ve giyim sektörü ihracat verilerine göre Türkiye 12 milyar dolar ihracat ile bu sektördeki önemli ülkelerden biri olmuştur. Bu sektör planlamadan üretime, lojistikten mağazacılığa kadar farklı fonksiyonları bünyesinde barındıran bu sektör, içerisinde moda ve trend etkisini barındırmasından dolayı hızlı değişimlerin yaşandığı dinamik bir sektördür. Bu dinamik sektörde değişikliklere hızlı adapte olmak, firmaların karlılıklarına çok önemli katkı sağlayacaktır.

Bu noktada firmaların faaliyetlerini düzenlemeleri ve gerekli tedbirleri önceden almaları için talep tahmini önemli bir rol oynamaktadır. Talep tahmini bu sektörde üretim planlamasından mağaza planlamasına kadar birçok noktaya etki etmektedir. Uzun dönemli talep tahminleri firma için iş gücü yatırımından finansal kaynakların yönetilmesi, yeni mağaza açılışından koleksiyon karışımının belirlenmesine kadar birçok faaliyete önemli bir girdi oluşturur. Bu sektörde kısa dönemli talep tahmininin etki ettiği en önemli faaliyetler ise lojistik ve mağazacılık faaliyetleridir. Doğru şekilde tahmin edilemeyen talep, lojistik depolarda dar boğaza veya atıl kapasiteye neden olurken, mağazacılık tarafında ise talepten az stok gönderilmesi durumunda satış kaybına, talepten çok stok gönderilmesi durumunda mağazada stok maliyetine ve fiziksel alan kısıtları nedeniyle mağazacılık operasyonlarının aksamasına neden olacaktır.

Bu çalışmada, 30 farklı mağazanın, 2 farklı stil bölümü için günlük seviyede talep tahmini yapılmıştır. Günlük seviyede yapılan bu tahmin, hangi mağazaya kaç adet ürün sevk edilmesi gerektiği kararı için bir gereklilik oluşturmaktadır. Verilen bu karar hem lojistik hem de mağazacılık faaliyetlerini ve karlılığı direk olarak etkilemektedir. Bu nedenle hem bu çalışmanın yapıldığı firmada hem de bu sektördeki farklı firmalarda uzun ve kısa dönemli talep tahmini üzerine çalışan ekipler bulunmaktadır.

Gittikçe rekabetçi hale gelen bu sektörde daha doğru talep tahmini yapmak kritikleşmiştir. Bu amaçla, bu çalışmada farklı ekonometrik ve makine öğrenmesi metotları kullanılarak talep tahmini yapılmış ve bu sektör için doğruluğu en yüksek olan metot belirlenmeye çalışılmıştır. Klasik metotlar tahmin yaparken doğrusal algoritmalar kullanırken, makine öğrenmesi metotları doğrusal olmayan algoritmalar kullanmaktadır. Bu nedenle doğrusal olmayan etkilerin yüksek olduğu verilerde

makine öğrenmesi metotları kullanmak doğruluğu artırmak ve bu doğrusal olmayan etkileri model içerisinde açıklayabilmek açısından faydalı olacaktır. Her iki metot, tahmin doğruluğunu artırmayı amaçlamaktadır. Klasik ekonometrik modellerin temeli matematik ve istatistiğe oturmaktadır. Regresyon, Üstsel Düzleme, Otoregressif Bütünleşmiş Hareketli Ortalama (ARIMA), Vektör Otokorelasyon ve Vektör Hata Düzeltme Modeli (VECM) metotları sıklıkla kullanılan klasik ekonometrik metotlardır. Diğer yandan yapay zekânın bir kolu olan makine öğrenmesi tahmin alanında önemli algoritmalara sahiptir. Özellikle son 10 yılda makine öğrenmesi alanında yapılan çalışmalar ve bu alanda geliştirilen algoritmalar önemli düzeyde artmıştır. Bu artışta, bu algoritmaların ortaya çıkardığı başarılı sonuçların ve pratikte uygulanabilme kolaylığının etkisi oldukça yüksektir. Fakat bu metotların bilgisayar ve bilgisayar bilimlerine fazlaca bağlı olması bu metotların dezavantajı olarak sayılabilir. Ayrıca algoritmaların işleyişindeki kara kutu özellikleri, bu algoritmaların bir diğer dezavantajı olarak yorumlanabilir. Buna karşın klasik ekonometrik yöntemler, uygulamalarının açıklama kolaylığı ile avantaj sağlamaktadır.

Bu çalışmada tahmin amaçlı lineer regresyon ve ARIMA klasik ekonometrik yöntemler olarak kullanılırken makine öğrenmesi metotları olarak Torbalama Regresyon Ağacı, Aşırı Rassal Ağaçlar, Rastgele Ormanlar, Gradyan Artırma Algoritması, Uyarlamalı Artırma Algoritması ve Aşırı Gradyan Artırma Algoritması kullanılmıştır.

Klasik ekonometrik yöntemlerden, yaygınlıkla kullanılan ARIMA modeli uygulanmıştır. Zaman serilerinde tahmin yapabilmek için serilerin durağan olması gerekmektedir. Bu nedenle, serilerin durağanlığını test etmek amacıyla Artırılmış Dickey Fuller (ADF) testi ve KPSS testi uygulanmıştır. Bu testler sonucu durağan çıkan veri setine ARMA modeli uygulanmıştır. Ayrıca veri setinin otokorelasyon ve kısmi otokorelasyon grafikleri incelenerek ARMA modeli parametreleri belirlenmiştir. En uygun parametreler model sonucunda çıkan Akaike Bilgi Kriteri'ne göre karar verilmiştir.

Makine öğrenmesi modelleri, bağımlı değişkeni etkileyebilecek bağımsız değişkenleri kullanarak modellerin kurması açısından ARIMA'dan farklılaşmaktadır. Bu çalışma veri seti hazırlanırken, bağımlı değişkeni etkileyebilecek tüm bağımsız değişkenler veri setine dâhil edilmiştir. Fakat bu açıklayıcı değişkenlerden önemsiz olanlar Spearman, ikili korelasyon, varyans ve rastgele orman algoritması metotları ile belirlenerek veri setinden çıkarılmıştır. Modeller uygulanmadan veri seti eğitim verisi, doğrulama verisi ve test verisi olarak 3 parçaya ayrılmıştır. Veri setinin %20'si test verisi olarak kullanılmıştır. Makine öğrenmesi metotların aşırı uyma probleminin önüne geçmek için çapraz doğrulama uygulanmıştır. Makine öğrenmesi yöntemlerini değerlendirmek için bir örnekleme prosedürü olan çapraz doğrulama, aynı zamanda tahmin performansını da genellemeye yardımcı olur. Bu amaçla k katlı çapraz doğrulama metodu, k = 5 için uygulanmıştır.

İkisi klasik metot, altısı makine öğrenmesi metodu olan 8 farklı metot, test verisinde ortalama mutlak sapma, hataların ortalama karekökü değerlerine göre, tahmin performansları karşılaştırılmıştır. Uyarlamalı Artırma Algoritması, verisinde ortalama mutlak sapma, hataların ortalama karekökü değerlerinde sırasıyla 38.42 ve 57.60 sonuçları ile en iyi performans gösteren metot olmuştur. Bu metodu Gradyan Artırma Algoritması takip etmiştir. Klasik metotlardan Lineer Regresyon metodu ortalama bir performans gösterirken, ARIMA metodu en vasat performansın sahibi olmuştur.

1. INTRODUCTION

The main purpose of the companies is to make a profit. Companies needs to fulfill the their main activities to reach that aim. In textile retail sector, organizations works to have the right product, at the right time, in the right store, at the right amount, and at the right price in order to make profit. Therefore, it is important to avoid stock-out or overstock by meeting the specified criteria. At that point, sales forecasting becomes prominent for such companies. By the help of sales forecasting, companies be able to define optimum stock level and direct their operational functions from production to logistics.

In fashion retail sector customers' personal tastes varies highly which makes sales forecasting insistent topic (Allenby et all., 2014). There is a lot of product variety in this sector and the vast majority consist of new products that have not shown sales performance previously. Furthermore, sales of this sector is significantly affected by many factors such as climate conditions, precipitation and temperature conditions, special days, stock mix and economic environment (Ni & Fan, 2011). All of this factors induce rapid changes in sales. To illustrate, people's demand for clothing increases during religious holidays and result in significant increase in sales. In addition, economic imbalances and uncertainties can affect this sector rapidly, causing sales to decline.

Today, most of the market becomes more competitive. Furthermore, globalization makes most of the function of the company most important such as supply chain, distribution due to prolonged delivery times. Therefore, this study purposes to solve the problems with sales forecasting of fashion retail sector by developing methods and models. In the scope of this study, large variety of variables, which are taken into consideration, have important effects on sales performance. Most of the studies in that field exercises with the limited number of variables. In addition, different machine learning algorithms and classical econometric methods are utilized to establish most accurate models for sales forecasting in desired levels. Also, different combination of these algorithms can be used to get new models by the covering disadvantage of each

methods.

Xia et al. (2012) conducted a study on sales forecasting and expressed that there are large variety of sales forecasting methods lately. Two classes that are addressed for sales forecasting classical methods. The first one relying on mathematical and statistical models such as regression, ARIMA, exponential smoothing. The second class is utilized artificial intelligence methods like ANN. Makridakis et al. (1998) indicated that most of the times, dealing of linear models of classical methods are not sufficient to satisfy the fashion retailing forecasting problems.

Markidakis (2018) conducted a study on ML methods and statistical methods. It is stated that the ML methods and statistical methods aim same goal that is increasing accuracy of forecasting with reducing sum of squared errors. The dissimilarity of ML methods and econometrics methods is that reducing errors are accomplished by non-linear algorithms and linear algorithms by ML methods and econometrics methods, respectively. ML methods need computer science more than statistical methods. Therefore, ML methods, which is positioned at the concurrence of computer science and statics, necessitate greater dependence on computer science to be accomplished.

Çatal et al. (2019) performed a study on sales forcasting on Walmart, which is one of the biggest retail company. Researchers investigates the comparision of regression and time series methods. According to this study, they figured out those regression methods provide better performance than time series analysis techniques. Moreover, the best predictor for sales forecasting is Boosted Decision Tree Regression algorithm and it does not necessitate much optimization to render high performance.

Jain et al. (2015) stated the results for sales forecasting of retail chains. This research study includes the sales forecast of Rossmann retail firm. The Rossmann store managers' duty is that estimating the daily sales making the problem vital. Therefore, researchers performed the different data mining techniques to predict sales at stores. This study draw attention to XGBoost that accomplish the best prediction when the result of it compared with the other methods like linear regression and random forest regression. In addition, wide range of variables are utilized from store to customer and special days to socio-geographical structure.

Makridakis et al. (2018) performed study on machine learning and they indicated that over the last decade, studies on Artificial intelligence (AI) have increased significantly. Utilization of Artificial Intelligence provides qualified practices in various fields. The achievement of AI can be explained by the utilization of algorithms that are develop

their success over time by the help of learning from trial and error. It is not only the through step-by-step guidance of code, but also usual programming domain based on logic like decision trees, if then rules. Given the fact that mentioned success of AI, significant amount of study executed on how remarkable part of it, using machine learning methods (ML), can be used to advance time series forecasting. In fact, methodocigal progress and enhancement in accuracy are proposed by the hundreds of research papers with the new ML algorithms.

By virtue of above analysis, machine learning methods such as bagging regression tree, extremely randomized trees, random forest algorithm, gradient boosting, AdaBoost, extreme gradient boosting algorithm are presented for fashion retailing forecast in this paper. Moreover, mostly used classical forecasting method being ARIMA is also utilized to forecast. This paper aims not only help to understand and apply of ML methods of such complex forecasting aims but also offers opportunity to confront the results of modern forecasting methods with classical method. The literature review and methodology parts where fundamental principles of the utilized methods are acquaint, data description, the result and discussion of empirical results.

2. LITERATURE REVIEW

The global textile and clothing industry is about USD 920 billion in 2018. The value of the leading textile exporters worldwide in 2018 by country is shown in the Figure 2.1. Turkey is one of these countries with the USD 12 billion in 2018.

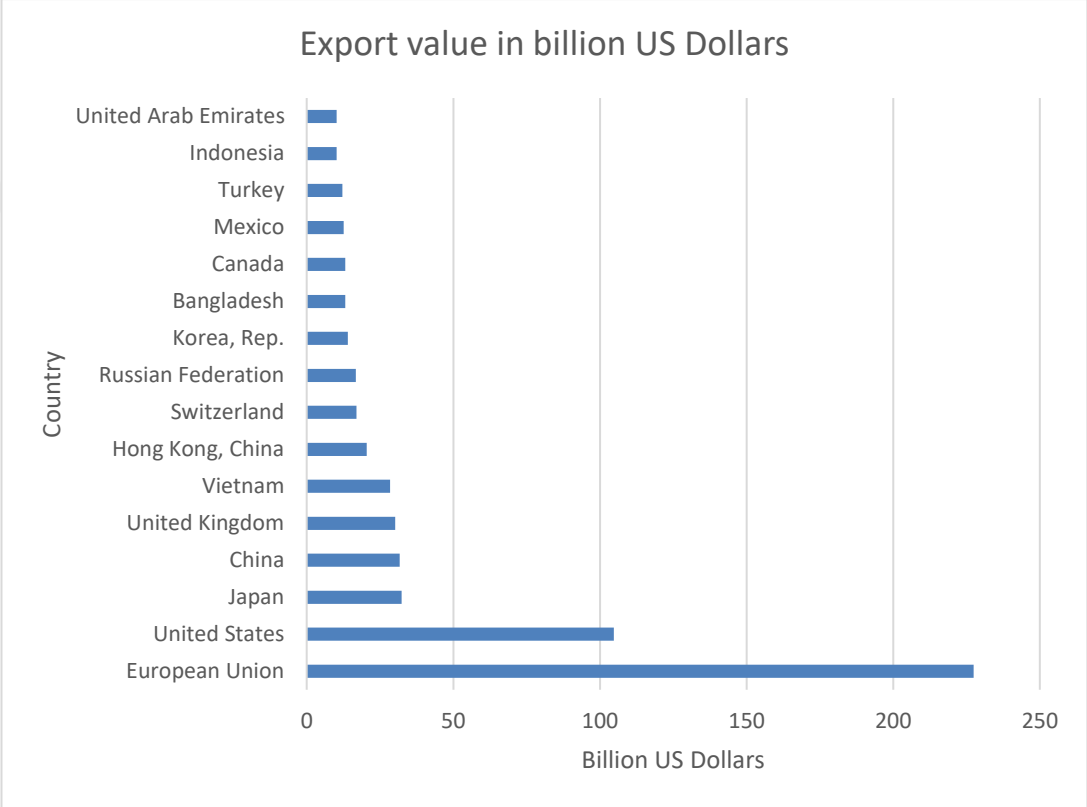


Figure 2.1 : World textiles and clothing exports by country 2018.

Source: WorldBank World Integrated Trade Solutions, 2018.

There are three markets that are specified in the textile sector are appeal, home textile and technical textile. Appeal textile comprise of the clothing of women, men and children. Thomassey (2010) investigates items that are sold under the category of appeal into three categories according to their sales potential and fashion dependency as following:

– Basic needs of consumer such as black t-shirt and sold all year or season are called basic items.

– Items that are highly depend on seasonal fashion trends and have short life are called fashion items. That are generally sold for one season and not reproduced at future seasons.

– Items that are sold for more than one season with the little alterations at different seasons according to trends. These items are names as best selling items. To illustrate, girls legging with different flowers patterns.

The big part of sales are provided by basic and best-selling items in the most appeal textile retail companies. Therefore, forecasting does not perform on fashion items(Thomassey, 2010). Although the historical sales of basic items are available, it is not available for the bestselling items. It is hard to predict the new items with modifications according to its original one because consumer's tastes can react differently for new item.

The appeal textile sector includes high number of items and SKUs at each seasons. The aggregation of data is necessary at different product level due to the high number of SKUs, the short life span of items and lack of references of new or modified items. Therefore, aggregation of data at the right level is important for sales forecast (Thomassey, 2010). Thomassey (2010) introduces the commercial product hierarchy in the clothing industry as Figure 2.2. In this study, the sales forecast is performed at style level.

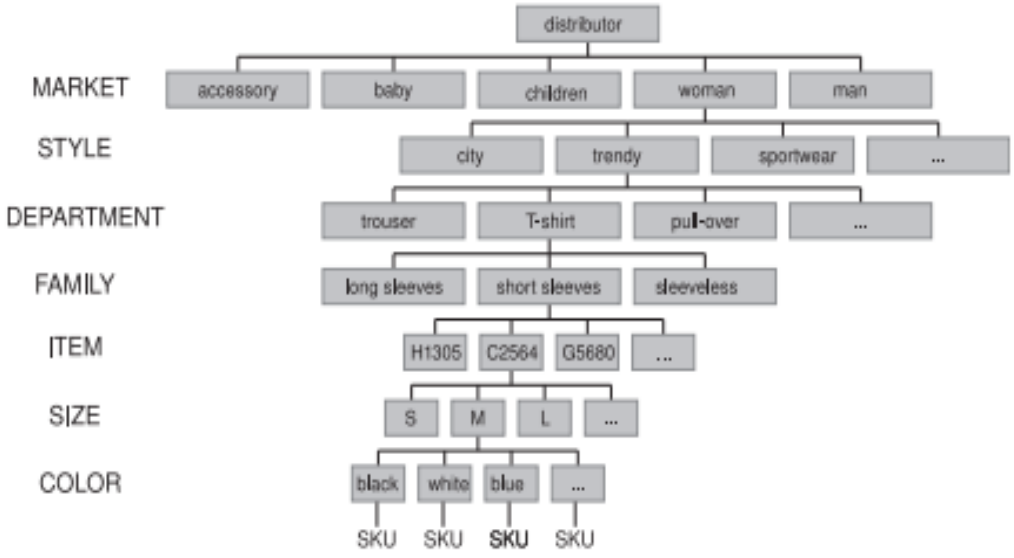


Figure 2.2 : Commercial product Hierarchy in clothing industry (Thomassey, 2010).

Due to having high changefulness and rely on too many factors, sales forecasting is compelling for all times. From a scientific point of view, the history of the sales forecasting dates back to about 60 years ago (Boulden, 1958).

Sales forecasting can be defined as estimating future sales based on historical sales data. Demand forecasting plays an increasingly substantial role in a business enterprise decision support system due to competition and globalization. The ideal sales forecasting for fashion retail company is that forecasting at the level of day, store and stock keeping unit (SKU). Because fashion product groups consist of options being at high numbers and leading to many distinctive SKU. Moreover, the company's internal structure, product characteristics and external factors hardly affect the sales and this makes sales forecasting difficult topic. For instance there are lots of factors that cannot be controlled by companies such as holidays, climate and economic conditions, trends, activity of marketing, campaigns influencing on sales highly (Liu et al, 2013). Moreover, sales performance of a product is highly interconnect to consumers personal pleasure which is differentiate broadly from person to person. Besides, past sales data is not reachable for a clear majority of items due to having short life cycle and many new products are released (Choi et al, 2011). Therefore, improving more reliable forecasting models and methods is getting a significant investigation subject.

Manufacturing and stock cost and selling price calculations by the decision makers can

be more efficaciously performed by the assist of sales forecast. More than that, the target of just-in-time distribution is replied quickly and accomplished with optimum inventory levels. Over stock and out-of-stock cost can be eliminated by the aid of appropriate forecasting method and models (Sun et al, 2008). Organizations' important functions such as sales, marketing, planning and finance needs forecasts for make a decision in both short and long term. Furthermore, for the different level of dispensation and allocation plan such as national and regional level, forecasts ensure base (Fildes et al, 2019). All of this reasons make sales forecasting is a crucial part of study at fashion retail sectors. Most of the developed fashion retail companies allocates substantial resources for sales forecasting. They work to integrate sales forecasting in the system they use in many fields from procurement to price determination and distribution.

Linear statistical methods such as ARIMA, became not sufficient to solve the real world's problems in the early 1980s. In the past two decades, machine learning models have drawn attention to be a serious competitors to classic statistics models as in the prediction community. Nonparametric nonlinear models that find out the stochastic dependence by utilizing the past data named black-box or data-driven models (Mitchell, 1997).

The theory of statistical forecasting suggest that an examined sequence is a particular realization of a random process with randomness derives from the many linearly interacted independent degrees of freedom (Anderson, 2011). In dynamic systems theory, nevertheless, haphazard action can be caused by the deterministic systems with only a low number of degrees of freedom is the emerging view, sympathizing nonlinearly (Farmer & Sidorowich, 1987). This aperiodic and confused behavior is also named as deterministic chaos (Schuster, 1988).

In these studies classical time series and some machine learning methods are utilized to forecast the sales in a fashion retail company and compare the method according to some forecasting evaluation criteria. Armstrong and Yokuma (1995) stated the criteria for selection of a forecasting method. According to the survey results, where 94 researchers, 55 educators, 133 practitioner and 40 decision makers participated, accuracy and timeliness in providing forecasts emerges the most important ones. All criteria that are investigated in this study is figured out at Figure 2.3. According to the literature review, the following methods that are applied to sales forecast are ARIMA,

Bagging Regression Tree, Extremely Randomized Trees, Random Forest Algorithm, Gradient Boosting Algorithm, AdaBoost and Extreme Gradient Boosting Algorithm.

One of the object issue is the using the linear or nonlinear methods. It is stated that it hinge on the problem. The properties and response have linear relationship approximately, the linear model provides better solution then decision trees. However, decision tree perform better if nonlinear and complex structure are observed between properties and response pointed out by the model (James et al., 2013).

	Mean Importance Rating (number responding)				
Criteria	Researcher (94)	Educator (55)	Practitioner (133)	Decision Maker (40)	Average (322)
Accuracy	6.39	6.09	6.10	6.20	6.20
Timeliness in providing forecasts	5.87	5.82	5.92	5.97	5.89
Cost savings resulting from improved decisions	5.89	5.66	5.62	5.97	5.75
Ease of interpretation	5.54	5.89	5.67	5.82	5.69
Flexibility	5.54	5.35	5.63	5.85	5.58
Ease in using available data	5.59	5.52	5.44	5.79	5.54
Ease of use	5.47	5.77	5.39	5.84	5.54
Ease of implementation	5.24	5.55	5.36	5.80	5.41
Incorporating judgmental input	4.98	5.12	5.19	5.15	5.11
Reliability of confidence intervals	5.09	4.70	4.81	5.05	4.90
Development cost (computer, human resources)	4.70	5.02	4.83	5.10	4.86
Maintenance cost (data storage, modifications)	4.71	4.75	4.73	4.72	4.73
Theoretical relevance	4.81	4.20	4.43	3.72	4.40
Ability to compare alternative policies	–	–	–	–	–
Ability to examine alternative environments	–	–	–	–	–
Learning	–	–	–	–	–

Figure 2.3 : Criteria for selection of forecasting methods (Armstrong & Yokuma, 1995).

Over the past decade, machine learning becomes a key component of many industries' services. The base of the machine learning consist of the learning from the past data and utilize the date to forecast for future. There is omnipresent cost constraint makes automation necessary and technological advances which have revolved data streams into fire hydrants. Machine learning algorithms are utilized at law enforcement, employment decisions, admissions, credit scoring, social networks, search results, and advertising ordinarily (Roelofs, 2019). Machine learning facilitate multifarious of applications such as health analytics, image processing, language processing,

computer vision, advertising among other makes it a significant technology. Moreover, robustness, reliability and understandability of the machine learning algorithms are essential owing to influencing high number of systems and people.

Supervised and unsupervised learning are the subdivision of machine learning algorithms. The methods depending upon training a data sample from data source with already appointed correct classification calling supervised. The methods where unmarked training data is utilized to categorize unseen patterns is called as unsupervised data (Puica & Florea, 2013).

One of the weakness of the machine learning models is being highly susceptible to little diversification of data dispersion. The generalizing or handling of unobserved or unusual states ability of the algorithm is delayed by this susceptible to data dispersion. The deprivation generalization of the algorithm causes to make wrong vote with some the strict outcome. It makes the algorithm defenseless to assault by the abuseful people by the substitute the underlying data with the way of indiscernible methods. Therefore, easy handling of particular predictions executed by the algorithm can be performed by the hostiles. Arduousness of the machine learning is that exposition of it. Exposition arduousness of machine learning algorithm is blind side of it. Inescapable fault of machine learning algorithms cause to confrontation of describing the reasons of fault rising from data or system and fixing the problem. Particular decisions of machine learning algorithms is not easily understood, even amongst the experts in machine learning experts. Comparatively, with the little counseling of the theory, driven by empirical progress determines the achievement of the machine learning (Roelofs, 2019).

Machine learning brings together two important terms which are generalization and overfitting. Generalization is defined as the ability of expressing the new cases that have not been met before datasets. In general terms overfitting is utilized the express the undesirable decrease in performance (Roelofs, 2019). Studies researching decision tree induction indicated that success on training data can be a deceptive predictor of true predictive precision. A complex tree with high accuracy in a training set gets worse than a simple tree with less flashy performance in fresh data. This situation is defined as overfitting the training data with the complex tree by the researchers. It means that models represent both the inherent relationships and figures emerging by chance. In this sense, much effort has been made to develop over fit reducing methods for tree

induction, usually in the form of the pruning strategies, as overfitting reduces estimation accuracy (Schaffer, 1993).

Cross validation which is exploited to deal with the generalization and over fit error standard statistical technique. The usable dataset divided into parts that are called as training and validation or test in order to use at learning proceses and performance evaluation, respectively (Karystinos & Pados, 2000).

To illustrate, there have been some essays with the cross validation method restriction in order to develop the generalization ability of multilayer perceptron. The first one which is addressed by Akaike (1992) is using Akaike's intelligence measure. Rissannen (1978) stated the minimum description length a criterion as second. Linear systems theory such as autoregressive or moving average models are prompt these criteria being linear in the some free network arguments.

3. METHODOLOGY

As mentioned before, this study examines the sales forecasting of fashion retail company with classical and machine learning methods which are ARIMA, Bagging Regression Tree, Extremely Randomized Trees, Random Forest Algorithm, Gradient Boosting Algorithm, AdaBoost and Extreme Gradient Boosting Algorithm. In order to perform the forecasting initial dataset with historical sales data is prepared and data pre-processing is applied. Dataset is splitted into train and test data. Then, forecasting methods are performed with Python. Finally, models formed from different methods are compared with performance measurements. Flow chart of applied methodology is shown in Figure 3.1

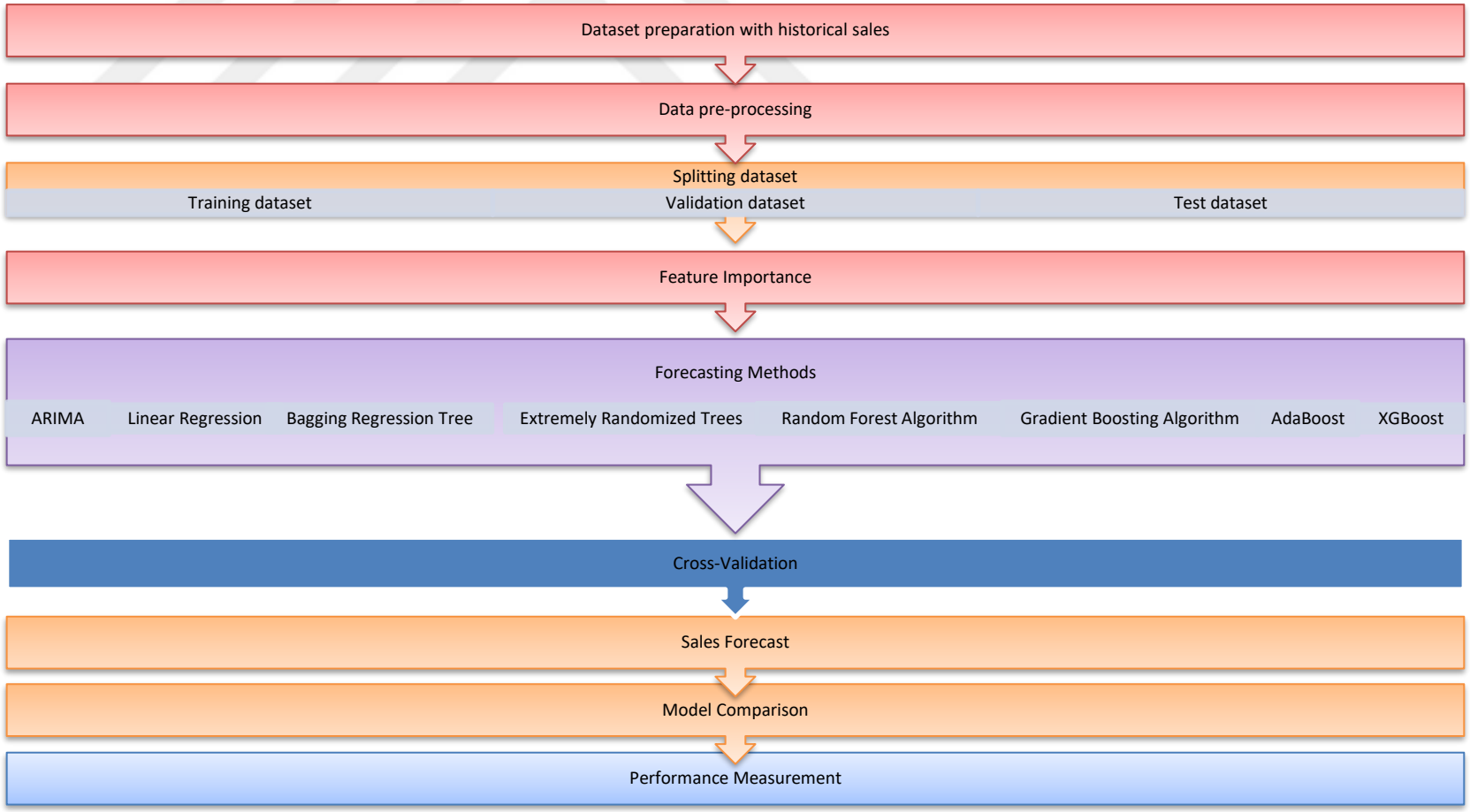


Figure 3.1 : Flow chart of applied methodology.

3.1 Autoregressive Moving Avarage Model

ARIMA models that are approved as one of the miscellaneous parts of models for estimation of time series many disparate kind of stochastic seasonal and non-seasonal time-series can be substitute of them. ARIMA needs stationary data to be utilized with the parts of autoregressive (AR), moving average (MA) processes (Oliveira & Ramos, 2019). Autocorrelations are purposed to explain with ARIMA. Iterative three stages transform of model choice is contained at ARIMA models whose framework is devised by Box et al (2015). The Box–Jenkins methodology which is a useful functional time series modeling method has been proved. ARMA (p, q) which is a model for stationary time series and defined as following (Shumway & Stoffer, 2017). x_t (t = 0, ±1, ±2,…) is a time series and the autoregressive and moving average orders are represented by p and q, respectively.

$$x_t = \emptyset_1 x_{t-1} + \cdots + \emptyset_p x_{t-p} + w_t + \varphi_1 w_{t-1} + \cdots + \varphi_q w_{t-q}, \quad (3.1)$$

where $\emptyset_p \neq 0, \varphi_q \neq 0$

$\emptyset_1, \emptyset_2,\ldots, \emptyset_p$ are constants at AR process while $\varphi_1, \varphi_2,\ldots, \varphi_p$ are constants are constants at MA process.

Steps of building an ARIMA model can summarized as below.

- Check the stationary of time series
- Identify the ARIMA model by using ACF and PACF
- Predict ARIMA model parameters
- Choose most suitable ARIMA model by using Akaike Information Criteria.
- Forecast

ARIMA models presents with the orders (p, d, q). Degree of autoregressive process (AR), moving average process (MA) and stationarity of time series are represented by the orders p, q, and d, respectively. Time series must be stationary to apply these forecasting methods. When the time series is stationary, no need to take the seasonal difference and d become zero. Therefore, ARIMA model reduced to ARMA. When

p=0 then model is reduced to MA process. When q=0 then the model reduced to AR process. Non-stationary time series must be evolve the stationary via taking seasonal difference with using ARIMA model. Stability is about the AR part. When the MA part equation has roots inside the unit circle with absolute values smaller than one, it is invertible. If series does not satisfy the being invertible ARMA model cannot be applied. If it is invertible ARMA process can be applied by selecting the best model using.

Stationarity of dataset must be tested before applying ARMA models. The presence of a unit root of the autoregressive polynomial is tested so as to determine whether or not the time series should be differenced. ADF and KPSS tests are utilized to test stationarity.

The ADF test investigates the null hypothesis of an nonstationary ARIMA (p, 1, 0) process while the altervative hypothesis is stationary ARIMA (p+1, 1, 0), setting an autoregressive AR model. In short, the null hypothesis is series is nonstationary, in other words, there exists unit root in series. According to the result of ADF test of daily store sales of fashion retail company, the null hypothesis is rejected. That means, our time series is stationary. Therefore, ARMA model can be applied. The results of ADF test is shown at Table 3.1.

Table 3.1 : ADF Test result

ADF Statistic	-3.21
p-value	0.019
Critical Values	3.44*
	2.86**
	2.57***

*** 1 %, ** 5% and * 10% significance level.

In the ADF test, the null hypothesis tests the presence of unit root in time series. On the contrary of ADF test, null hypothesis of the KPSS test tests the absence of unit root. Beveridge-Nelson decomposition from a basis for KPSS test (Neusser, 2016). According to the result of ADF test of daily store sales of fashion retail company, the null hypothesis cannot be rejected which means series is stationary. KPSS test result is figured out at Table 3.2. Both ADF test and KPSS test results are compatible with each other. Therefore, our time series is stationary.

Table 3.2 : KPSS Test result

KPSS Statistic	0.14
p-value	0.062
Critical Values	0.12*
	0.15**
	0.22***

*** 1 %, ** 5% and * 10% significance level.

ACF and PACF utilizes to determine ARIMA model's Autoregressive (AR) and moving average (MA) process order. Also, this graphs help to investigate stochastic process.

ACF is an autocorrelation function and gives information about relation of existing values with its past values. ACF considers the components of time series such as seasonality, trend, cyclic and residuals. ACF plot gives idea about the MA process.

Partial autocorrelation function is represented by PACF. It figures out the residual correlations with the next lag value. Therefore, it is called partial while ACF called complete autocorrelation function. PACF helps to keep next lag as a future while modelling if there is any embowered information in the residual being modeled by next lag. PACF informs about AR process.

ACF and PACF are plotted for investigated data at daily level in order to figure out the relation of current and past values as explained before. It can be deduced from the ACF plot that there is weekly seasonality due to having relatively high correlations for 7, 14, 24 etc. order lag as shown in Figure 3.2.

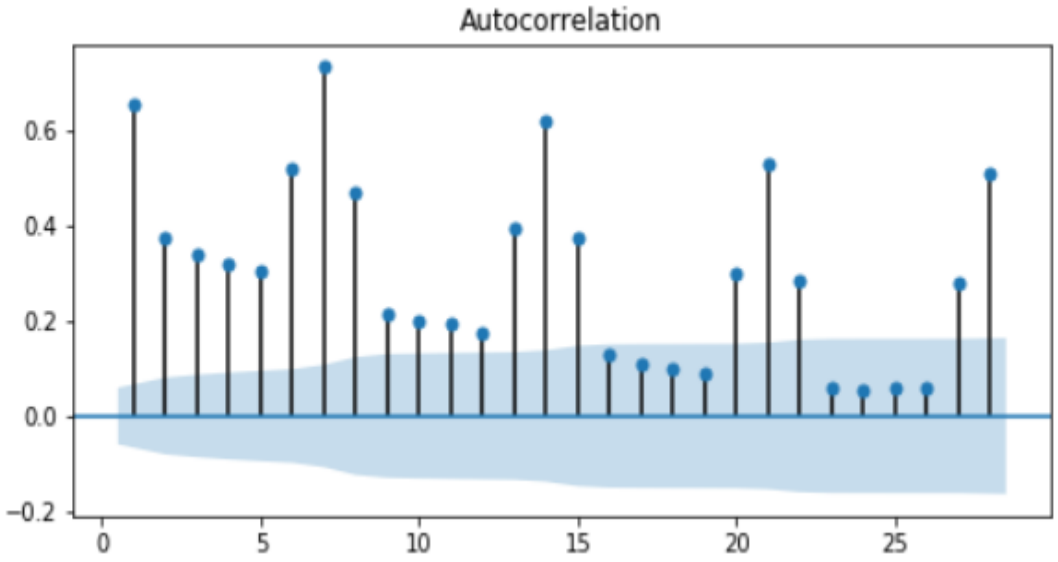


Figure 3.2 : ACF of daily sales.

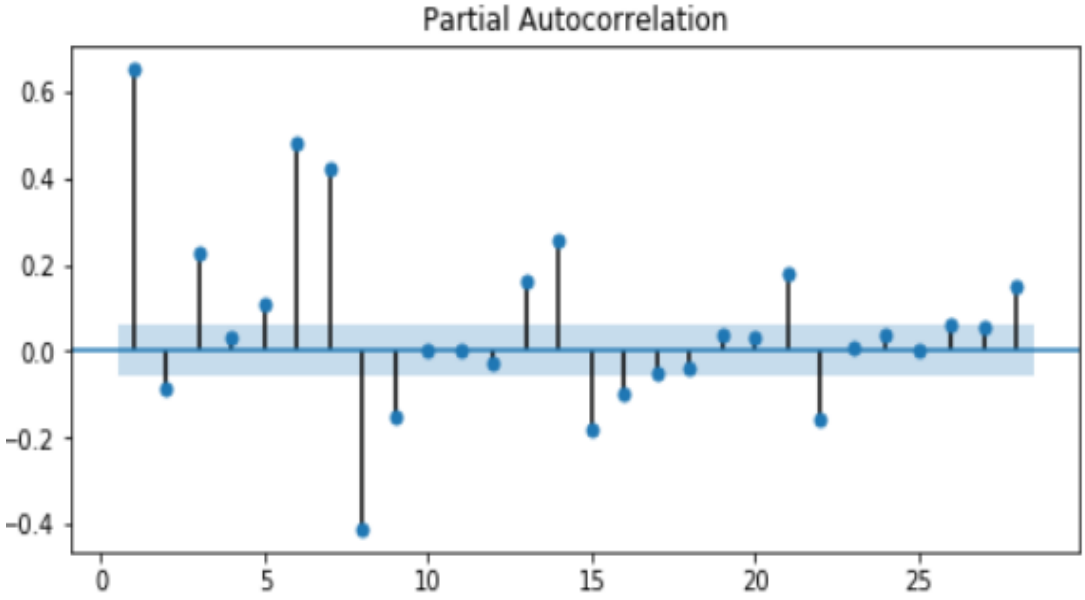


Figure 3.3 : PACF of daily sales.

The order selection of ARMA (p, q) model aims to minimize information criterion at particular values of p and q. The consecutive increment of the orders of p and q raises the fit of model as variance of residuals continual declines with the fixed number of observations. Overfitting tendency is indemnified by the punishment (Neusser, 2016). Akaike Information Criteria (AIC) is one of the most common information criteria and it is utilized in this study while selecting the AR and MA order of (p, q). The model with the minimum AIC is result with best fit. ARMA (p, q) process is used in this study

with the daily sales with two different (p, q) these are (3, 6) and (1, 2). The formula of AIC presented as following (Neusser, 2016).

$$AIC(p,q) = ln\sigma_{p,q}^2 + (p+q)\frac{2}{T} \quad (3.2)$$

3.2 Bagging Regression Tree

The usage of identical training algorithm for every predictor and treat them on particular random subsets of the training set to yield variety of classifiers. Bootstrap aggregating (bagging) is the method where samples are created by the replacement. Along the multiple predictors, bagging provides samples to be epitomized for several times. Moreover, samples of the training to be epitomized for several times for the identical predictor. Bagging tree regressor aggregates the single predictions from the different decision trees by meaning the definitive prediction. Because of meaning the different models, prediction from the bagging has smaller variance then single predictions (Gihimire et al., 2012).

3.3 Random Forest Algorithm

One of the most popular machine learning forecasting algorithm is the Random Forest Algorithm. Random forest which is exercised via bagging method usually is an ensemble of decision tree. Farnaaz & Jabbar (2016) conducted a study on Random forest tree where accretion provides supernumerary randomness via examining for the best characteristic through the random subset of characteristics in place of very best characteristic when dividing a node. The algorithm works out major tree variety that cause a higher bias for a lower variance, usually provides an overall better model. In random forest algorithm low variance is achieved via the both choosing a random sample of the data and random subset of the properties to grow decision tree. The significance of each predictor can be received from the random forest regressor by sorting the predictors conformable to the diminishing in residual sum of square owing to separating over the stated predictor over the all trees like bagging tree regressor. Random forest comprise of lots of decision trees. Nodes are separated by utilizing the some parameters such as tree count and characteristics. Random forest algorithm advantages can be listed as following.

– Trees that are created during before modelling can be used for fortune reference.

– Over fitting problem is minimized by the random forest algorithm.
– Accuracy and feature selection according to significance is performed by random forest algorithm.

3.4 Gradient Boosting Algorithm

Some feeble learners unified into powerful learner is called as boosting which mean to any ensemble method. In boosting methods, predictors are trained consecutively, each attempt to fix its predecessor (Géron, 2017). Boosting algorithm utilizes the all training data set for regression instead of resampling some part of training data contrary to bagging and random forest. During the first iteration training samples are specified same weights, then the weights are redistributed respectively according to the compatibleness the training data in respective iterations of the regression tree (Gihimire et al., 2012). There are some boosting algorithms are available. In the scope of this study Gradient Boosting, Adaptive Boosting and Extreme Gradient Boosting algorithms are used.

Gradient Boosting is a machine learning method for both regression and classification. Weak prediction models are used to generate an estimation model by using decision trees. Furthermore, it can be defined as the optimization algorithm for appropriate cost function. Gradient Boosting method unify the weak learners to obtain a unique strong learner with iteration (Jain et al., 2015). Friedman (1999) conducted that Gradient Boosting is obtained by the addition on the regression models with accommodation of the basic parametrized function to present residuals by least square at each iteration. Loss functional is diminished where pseudo residuals are the gradient, corresponding with the model values at each measured current step at training data point.

3.5 Adaptive Boosting Algorithm

Freund & Schapire (1996) was proposed the Adaptive Boosting Algorithm. They suppose that weak learner creates a hypothesis that has little error regarding the distribution. Boosting algorithm creates the next weight vector via this new hypothesis with n times iterations. It is end up with the final hypothesis utilizing a weighted majority vote. In essence, AdaBoost forms a single strong learner by summing their weak learners' probabilistic predictions.

Former learners helps to intensify on the misclassified observations at AdaBoost. This is achieved via base learner enhancing the importance of misclassified investigations and changing adaptly at the training data (Jiang et al., 2019).

3.6 Extreme Gradient Boosting Algorithm

XGBoost that is utilized to estimate the target variable via the usage of training data is a one of the supervised method. The complication of the model is controlled by the regularization keeping away from overfitting. A group of classification and regression trees (CART) are used by XGBoost. Family members on data are grouped into disparate leaves and they are scored at comparable leaf. Dissimilarly from decision trees, in addition to the decision value, an actual comes from each leaves in CART. This provides deeper definitions above classification. It is formed from 3 steps defined as following (Jain et al., 2015).

1. Additive Training: Parameters of trees are described at that phase. Those functions hold data about the tree construction and leaf score. Using an additive optimized approach: learned thing are pinned and combined at new tree also.
2. Model Complexity: It helps to parameter regularization and invoke a penalty against specific situations.
3. Structure Score: Finest split condition can be predicted via structure score by considering complexity. The greater collision on the purity is belongs to the first split of the tree and subsequent splits focuses on the little parts of dataset being the first tree classifiers.

3.7 Extremely Randomized Trees Algorithm

Extremely randomized trees is one of the supervised regression and classification method. Also, it is a type of bootstrap aggregation in ensemble learning methods. Geurts et all. (2005) is explained that algorithm by dividing into randomized and extreme parts. When splitting a tree node, both attribute and cut-point selection makes heavy randomized. It makes totally randomized trees whose structure are independent of the output values of learning sample in the extreme part. Moreover, they suppose that Extremely Randomized Trees can help to decrease bias and variance while decision trees have high standard deviations of the errors.

4. DATA DESCRIPTION

Within the scope of this study, sales of fashion retail company's are forecasted. This company is one of the biggest fashion retail company in the Turkey with hundreds of stores in domestic and abroad. As mentioned previous parts, SKU or item level prediction is very difficult because of most of the items sold are new with no historical sales data. Therefore, product level prediction is exercised at style level which is the described at previous parts. The dataset is a time series data where sales data is collected at daily level for 30 stores and 2 styles. Forecasting is performed for each stores' each styles' at daily level sales quantity. Therefore, the dependent variable of this study is store-style daily level sales. Dataset includes data from January 2017 to February 2020. Figure 4.1 shows the daily store sales quantities for the given time interval. Descriptive statics of the dataset is represented at Table 4.1.

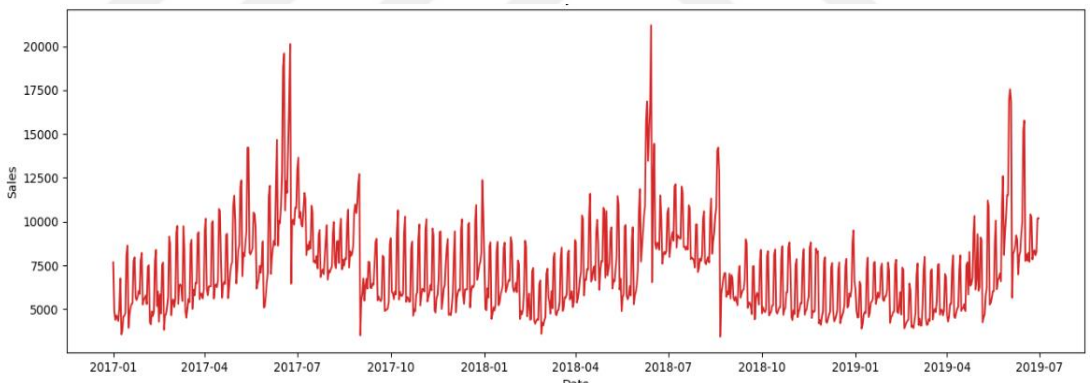


Figure 4.1 : Daily store sales

Table 4.1 : Descriptive statics.

	Sales	Day Sales	Month Sales	Day All Store Sales	Day Store Style Sales	Last Year Stock	Last Year Sales	Lagged Sales
Count	36,440	36,440	36,440	36,440	36,440	36,440	36,440	36,440
Mean	181	7,243	229,364	4,057	3,625	3,843	187	181
Standard Deviation	109	2,495	53,147	2,520	1,572	1,646	114	109
Minimum	11	3,437	126,333	1,236	1,111	518	11	11
25%	102	5,365	192,290	2,346	2,496	2,659	104	102
50%	160	6,689	231,068	3,332	3,378	3,589	164	160
75%	235	8,563	260,154	5,046	4,537	4,839	243	235
Maximum	1,269	21,201	328,293	33,082	11,734	12,289	1,367	1,269

5. RESULTS

5.1 Forecasting Error Measures

There exist some forecasting error measures that are widely used to assess the accuracy of forecasting models. Moreover, these are useful to compare the different forecasting models between each other. Actual observation for time t and forecasted value are represented by y_t and $\hat{y}_t$, respectively. These accuracy measures are performed at the test data, which is out of sample data. Total number of observations are denoted by m while number of test data observations are denoted by n where m > n. Mean error (ME), the mean absolute error (MAE) and the root mean squared error (RMSE) which are usually applied scale – dependent statics specified as below (Peña et al., 2011).

$$ME = \frac{1}{n-m} \sum_{t=m+1}^{m} (y_t - f_t) \tag{5.1}$$

$$MAE = \frac{1}{n-m} \sum_{t=m+1}^{m} (|y_t - f_t|) \tag{5.2}$$

$$RMSE = \sqrt{\frac{1}{n-m} \sum_{t=m+1}^{m} (y_t - f_t)^2} \tag{5.3}$$

As seen above MAE is simple to understand when it is compared with RMSE when comparison of the performance of forcasting methods on the same data set. Infrequent major errors can be more easily detected by the RMSE which result in irregular weight to very large errors makes more worthful. According to the evaluated variable and forecasting accuracy, RMSE and MAE change and there is no complete standard for these values (Peña et al., 2011).

As is seen from above formulas MAE and RMSE are size dependent statics which makes cause disadvantages at different size of samples. Thus, size independent statics based on percentage values provide an advantage to measure accuracy and they are

mostly utilized to evaluate the performance of forecasting methods at different data sets. Mean percentage error (MPE) and the mean absolute percentage error (MAPE) are the conventional methods that are utilized and formulated as below (Peña et al., 2011). The disadvantage of percentage error measures are addressed as being $y_t = 0$ and cause to be infinite of undefined for ant t in the time series and causing to zero value when any y_t is very big. Often, when the performance of forecasting methods are evaluated to compare, accuracy metrics result in contrary results (Ramos et al., 2014). In our study there are some zero values while some observations are very high. Thus, this performance measures are nor suitable for this study.

$$MPE = \frac{1}{n-m} \sum_{t=m+1}^{m} \left(\frac{y_t - f_t}{y_t} \right) \times 100 \tag{5.4}$$

$$MAPE = \frac{1}{n-m} \sum_{t=m+1}^{m} \left| \frac{y_t - f_t}{y_t} \right| \times 100 \tag{5.5}$$

5.2 Variable Importance

Machine learning methods utilize the independent variables that have effect on the target variable. Independent variables used in machine learning methods directly affect the performance of these methods. In this study, both numerical and categorical variables are included in dataset. In univariate analysis, mean, median, minimum, maximum and standard deviations of each numerical variable are investigated. Correlations between numerical variables were investigated with the Spearman method. According to the results of the Spearman correlation analysis sales values at different level features are highly correlated with each other. Figure 5.1 represents the results of Spearman correlation analysis

Variable importances are evaluated by Pairwise correlation, variance and feature importance with random forest algorithm. Most important varibles are shown in Figure 5.2. Less important variables are removed from data set and and then machine learning methods are applied.

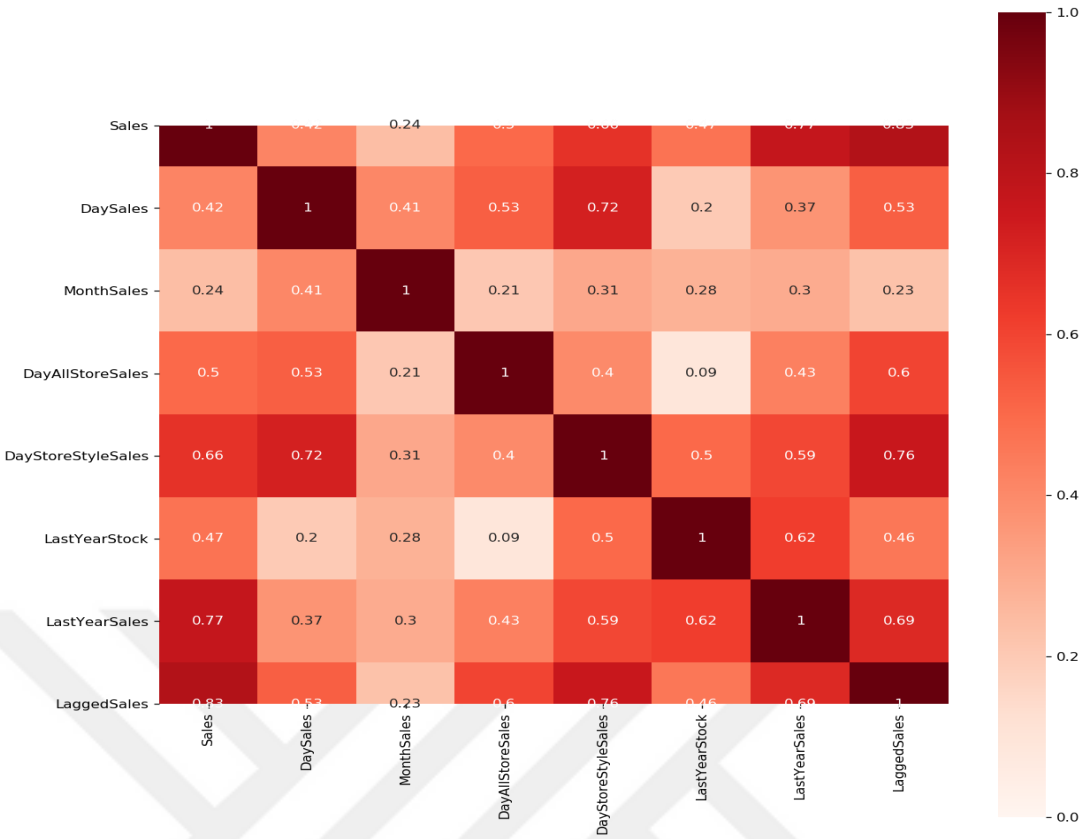


Figure 5.1 : Spearman correlation analysis.

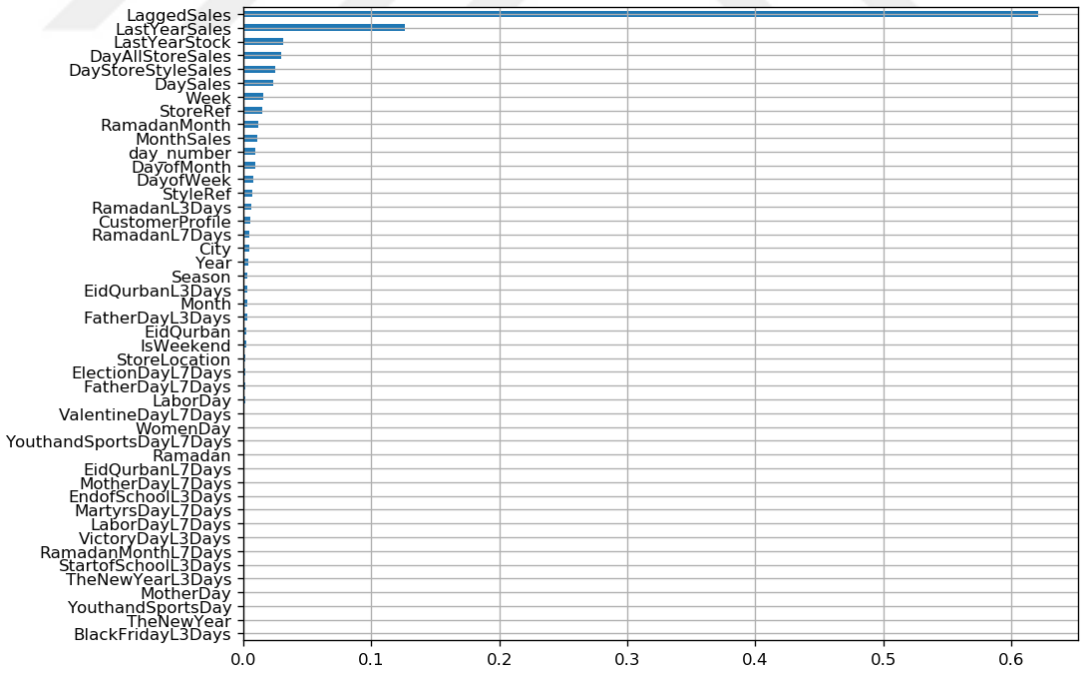


Figure 5.2 : Variable importance from random forest algorithm.

5.3 Model Comparison

ARIMA, Bagging Regression Tree, Extremely Randomized Trees, Random Forest Algorithm, Gradient Boosting Algorithm, Adaboost and Extreme Gradient Boosting Algorithm are utilized to forecast fashion retail daily sales. Models are established with historical data set and tested.

Cross validation is a resampling procedure in order to evaluate the machine learning methods and helps to generalize performance. K-Fold cross validation method is used to prevent biased prediction of the model performances. K that is the parameter of this method determines the number of groups that the data sample is to be split into. Low k values can cause to biased results while the high k values can suffer from high variability. Therefore, 5-fold cross validation is applied.

Model results without cross-validation are summarized at below Table 5.1. XGBoost and Gradient Boosting algorithms result with lowest MAE, RMSE and highest R^2 values where it is very close to each other. Linear regression model also has very close results to other machine learning model results. However, as seen from results ARMA model has lower performance at all performance measures when compared machine learning algorithms and linear regression.

Table 5.1 : Model performances without cross-validation.

Model	MAE	RMSE	R2
XGBoost	29.63	43.5	0.83
Gradient Boosting	35.86	43.77	0.83
Extremely Randomized Trees	35.86	52.05	0.76
Bagging	35.88	52.41	0.76
Random Forest	35.88	52.4	0.76
Linear Regresssion	36.23	53.57	0.75
AdaBoost	37.14	52.91	0.75
ARMA	81.86	92.73	0.30

Model results with cross-validation are summarized at below Table 5.2. AdaBoost and Gradient Boosting algorithms result with lowest MAE, RMSE and highest R^2 values where it is very close to each other. While XGBoost has best performance results without cross validation, its performance decreased after cross-validation. This result shows the importance of applying cross-validation to machine learning methods. When they are compared, it is obvious that without cross-validation, results are optimistic.

Table 5.2 : Model performances with cross-validation.

Model	MAE	RMSE	R2
AdaBoost	38.42	57.6	0.72
Gradient Boosting	38.79	58.27	0.71
Extremely Randomized Trees	40.62	62.32	0.69
Random Forest	41.16	62.92	0.70
Linear Regresssion	41.18	62.95	0.70
Bagging	43.4	65.12	0.69
XGBoost	43.44	65.17	0.69
ARMA	81.86	92.73	0.30

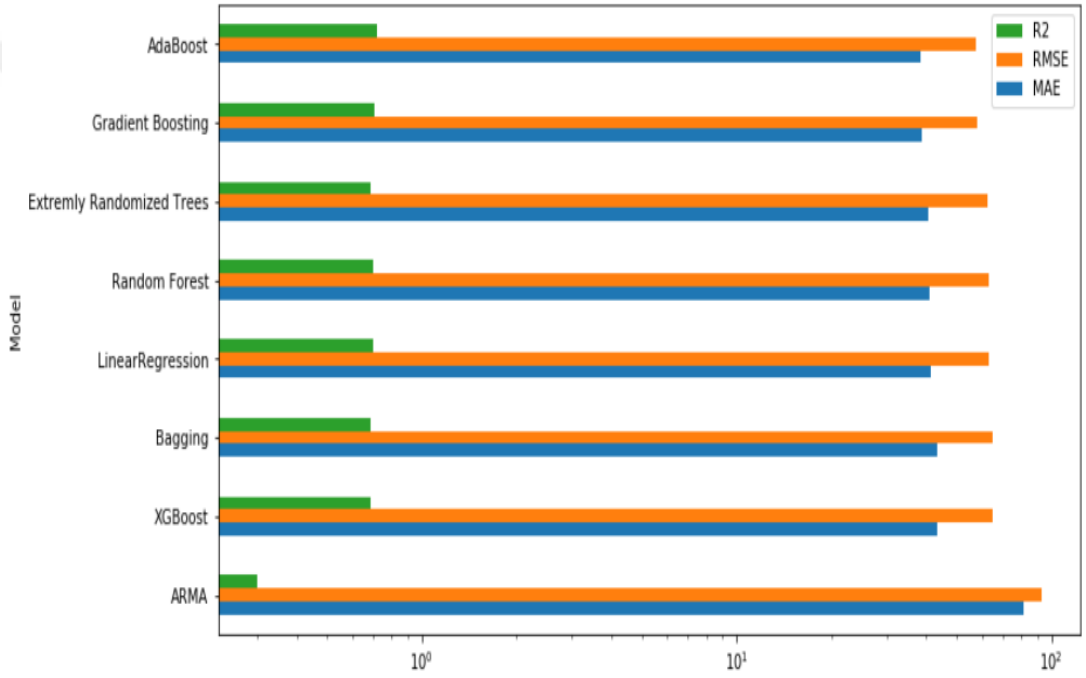


Figure 5.3 : Comparison of models with cross-validation.

Alon et all. (2001) conducted a research on retail sales forecasting by utilizing and comparing of both ANN and traditional methods. This study performs Winters exponential smoothing, ARIMA model and multivariate regression as traditional methods. According to the conclusion of this study, ANN method gives result best across the traditional methods for different time intervals. ANN method is a kind of deep learning algorithm which is one of the powerful branch of machine learning. The results of this study that is applied on retail sales is compatible with the result of current study. Both study state that machine learning algorithms perform better than traditional methods at retail sales forecasting.

Loureiro et all. (2018) is another study where wide variety of methods are used to predict the future sales at fashion retail sector. Decision Trees, Random Forest, Support Vector Regression, Artificial Neural Networks and Linear Regression. Deep Neural Network and Random Forest that gives very close results to each other and present the best results while linear regression presents worst results within the all models. This study is very close to current applied study within both concept, data and applied sector. Moreover, results in both studies support each other with the comparison of linear regression method with machine learning algorithms.

Mullainathan and Spiess (2017) stated the study on machine learning methods with an applied econometric method. This study describes how machine learning methods works and differentiate from traditional methods. Within the scope of this study house values are forecasted with the OLS, Regression tree, LASSO, Random Forest and Ensemble methods. Comparison of forecasting performance of these methods are performed via R^2 values of models. Performance of Random Forest algorithm is the best with serious difference against traditional methods at training sample. However, authors indicates that the machine learning methods have powerful inclination of overfitting and it cause to overstate of performance comparison. Out of train sample forecasting performance results affirms that assertion. Although machine learning algorithms still perform best with respect to traditional method, its performance declined against to train sample performance relatively. In short, this study figures out that, traditional methods has relatively low goodness of fit of forecasting as other studies mentioned before.

Shahrabi et al. (2009) conducted similar research on comparison of machine learning methods and traditional methods. The difference of this study is the applied area which is car company supply chain demand forecasting. Long term car demand of supply chain is forecasted with the methods that are ARIMA, Exponential Smoothing, Exponential Smoothing with Trend being traditional methods and ANN and Support Vector Machines being machine learning methods. The results of this study that is not surprise is coherent with each study referred before. The machine learning methods are more accurate than the traditional forecasting techniques.

Another study at supply chain demand forecasting is conducted by the Carbonneau et all. (2004). As Mullainathan and Spiess (2017) study results, machine learning methods demonstrates best performance. However, this study stated that their

forecasting accuracy is not significantly better than traditional methods such as regression method.

6. CONCLUSIONS

Fashion retail sector is one of the most dynamic sector due to changing trends very rapidly. Therefore, products have high renewal rate which cause to difficulty of accessing historical data. Sales forecasting plays important role from production, distribution to merchandising in retail sector needs historical data. Thus, both retail companies and academic research are interested closely with sales forecasting. Fashion retail sales are considered as time series and classical statistical methods are utilized for long years. Machine learning algorithms which give opportunity to predict sales with different approaches. This study helps to understand both classical methods and machine learning algorithms while comparing the performance of these methods with application of real data. In the content of the study ARIMA, Linear Regression, Bagging Regression Tree, Extremely Randomized Trees, Random Forest Algorithm, Gradient Boosting Algorithm, AdaBoost and Extreme Gradient Boosting Algorithm are applied with daily style data of one of the biggest fashion retail company in the Turkey. According to empirical results, AdaBoost and Gradient Boosting which are boosting algorithms have the best performance. Moreover, all the machine learning methods are performed better than ARIMA. There are one of the hot topic which is deep learning can be utilized for sales forecasting at the following studies. Moreover, sales forecasting based on item attributes such as color, fashion level, material can provides added value results for fashion retail sector.

REFERENCES

Akaike, H. (1970). Statistical predictor identification. Annals of the institute of Statistical Mathematics, 22(1), 203-217.

Alon, I., Qi, M., & Sadowski, R. J. (2001). Forecasting aggregate retail sales:: a comparison of artificial neural networks and traditional methods. Journal of retailing and consumer services, 8(3), 147-156.

Anderson, T. W. (2011). The statistical analysis of time series (Vol. 19). John Wiley & Sons.

Aoki, M. (2013). State space modeling of time series. Springer Science & Business Media.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Choi, T. M., Hui, C. L., Ng, S. F., & Yu, Y. (2011). Color trend forecasting of fashionable products with very few historical data. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(6), 1003-1010.

Farnaaz, N., & Jabbar, M. A. (2016). Random forest modeling for network intrusion detection system. Procedia Computer Science, 89(1), 213-217.

Fildes, R., Ma, S., & Kolassa, S. (2019). Retail forecasting: Research and practice. International Journal of Forecasting.

Freund, Y., & Schapire, R. E. (1995, March). A desicion-theoretic generalization of on-line learning and an application to boosting. In European conference on computational learning theory (pp. 23-37). Springer, Berlin, Heidelberg.

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine

learning, 63(1), 3-42.

Ghimire, B., Rogan, J., Galiano, V. R., Panday, P., & Neeti, N. (2012). An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA. *GIScience & Remote Sensing*, *49*(5), 623-643.

Jain, A., Menon, M. N., & Chandra, S. (2015). Sales Forecasting for Retail Chains.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, pp. 3-7). New York: springer.

Jiang, H., Zheng, W., Luo, L., & Dong, Y. (2019). A two-stage minimax concave penalty based method in pruned AdaBoost ensemble. Applied Soft Computing, 83, 105674.

Karystinos, G. N., & Pados, D. A. (2000). On overfitting, generalization, and randomly expanded training sets. IEEE Transactions on Neural Networks, 11(5), 1050-1057.

Liu, N., Ren, S., Choi, T. M., Hui, C. L., & Ng, S. F. (2013). Sales forecasting for fashion retailing service industry: a review. Mathematical Problems in Engineering, 2013.

Loureiro, A. L., Miguéis, V. L., & da Silva, L. F. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. Decision Support Systems, 114, 81-93.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). Forecasting Methods and Applications, New York: Jhon Wiley & Sons.

Mitchell, T. M. (1997). Machine Learning, McGraw-Hill Higher Education. New York.

Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. Journal of Economic Perspectives, 31(2), 87-106.

Neusser, K. (2016). Time series econometrics. Switzerland: Springer.

Ni, Y., & Fan, F. (2011). A two-stage dynamic sales forecasting model for the fashion retail. Expert Systems with Applications, 38(3), 1529-1536.

Oliveira, J. M., & Ramos, P. (2019). Assessing the Performance of Hierarchical

Forecasting Methods on the Retail Sector. Entropy, 21(4), 436.

Peña, D., Tiao, G. C., & Tsay, R. S. (2011). A course in time series analysis (Vol. 322). John Wiley & Sons.

Puica, M. A., & Florea, A. M. (2013). Emotional belief-desire-intention agent model: Previous work and proposed architecture. International Journal of Advanced Research in Artificial Intelligence, 2(2), 1-8.

Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. Robotics and computer-integrated manufacturing, 34, 151-163.

Rissanen, J. (1978). Modeling by shortest data description. Automatica, 14(5), 465-471.

Roelofs, R. (2019). Measuring Generalization and Overfitting in Machine Learning (Doctoral dissertation, UC Berkeley).

Schaffer, C. (1993). Overfitting avoidance as bias. Machine learning, 10(2), 153-178.

Schuster, H. G. (1988). Deterministic Chaos: An Introduction, Physik-Verlag Gmbh.

Shahrabi, J., Mousavi, S. S., & Heydar, M. (2009). Supply chain demand forecasting: A comparison of machine learning techniques and traditional methods. Journal of Applied Sciences, 9(3), 521-527.

Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications: with R examples. Springer.

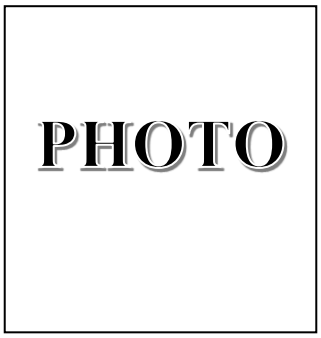
Sun, Z. L., Choi, T. M., Au, K. F., & Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. Decision Support Systems, 46(1), 411-419.

WORLDBANK. World Textile and Clothing Exports By Country 2018. Retrieved from https://wits.worldbank.org/CountryProfile/en/Country/WLD/Year/2018/TradeFlow/Export/Partner/by-country/Product/50-63_TextCloth#

Xia, M., Zhang, Y., Weng, L., & Ye, X. (2012). Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs. Knowledge-Based Systems, 36, 253-259.

Yokuma, J. T., & Armstrong, J. S. (1995). Beyond accuracy: Comparison of criteria used to select forecasting methods. International Journal of Forecasting, 11(4), 591-597.

CURRICULUM VITAE



Name Surname **: Hanife Işık**

Place and Date of Birth **: Istanbul – November 21, 1992**

E-Mail **: hanifeeisik@gmail.com**

EDUCATION

- **B.Sc.** **:** 2016, Istanbul Technical University, Management Engineering

  2015, Istanbul Technical University, Food Engineering

- **M.Sc.** **:** 2020, Istanbul Technical University, Economics

PROFESSIONAL EXPERIENCE AND REWARDS:

- February 2020 – Present: Business Analytics Specialist at LC WAIKIKI RETAIL
- March 2018 – February 2020: Retail Analytics Specialist at KOTON RETAIL
- May 2016 – March 2018: Store Planner at LC WAIKIKI RETAIL
- Graduation with the 1st place in the Management Engineering Programme Ranking, 2016
- Graduation with the 1st place in the Faculty Ranking, 2015
- Graduation with the 1st place in the Food Engineering Programme Ranking, 2015
- Ömer Duruk Achievement Award with First Degree, 2015
- Pelin Köz Achievement Award with First Degree, 2014
- 'TUGIS' Food Industry Achievement Awartd with First Degree, 2012

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Çelebi, D., Yörüsün, A., & Işık, H.** (2018). Bicycle sharing system design with capacity allocations. Transportation research part B: methodological, 114, 86-98.