YAŞAR UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
BUSINESS ADMINISTRATION PROGRAMME

PhD THESIS

A BIG DATA ANALYTICS BASED METHODOLOGY FOR STRATEGIC MARKET ANALYSIS

MURAT ÖZEMRE

THESIS ADVISOR: ASST. PROF. DR. ÖZGÜR KABADURMUŞ

2019 İZMİR

PhD/PROFICIENCY IN ART THESIS JURY APPROVAL FORM

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of PhD/Proficiency in Art.

26.08.2019

Asst. Prof. Dr. Özgür KABADURMUŞ

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of PhD/Proficiency in Art.

26.08.2019

Prof. Dr. Sabri ERDEM

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of PhD/Proficiency in Art.

26.08.2019

Prof. Dr. Okan TUNA

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of PhD/Proficiency in Art.

26.08.2019

Assoc.Prof. Dr. Güzin ÖZDAĞOĞLU

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of PhD/Proficiency in Art.

26.08.2019

Asst. Prof. Dr. Görkem SARIYER ATAMAN

Assoc.Prof.Dr. Çağrı BULUT

Director of the Graduate School of Social Sciences

ABSTRACT

A BIG DATA ANALYTICS BASED METHODOLOGY FOR STRATEGIC MARKET ANALYSIS

Murat ÖZEMRE
PhD Program in Business Administration
Advisor: Asst. Prof. Dr. Özgür KABADURMUŞ
2019

Today's competitive business environment forces companies to make better predictions and decisions for their business environments. Therefore, strategic market analysis is one of the most critical tasks for companies. However, business decision-makers should absorb the high volume of data with different views before making their strategic decisions. This dissertation presents a novel and holistic methodology for strategic market analysis by using Big Data Analytics. The proposed methodology of this dissertation employs two different machine learning algorithms, Random Forest (RF) and Artificial Neural Networks (ANN), to forecast the export volumes using an extensive amount of open trade data. Then, the forecasted values are included in the Boston Consulting Group (BCG) Matrix to conduct strategic market analysis. To demonstrate the effectiveness of the proposed methodology, two hypothetical case studies of a Turkish and Chinese company exporting refrigerators and freezers to the United Kingdom are considered and the managerial implications after implementing the proposed methodology are presented.

Keywords: Big Data Analytics, Strategic Market Analysis, Trade Volume Forecasting, Boston Consulting Group Matrix, Machine Learning

ÖZ

BÜYÜK VERI ANALITIĞI YÖNTEMİYLE STRATEJİK PAZAR ANALİZİ

Murat ÖZEMRE
Doktora Tezi, İşletme Anabilim Dalı
Danışman: Dr. Öğr. Üyesi Özgür KABADURMUŞ
2019

Günümüzün rekabetçi iş ortamı, şirketleri içinde bulundukları iş ortamları ile ilgili daha iyi tahmin yapmaya ve daha doğru kararlar almaya zorlamaktadır. Bu nedenle, stratejik pazar analizi şirketler için en kritik görevlerden birisi olmaktadır. Ancak, karar vericiler stratejik kararları vermeden önce ellerinde bulunan yüksek hacimli veriyi, farklı bakış açılarını da katarak iyice özümsemelidirler. Bu tez, Büyük Veri Analitiği kullanarak stratejik pazar analizi için bütünsel bir yöntem sunmaktadır. Sunulan yöntem, açık ticaret verilerinin kullanılması ve iki farklı makine öğrenme algoritması (Random Forest (RF) ve Yapay Sinir Ağları (YSA)) yardımıyla, ülkeler arasında ürün bazında ihracat hacmini tahmin etmektedir. Ardından, elde edilen sonuçlar stratejik pazar analizi yapmak için Boston Consulting Group (BCG) Matrisine dahil edilir. Önerilen metodolojinin etkinliğini göstermek için, İngiltere'ye buzdolabı ve dondurucu ihraç eden farazi bir Türk ve bir Çin şirketi üzerinde uygulanmıştır. Uygulamanın ardından elde edilen sonuçlara göre yönetsel çıkarımlar sunulmuştur.

Anahtar sözcükler: Büyük Veri Analitiği, Stratejik Pazar Analizi, Ticaret Tahminleme, Boston Consulting Group Matrisi, Makina Öğrenmesi

ACKNOWLEDGEMENTS

I would like to express my sincere indebtedness to Asst. Prof. Dr. Özgür KABADURMUŞ for his constant guidance and supervision in the course of this dissertation.

I would like to thank Prof. Dr. Durmuş ÖZDEMİR and Prof. Dr. Sabri ERDEM for their valuable advice during dissertation monitoring studies.

I would like to thank you my BİMAR colleagues and my other friends for their continuous support.

A big thank you to my wife Kübra and my son Kerem for their endless patience and invaluable support.

Murat ÖZEMRE

İzmir, 2019

TEXT OF OATH

I declare and honestly confirm that my study, titled "A Big Data Analytics Based Methodology for Strategic Market Analysis" and presented as a PhD Thesis, has been written without applying to any assistance inconsistent with scientific ethics and traditions. I declare, to the best of my knowledge and belief, that all content and ideas drawn directly or indirectly from external sources are indicated in the text and listed in the list of references.

Murat ÖZEMRE

Signature

....

September 19, 2019

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	: Artificial Neural Networks.
APEC	: Asia-Pacific Economic Cooperation
AR	: Auto Regressive
ARIMA	: Auto Regressive Integrated Moving Average
AUT	: Austria
BCG	: Boston Consulting Group
BCI	: Business Confidence Indicator
CCI	: Consumer Confidence Indicators
CHN	: People's Republic of China
CHZ	: Czechia
CLI	: Composite Leading Indicators
CNN	: Convolutional Neural Networks
CPB	: Netherlands Bureau for Economic Policy Analysis
CPI	: Consumer Price Indices
CRISP-DM	: CRoss Industry Standard Process for Data Mining
DEU	: Germany
EFTA	: European Free Trade Association
EG	: Engle–Granger.
EPU	: Economic Policy Uncertainty
EXC	: Exchange Rate
FGRTF	: Forecasted Growth Rate of Trade Flow
FIT	: Forecasting International Trade
FRA	: France
FRMS	: Forecasted Relative Market Share
GDP	: Gross Domestic Product
GRTF	: Growth Rate of Trade Flow
GRU	: Gated Recurring Units
HS	: Harmonized System.
IRQ	: Iraq
ITA	: Italy
LSTM	: Long Short-Term Memory
MA	: Moving Average

MLP	: Multi-Layer Perceptron
NLD	: Netherlands
OECD	: Organization of Economic Development and Corporation.
PPI	: Producer Price Indices
RF	: Random Forest.
RMS	: Relative Market Share
RMSE	: Root Mean Square Error
RNN	: Recurrent Neural Networks
SCM	: Supply Chain Management
SMA	: Strategic Market Analysis
SWOT	: Strength, Weaknesses, Opportunities, Threats
TUR	: Turkey
UK	: United Kingdom
USA	: United States of America
VAR	: Vector Auto Regression.
VECM	: Vector Error-Correction Models.
WTO	: World Trade Organization

INTRODUCTION

With the rise of globalization, supply chains operations became more complex and therefore harder to manage. Nowadays, the companies not only interact with the companies within their supply chains and but also interact with outside supply chains. They also constantly compete with global supply chains. This fierce global competition increases the importance of the effective management of supply chain operations. Many supply chain operations, such as production, procurement, sales, warehousing, and forecasting, cannot be effectively designed with considering only local parameters. Today's supply chains are interconnected with global companies and supply chains, and this requires to think globally when designing and managing supply chain operations. Thus, the supply chain operations must be designed and managed with the global parameters and the effective management of global supply chain operations has become very important (Ozemre & Kabadurmus, 2019).

On the other hand, business decision-makers should absorb the high volume of data with different views before making their decisions. Using big data enables managers to decide on the basis of evidence rather than intuition. As someone has more data and make better use of that data, the decision making process would yield better decisions. Therefore, a company should start to collect as much internal and external data as possible (Hofmann, 2017). Finding the right open/public data sources, making regular updates, and integrating them with the internal data with the correct measure, scale and period are considered as some of the obstacles for businesses. Companies that overcome these obstacles and smooth this process would get a competitive advantage in their market. The process of using advanced technologies to examine big data in order to uncover useful information defined as Big Data Analytics (BDA) (D. Q. Chen, Preston, & Swink, 2015).

To effectively design logistics activities, the availability of accurate forecast data is crucial. For example, the effective resource allocation of a company for the distribution of its goods depends on the sales forecast data. Similarly, the quality of the sales forecast affects the performance of production scheduling and resource utilization. A good inventory management practice can help to achieve an agile response to the customer demand, however, it depends on the accurate forecast data.

All these company level forecast related issues are important, but the requirement of a good forecast becomes more significant for the design of supply chain operations. In supply chain management, accurate forecasts help to streamline the operations. For example, data sharing with the other echelons of the supply chain can lead to better forecasts and help reducing bullwhip effect, which may lead to increased inventories, poor customer service levels, poor resource allocation, and wrong logistics decisions (Kabadurmus, Erdogan, & Tasgetiren, 2017). Therefore, an effective forecasting practice can reduce the inventory levels without affecting the service level and improve supply chain performance.

Forecasting in global supply chain operations is more challenging than the forecasting in local supply chains due to the complexity of the global supply chain networks. For exporters (or importers), predicting export (or import) volumes are also important since their entire supply chain operations depend on the forecasted exports. The prediction of the total export volume of a country to a specific country may help to adjust their marketing strategies. If the exporting company can foresee that the total export volume would increase in the future, they can increase their production by adjusting their own supply chain operations. If they can predict that the export volume to a specific country to be reduced in the near future, they can search for alternative markets to sell their products and reshape their global supply chain operations without hindering the progress of supply chain strategies. Therefore, being able to make accurate forecasts is very important. However, in today's global and complex trade environment, forecasting has become even more difficult (Ozemre & Kabadurmus, 2019).

Global trade is one of the areas having various and mature data resources. Also, World Trade has been continuously growing. World merchandise exports have increased by 32 percent in value since 2006 (Piezas-Jerbi & Wardyn, 2017). Manufactured goods and agricultural products have the largest share in this increase. China became one of the major countries in the world trade where its merchandise trade contribution to world trade is 13 percent in 2018 and Turkey's contribution is 0.9 percent. However, the growth of China's exports has recently lost its momentum and is projected to decline in 2019 and 2020 (OECD, 2019a). With the recent escalations of trade tensions, increased competition in world trade is expected. At the company level, strategic decision-making tools have become significant to survive in this competitive environment.

With the increased availability of wide data sources, exporters and importers can assess the progress of the trade volumes between countries and adjust their global supply chain operations accordingly. Nevertheless, they usually try to reach some statistics about a specific market country, period and product, and therefore, obtain very limited information about market behavior. With the help of a holistic approach, they could continuously assess the market situation from different perspectives, and then they could use this market assessment in their decision-making process by simultaneously considering the market, product and time-related factors.

Problem Definition

To make strategic market analysis, market, product, and time-related factors can be combined with the available wide data sources using BDA to have a clearer vision about the specific country and product pairs that they should focus on in the future. According to the above motivations, the problem definition of this dissertation is to build a decision support system for export/import companies to help their Strategic Market Analysis. While building this system, both internal and external data should be used. The system analyzes the product groups and countries that exporter/importers operate. The proposed model forecasts market behavior for each product and country using the available wide data sources with BDA algorithms. The analysis results give a clear insight about the current positon in the market for each product and country pairs. According to the forecasted market behavior and different sales scenarios of the company, future situation of the company in the market can be positioned. By using this information, recommendations for strategic market analysis can be generated.

A more detailed explanation of the problem definition is given in Chapter 2.

Research Questions

To solve the above problem, the following research questions are determined in this dissertation:

- How open data sources can be used in strategic market decisions for exporters/importers?
- Which strategic management models and open data can be used together for strategic market analysis?

- How an additional value can be added by using Big Data Analytics in selected strategic market analysis method?
- Which factors are affecting the bilateral trade forecasting?
- How accurate forecast results in the selected machine learning method can be achieved?

These are the main research goals that are aimed to solve by using the proposed big data analytics framework developed in this dissertation. To answer these questions, various international trade data are considered by using BDA and managerial implications are devised for a company. The Boston Consulting Group (BCG) Matrix has been adopted as a strategic management tool in this dissertation. This dissertation also provides a non-parametric forecasting method using machine learning algorithms with an extensive amount of data. Two different machine learning algorithms, Random Forest (RF) and Artificial Neural Networks (ANN) have been applied to forecast the export volumes. The previous studies in the literature only developed forecasting models, but this dissertation uses forecasting as an input to the proposed strategic market analysis model and developed a holistic Big Data Analytics framework for strategic market analysis. Different than the other models in the literature, the proposed Big Data Analytics approach employs more variety (and amount) of data sources and machine learning features to forecast export volumes. To demonstrate the effectiveness of the methodology developed in this dissertation, a hypothetical case study of a Turkish company exporting refrigerators and freezers to the United Kingdom is presented. This product group is one of the main export products of Turkey, and United Kingdom is one of the main importers. The effectiveness of the proposed methodology is validated in two ways. The first one is, that it does not depend on one product, three product groups are selected to widen the product space. The second one is, that it does not depend on one single exporting country since a Chinese company is also tested as an additional case study.

Contributions

As discussed in detail in Chapter 1 Literature Survey, there are two main groups of studies on BCG Matrix or positioning map implementations. One group of studies bring new approaches to new dimensions. The second group made situation analysis with up to date information. Both groups are focused on the current situation of a company or organization. There is no case study including the forecasted market results with sensitivity analysis. There are various studies on using international trade data implementations on strategic market analysis using BCG Matrix type of tools. Mainly they are focused on country level analysis. There are no studies that use international trade data for company level analysis.

The literature on forecasting international trade is focused on the improvement of the forecasting results. The resulting models are not used to derive managerial implications or build a decision support system. Also, the machine learning models on international trade forecasting use limited external factors other than product information. In literature, there is no holistic approach or process model to employ Big Data Analytics in Strategic Market Analysis.

The main contribution of this dissertation to the literature is the developed holistic Big Data Analytics framework for Strategic Market Analysis. In this dissertation's scope, international trade data and BCG Matrix are used together, but the other data sources and market analysis tools can be easily applied by using this approach. Building a Machine Learning model with an extensive amount of open data to forecast international trade volumes is the second biggest contribution to the literature. Enhancement of the well-known BCG matrix with forecasted future position and with different scenarios is the other main contribution to the literature. In previous studies, international trade data was used to compose the country-level BCG matrix. This dissertation's contribution is using international trade data to compose company level BCG Matrix.

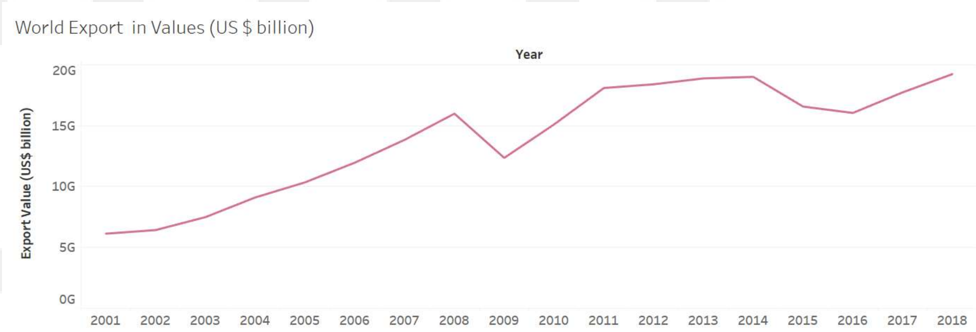
This dissertation organized as follows. The next chapter gives short information on Bilateral World Trade and then summarizes the literature on the forecasting international trade, big data analytics, strategic market analysis and machine learning models respectively. In Chapter 2, problem definition, the steps for determination of the case, selection of export and import countries and reference country and products are given in detail. In Chapter 3 the proposed holistic Big Data Analytics framework which combines strategic market analysis model and forecasting international trade is presented. Chapter 4 demonstrates the application of the proposed methodology and reports the results of the case study in detail. Conclusion and future works are summarized in the last chapter.

1. LITERATURE SURVEY

1.1. Bilateral World Trade

World trade is continuously growing by means of volume and value. Merchandise exports of World Trade Organization (WTO) members totaled 19.22 trillion US Dollars in 2018, where it was 17.72 trillion in 2017. Merchandise trade grew by 10.9 percent in value and 4.7 percent in volume in 2017 (World Trade Organization, 2018). Trade has been continuously growing for seven years in terms of volume. Figure 1 shows the yearly export value changes between 2001 and 2018. The effect of the 2009 crisis can be seen from this figure.

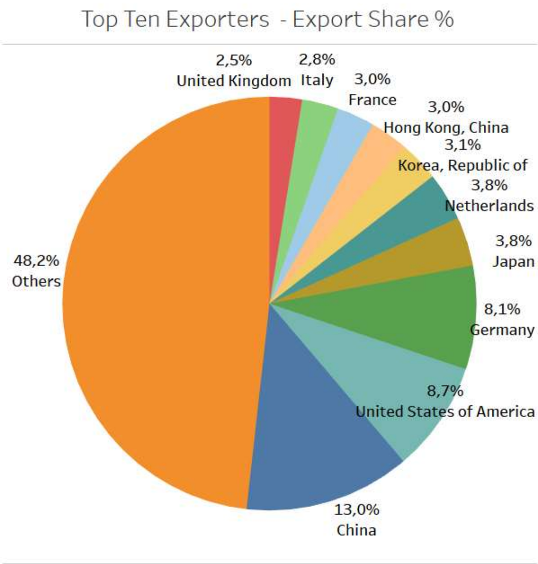
Figure 1. World Export in Value



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

The top ten traders are accounted for exports totaling almost 9.95 trillion US Dollars with a share of 51.8 percent. Figure 2 shows the export share of the top ten exporters and the rest of the world in 2018. China, the United States of America, Germany, Japan, Netherlands, Korea, Hong Kong, France, Italy and the United Kingdom are the top ten exporters in 2018.

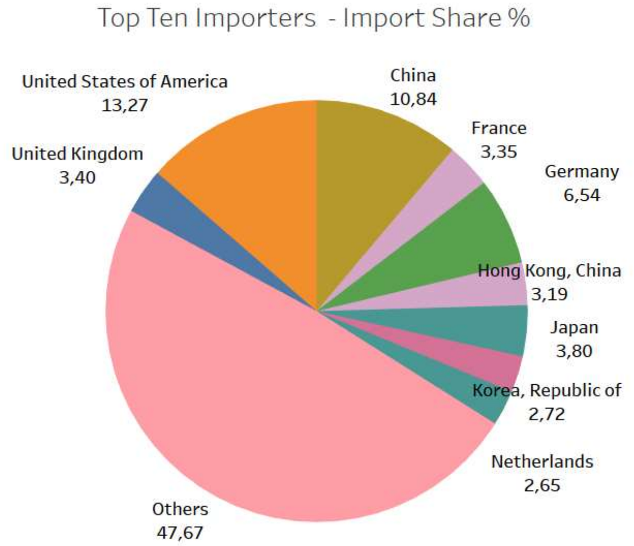
Figure 2. Export Share According to Export Values in 2018



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

The top ten traders are also accounted for imports totaling almost 10.3 trillion US Dollars with a share of 52.33 percent. Figure 3 shows the export share of the top ten importers and the rest of the world in 2018. The United States of America, China, Germany, Japan, The United Kingdom, France, Hong Kong, Korea, Netherlands and India are the top ten importers in 2018.

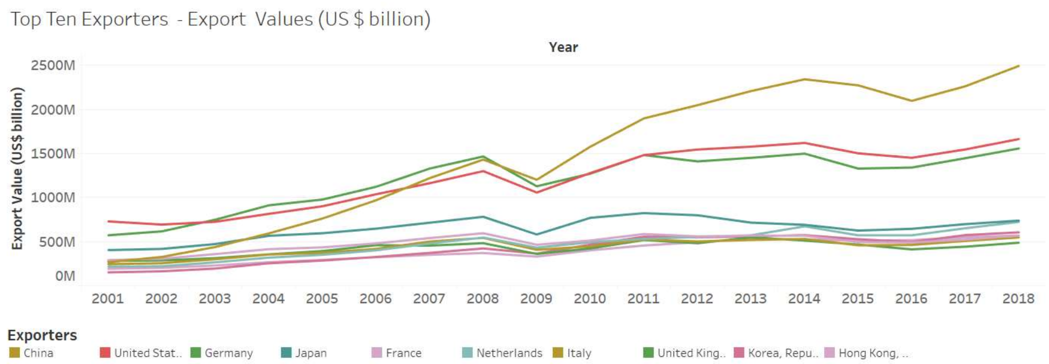
Figure 3. Import Share According to Import Values in 2018



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure 4 shows the export value changes in the top 10 exporters from 2001 to 2018. The rise of China can be easily seen to pass the USA and Germany after the 2009 crisis. The traders other than the top three countries (China, USA and Germany) remain in their steady positions.

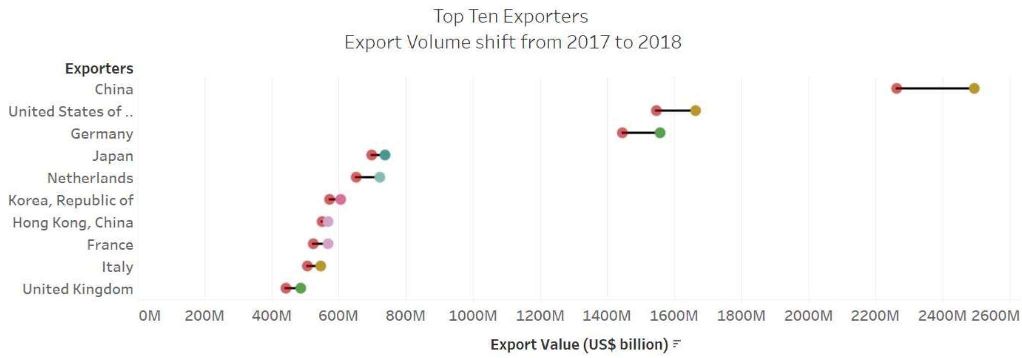
Figure 4. Export Value Changes of Top 10 Traders



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure 5 shows the export value changes in US Dollars between 2017 and 2018. The changes are mostly proportional to their export volumes. Especially, the increase in the export volume of China significantly exceeds the others.

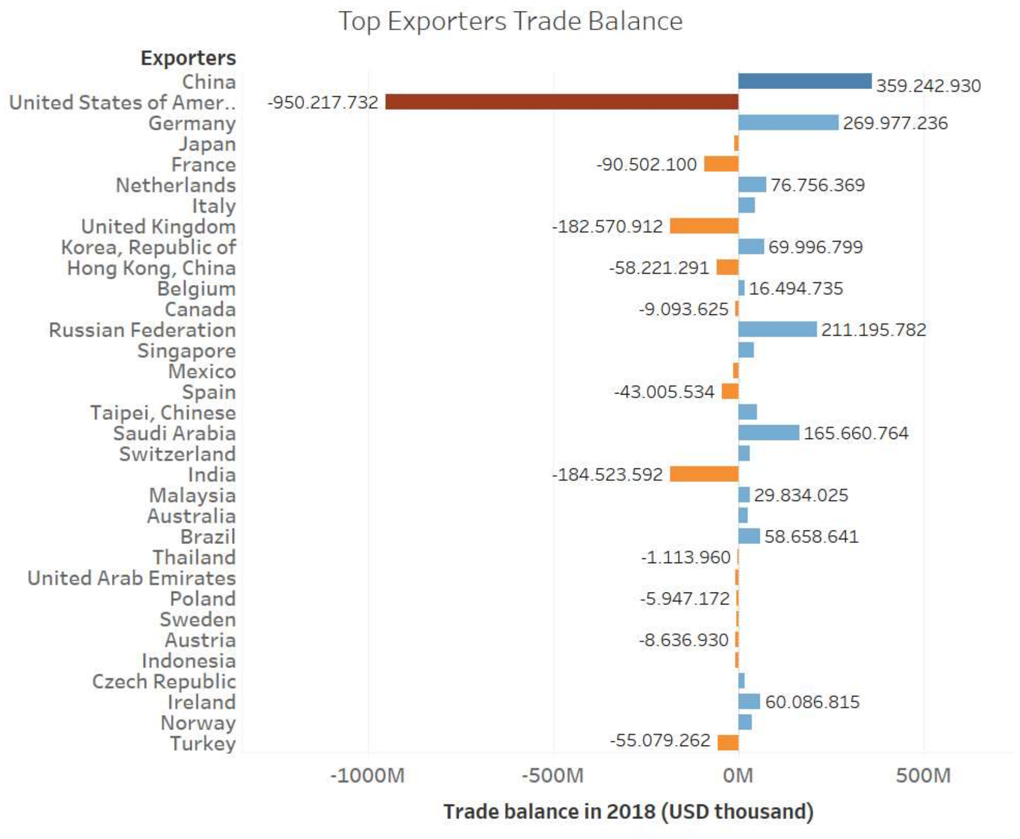
Figure 5 Export Volume Changes of Top Ten Exporters from 2017 to 2018



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

To evaluate country trades from another perspective, the trade balance of the countries are evaluated. Figure 6 shows the trade balance of the top exporters. While some countries such as China, Germany, Netherland, Italy and Korea have a positive trade balance, some countries such as the USA, France and the United Kingdom have a negative trade balance. According to this figure, China is the major exporter in the world.

Figure 6. Trade Balance of Top Exporters



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Successful portfolio management is important to be one step ahead in today's competitive business environment. Using the right data and the ability to conclude meaningful results have become important for business decision-makers (Bohanec, Robnik-Sikonja, & Kljajic Borstnar, 2017). Exporters, importers and logistics service providers are considered as the potential users of the proposed methodology. Therefore, evaluation of the current and future business situations in international trade is important for decision-makers. International trade is evaluated in three different aspects:

- The first aspect is the representation of the market. Single country, a geographical region (North Africa, Middle East, etc.) or group of countries with common perspectives (OECD, EFTA, etc.) can be considered as a market.

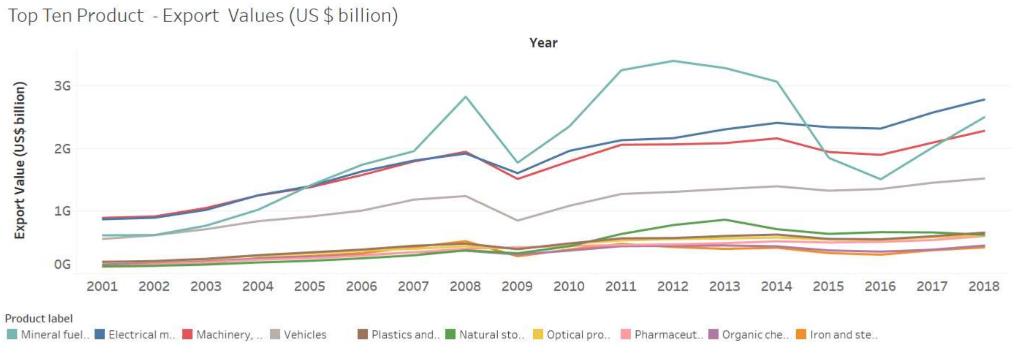
- The second one is the prediction horizon, which can vary from a single month to years.
- The third aspect is the product granularity. Product granularity is handled according to the Harmonized Commodity Description and Coding System (HS) levels.

The Harmonized System is an international nomenclature for the classification of products (Council, Customs, & Co-operation, 2017). HS was introduced in 1988 and has been adopted by most countries worldwide. It has undergone several changes in the classification of products. The HS comprises approximately 5,300 articles/product descriptions that appear as headings and subheadings, arranged in 99 chapters, grouped in 21 sections. The products are represented as six-digit codes in the HS. The six-digit codes can be broken down into three parts. The first two digits of an HS code represents the chapter of the goods is classified in (e.g., “84” is “Machinery”). The next two digits represent the grouping within that chapter. For example, “84.18” code the “Refrigerator and freezers” group in the machinery chapter. The last two digits identify a specific product within that group. For example, “84.18.40” represents the upright type freezers with a maximum capacity of 900 liters.

Manufactured goods are accounted for 70 percent of all merchandise exports. The largest group is fuels and mining products with 15 percent of the total. Agricultural products take 10 percent. Figure 7 shows yearly changes in the top 10 traded product groups in US Dollars. In addition, the reason for the rise and fall of the product group “mineral fuels” is the changes in oil prices over the years.

- Electrical machinery and equipment – HS 85
- Mineral fuels, mineral oils and products – HS 27
- Machinery, mechanical appliances – HS 84
- Vehicles – HS 87
- Natural stones, precious metals – HS 71
- Plastics and articles thereof – HS 39
- Optical products – HS 90
- Pharmaceutical products – HS 30
- Organic chemicals – HS 29
- Iron and steel – HS 72

Figure 7. Top Ten Product Group Value Changes in billion US Dollars



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

In the following chapter, the studies about forecasting international trade are given. Both parametric and non-parametric approaches and their implementation areas with their methods are given. In the last part of the chapter, other forecasting approaches in other domains are given.

1.2. Forecasting International Trade (FIT)

Supply Chain Management (SCM) consists of all the necessary steps to fulfill customer demand (Raisinghani & Meade, 2005). This makes supply chain management a critical tool to achieve a competitive advantage for cooperation (Stefanovic & Stefanovic, 2009). Demand management is one of the key processes of SCM (Barbosa, Vicente, Ladeira, & de Oliveira, 2018). Big Data Analytics enables companies to develop better relationships with their customers and suppliers based on a deep understanding of their market demands (Sanders, 2014). A proper understanding of the market demands helps designing supply chains to respond faster and more effectively to changing customer and supplier needs (Sanders, 2014). Demand has become more and more volatile over time due to the increasing volume of trade. Which creates difficulty to estimate demand (Speranza, 2018). According to the latest literature review studies, the number of publications on Big Data Analytics has increased dramatically for the last five years (Sivarajah, Kamal, Irani, & Weerakkody, 2017) in parallel to the increase in supply chain management studies (D.

Mishra, Gunasekaran, Papadopoulos, & Childe, 2018; Nguyen, Li, Spiegler, Ieromonachou, & Lin, 2018). Demand management by using Big Data Analytics covers 25 percent of all Big Data Analytics in the SCM studies (Barbosa et al., 2018). These studies mainly focus on forecasting, synchronizing and predicting customer demands.

Over the past years, various forecasting models were applied for trade forecasting that uses extrapolation, time-series and economic models, agent-based computational economics models, and machine learning algorithms (Nummelin & Hänninen, 2016). Two mainstream research approaches, parametric and non-parametric approaches, were used in trade forecasting models. Auto-Regressive Integrated Moving Average (ARIMA), Holt-Winter and their several variations are the most commonly used parametric time series forecasting techniques in trade volume forecasting. Dale & Bailey (1982) used the Box Jenkins method to forecast U.S. merchandise exports. Senhadji & Montenegro (1999) estimated export demand elasticities in developing and industrial countries by time-series techniques. Veenstra & Haralambides (2001) forecasted seaborne trade flows of four different commodity products (crude oil, iron ore, grain and coal) by using Vector Auto Regression (VAR). Akhtar (2003) investigated seasonality in Pakistan's merchandise exports and imports by using a univariate ARIMA model. Sahu & Mishra (2013) used ARIMA to estimate spice import and export volumes, and production behaviors for India and China. Khan (2011) used ARIMA, Holt-Winter and Vector Auto Regression (VAR) techniques to forecast Bangladesh's total imports. Similarly, Kargbo (2007) used ARIMA, VAR, Engle-Granger (EG) single-equation and vector error-correction models (VECM) to forecast South Africa's agricultural exports and imports. Emang, Shitan, Abd. Ghani, & Noor (2010) built univariate time series models to forecast the export demand of moulding and chipboard volume for Peninsular Malaysia. Ge & Wu (2019) uses multiple linear regression analysis in the prediction of corn price fluctuation.

The need for managing structural interdependencies and parametric assumptions make the usage of parametric models harder in realistic applications. In addition to the difficulties of applying parametric models, the availability of machine learning tools and open data sources are the other reasons to use non-parametric models. Machine learning models used in non-parametric solutions where they are versatile and flexible in making use of data to capture complex behaviors (Choi, Wallace, & Wang, 2018). The use of big data with non-parametric models to understand market behavior will use more inductive techniques than deductive techniques in the future (Erevelles, Fukawa, & Swayne, 2016). Co & Boosarawongse (2007) compared their Artificial Neural Network model with the parametric models (exponential smoothing) to forecast Thailand's rice export. Their study applied to four different types of rice products and showed that ANN yielded the best forecast results. Similarly, Pakravan, Kelashemi, & Alipour (2011) forecasted Iran's rice import by using ANN. Silva & Hassani (2015) used Singular Spectrum Analysis to forecast U.S. Trade before, during and after the recession of 2008. Nummelin & Hänninen (2016) applied support vector machines (SVM), ANN and Random Forests models to analyze and forecast global bilateral trade flows of soft sawn wood. Similar to the model developed in this dissertation, their model uses not only export volumes but also other economic indicators, such as exchange rates and GDP (Gross Domestic Product). However, their model includes a limited number of factors, whereas this dissertation's model employs a broad range of factors. Ozemre & Kabadurmus (2019) developed a focast model that uses various open data sources and employs Random Forest and Artificial Neural Networks to forecast the export volume of refrigerators and freezers from Turkey to the United Kingdom. Shibasaki & Watanabe (2012) forecasted cargo flow in Asia-Pacific Economic Cooperation (APEC) region by using relations between economic growth, trade, and logistics demand models. Sokolov-Mladenović, Milovančević, Mladenović, & Alizamir (2016) modeled economic growth using ANN based on trade, import and export parameters. Gupta & Kashyap (2015) forecasted inflations of G-7 countries using ANN. In the proposed model of this dissertation, economic growth and inflation to predict import and export volumes are used. Iebeling & Milton (1996) built an ANN model to forecast financial and economic time series.

In addition to the prediction in trade and economic models, the RF and ANN algorithms have been widely used in different fields (Palmer, Montaño, & Sesé, 2006). For example, Doganis, Alexandridis, Patrinos, & Sarimveis (2006) used ANN and evolutionary computing for sales forecasting of short shelf-life food products. Palmer et al. (2006) forecasted tourism time series using ANN. Chong, Li, Ngai, Ch'ng, & Lee (2016) developed an ANN model to forecast product sales of e-commerce companies by employing user-generated content. Law (2000) forecasted tourism demand by comparing the back-propagation neural network, regression models, time-series models, and feed-forward neural networks. Ayankoya, Calitz, & Greyling (2016) built a neural network model to predict grain commodity prices in South Africa.

Laptev, Yosinski, Li, & Smyl (2017) studied time-series modeling based on Long Short Term Memory (LSTM) to forecast extreme events for Uber.

In the following chapter, the relationship between Big Data and Big Data Analytics is discussed. In addition to the ongoing discussions about BDA, obstacles and opportunities are examined. Also, in the next chapter, the base of the proposed Big Data Analytics process and the progress steps are given.

1.3. Big Data Analytics (BDA)

Over the last decade, there has been a dramatic increase in the amount of data. The main enablers for this data explosion are the wide-spread usage of digital computers, connected devices and internet access. However, humanity has been dealing with data increase for many decades. For example, in 1944, the information in libraries was expected to double in size every sixteen years (Henkle & Lubetzky, 1946). In 2012, this doubling cycle was reduced to 2 years (Gens, 2012) and shorter doubling cycles are expected in the future with the aid of the digitalization wave. According to M. Chen, Mao, & Liu (2014), the amount of data generated in two days in 2011 is equal to the amount of data from the start of the civilization to 2003. One of the most expected outcomes of the digitalization wave is that everyone would be a data provider in the near future (Gens, Philip, & Bill, 2018). As most of the large enterprises may get additional revenues from their data-related services, Big Data will be one of the most important topics in the near future.

Although the discussions about the definition of big data still continue (Hu, Wen, Chua, & LI, 2014), there are many defining features and properties of big data. The size of the data is the most obvious property of big data (Russom, 2011), yet the main characteristics of big data are defined by five V’s: Volume, Velocity, Variety, Veracity and Value (Nguyen et al., 2018; Wamba et al., 2017). The size of the data determines the volume. Velocity is affected by the frequency of data. Variety is determined by the combinations of different types of structures within the data. Each of these factors introduces different issues (Fan, Han, & Liu, 2014). For example, a massive amount of data comes with its heterogeneity and a high number of variables increases noise accumulation by hiding the real variable, spurious correlation and incidental endogeneity. Because of these difficulties, Veracity has become an important factor (Addo-Tenkorang & Helo, 2016), which is one of the biggest challenges in Big Data Analytics (BDA). Although volume is the starting factor in big data concept, companies can get the highest benefit by focusing on variety in their BDA initiatives (Davenport & Dyché, 2013). The fifth V, Value, is important for creating the highest impact during the Big Data Analytics process. According to a recent survey (LaValle et al., 2011), the quality of the data is one of the main issues in BDA projects. To achieve real value from Big Data, a variety of techniques and disciplines, including statistics, data mining, machine learning, social network analysis, signal processing, pattern recognition, optimization methods, and visualization approaches can be used (C. P. Chen & Zhang, 2014).

Using the right data and the ability to conclude meaningful results have become important for business decision-makers (Bohanec, Robnik-Sikonja, & Kljajic Borstnar, 2017). This makes the strategic decision-making process more dynamic and adaptive (Merendino et al., 2018). Higher operational efficiency, better strategic decision-making, better visibility, improvement in customer service, better new products (or services), and enhancement of customer experience are the most commonly agreed big data opportunities (C. P. Chen & Zhang, 2014; Russom, 2011; Schoenherr & Speier-Pero, 2015). Although BDA can help reducing costs, being more agile and achieving higher service levels (Nguyen et al., 2018), less than 20% of companies adapted BDA in their supply chains due to people, culture or process-related challenges (Wang et al., 2016). Organizations that have been generating or dealing with large data sources for decades can become early adopters of BDA (Villars, Olofson, & Eastwood, 2011). In some industries, early adopters already started to enjoy these benefits. Government organizations, classic commercial high-performance computing spaces and financial services can be considered as such organizations. With the addition of new data sources and the ease of access to computational resources, some industries recently started to benefit from the above-mentioned big data opportunities. These industries include media/entertainment, healthcare, life sciences, surveillance, transportation, logistics, retail, utilities and telecommunication (Villars et al., 2011). To successfully attain these big data opportunities, some barriers and challenges should be overcome by business managers. Some of these challenges are technical and some are people, culture or process-related (C. P. Chen & Zhang, 2014; Hu et al., 2014; Russom, 2011; Schoenherr & Speier-Pero, 2015; Villars et al., 2011). According to Sivarajah, Kamal, Irani, & Weerakkody (2017), all these challenges can also be grouped as data challenges, process challenges and management challenges. Since many organizations are still new in BDA, the inadequate staffing and skills to architecting and building analytics systems, integrating with existing systems, overcoming the privacy, confidentiality, security and governance issues are technical and people related obstacles. Unclear BDA roadmap and business cases, lack of upper management support, resistance to change are also some of the common challenges about cultural and process-related issues.

Today's competitive business environment requires companies to find the root causes of problems faster, react to changing conditions immediately, and make better predictions for their business environments. To make effective decisions, they should adapt to their business environments. BDA is considered as a game-changer by enabling these capabilities (Wamba et al., 2017). Many companies focus on their internal data to make their strategic and tactical decisions. For example, performance measurements rely mostly on the previous year's performance of a company without considering realistic market insights. Therefore, many companies do not monitor their relative performances according to the actual behavior of the market and business conditions. Successful business leaders intuitively take the external environment into account, but business environments are changing faster and different networks are constantly interacting with each other. To deal with this complexity, companies started to include more data sources (internal/external) in their decision-making processes. Data sources were limited in the past, however, with the help of digitalization, today's companies have easy access to various open/public data sources. The combination of internal and external data also introduces some issues to the process. It is similar to the blind men and elephant analogy, where the elephant rapidly grows and continuously changing its pose (Wu, Zhu, Wu, & Wei, 2014).

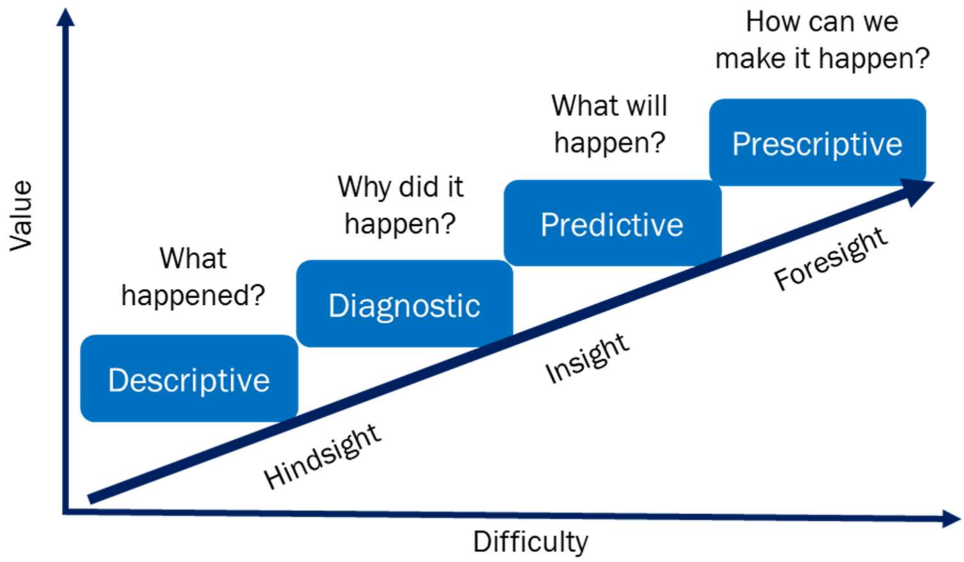
Big Data Analytics can be seen as data manufacturing and it is similar to traditional manufacturing (Hazen, Boone, Ezell, & Jones-Farmer, 2014) because raw data are converted into data and managerial decisions like converting raw materials into physical products. The proposed Big Data Analytics methodology of this dissertation is mainly based on the CRISP-DM (CRoss Industry Standard Process for Data Mining) methodology (Wirth & Hipp, 2000). In Figure 8, the adopted CRISP-DM process is presented.

Figure 8. CRISP-DM (CRoss Industry Standard Process for Data Mining) Process

Source: Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In The Fourth International Conference on the Practical Application of Knowledge Discovery

Big Data Analytics algorithms are very powerful to analyze data, diagnose and identify problems. BDA also helps to give foresight for the current and forthcoming conditions. As shown in Figure 9, there are four levels of Big Data Analytics (Delen & Zolbanin, 2018). The created value on each level is proportional with the difficulty to overcome that level. The first level, descriptive analytics, explains what happened, also called as business reporting or business intelligence. The second level is named as diagnostic, which answers why it happened. In that sense, diagnostic analytics uses traditional business intelligence and closely related to the descriptive analytics level. The third level, predictive analytics, explains what will happen. Predictive analytics comprise a variety of techniques that predict future outcomes based on historical and current data (Gandomi & Haider, 2015). The determination of demand levels can be taken as an example. Predictive analytics enables decision-makers to foresee future system behaviors. Predictive techniques allow decision-makers to look to the future, with knowledge of possible outcomes (Hazen, Skipper, Boone, & Hill, 2018). The final level is called as prescriptive analytics, which gives a solution for how to make it happen by using decision support systems and decision automation. Prescriptive analytics focused on decisions, while the other first three steps focused on information and insight (Delen & Zolbanin, 2018).

Figure 9. The steps of Big Data Analytics According to Gartner



Source: Laney, D., & Kart, L. (2012). Emerging Role of the Data Scientist and the Art of Data Science. Retrieved from https://www.gartner.com/doc/1955615?ref=mrktg-srch (date accessed: 20/07/2019).

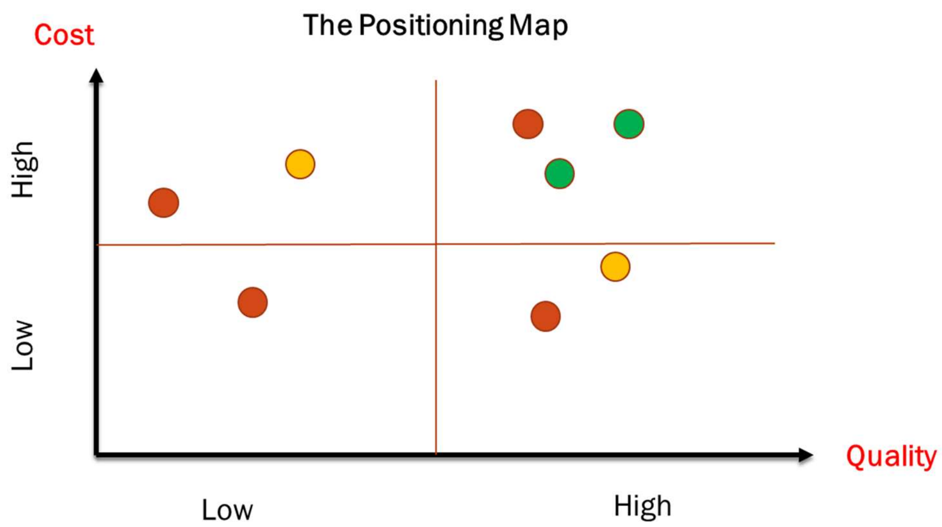
Tools for strategic market analysis and situation analysis are given in the next chapter. The common tools and their usages are discussed. The BCG Matrix, the selected model for this dissertation, and its relation with market strategy types are given.

1.4. Strategic Market Analysis (SMA)

Each company must find the game plan for long-run survival and growth that makes the most sense given its specific situation, opportunities, objectives, and resources. This is the focus of strategic planning the process of developing and maintaining a strategic fit between the organization's goals and capabilities and its changing marketing opportunities (Kotler & Armstrong, 2018). After defining the corporate-level mission and objectives purpose, the next step is to make an analysis of the situation. There are several approaches to conduct a situation analysis, where SWOT (Strengths, Weaknesses, Opportunities, and Threats) Analysis, Porter's Five Forces Analysis, are some well known methods. After getting a clear picture of the situation analysis, the next step is to formulate the priorities and moves of the marketing strategy. Priorities and moves are determined by the information obtained from the methods used in situation analysis (Miller, 1997).

SWOT Analysis helps companies to understand their businesses from a strategic view. It shows the advantages of opportunities, helps to avoid threats, and address weaknesses by using the company strengths (Shaw, 2012). Porter's Five Forces Analysis is used to understand the forces in the business environment or industry that can affect the company's profitability. It mainly focuses on competitive rivalry by looking at supplier power, buyer power, the threat of substitution and threat of new entry. Customer journey mapping is a way to visualize the user experience that helps when creating future marketing strategies (Kotler & Armstrong, 2018). It is mostly used in end consumer goods, services, and covers all story of the customer experience. It maps the steps the customer takes when engaging with a company. Positioning Map is used to compare products and determine their positioning in the market based on how customers perceive them (Wind, 1988). In positioning maps various dimensions are used such as Cost vs. Quality (Kotler & Armstrong, 2018), Brand Position vs Market Segments (Hassan & Craft, 2012), Supply Risk vs Profit Impact (Padhi, Wagner, & Aggarwal, 2012). Hassan & Craft (2012) used a positioning map to evaluate world market segmentation and brand positioning strategies with brand position vs market segment dimensions. Padhi, Wagner, & Aggarwal (2012) studied on a firm's purchasing strategy by using supply risk vs profit impact dimensions. Figure 10 shows an example of Cost and Quality Positioning map. During new product development process, a Cost and Quality positioning map can be used to see the gaps in the market.

Figure 10. Positioning Map



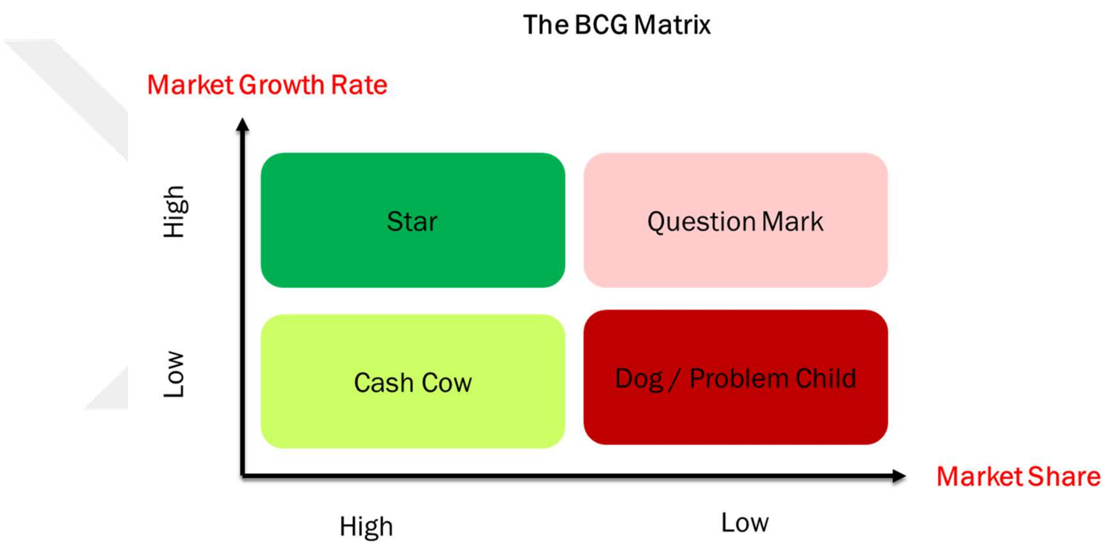
Source: Kotler, P., & Armstrong, G. (2018). Principles of marketing, Global edition (17th). London: Pearson.

As another method, Boston Consulting Group Matrix or Growth-Share Matrix is a framework for analyzing products according to growth and market share. It helps companies to make product portfolio analysis. Using the BCG Matrix, companies have an idea of what products they need to use to grow market share opportunities. Market growth rate and relative market share are two dimensions of BCG Matrix (Gite & Roy, 2014). Star, Cash Cow, Dog/Problem Child and Question Mark are the quadrant names in the BCG matrix. Figure 11 illustrates the BCG matrix.

In the BCG Matrix, stars are high market share and high market growth rate products. Therefore, companies are advised to invest in stars. Investment in those markets is crucial. High market share and low market growth rate are defined as Cash Cow. Some investment is recommended for cash cows to maintain a certain level of cash flow. Dogs are located in the low market share and low market growth rate quadrant. These product groups are prime candidates for downsizing (or even exiting the market). Low market share and high market growth rate define Question Marks. The matrix recommends investing in “question marks” if the product has the potential to move into stars (or divesting, otherwise). There are several examples of BCG matrix implementations. For example, Gite & Roy (2014) identified the export markets for the Indian carpet industry and classify them into four segments. Mutandwa et al.

(2009) applied the BCG matrix for coffee export marketing in Rwanda. Svatos, Smutka, & Miffek (2010) applied BCG matrix for commodity structure of agrarian trade of the EU countries. They investigated it on two levels that are the EU internal market and the world market. Ioana, Mirea, & Balescu (2009), used BCG Matrix to analyze service quality management in materials industry. Smith (2002), applied BCG Matrix in customer profitability analysis. Based on BCG Matrix, Calandro & Lane (2007) developed a new competitive analysis tool, with relative profitability and relative growth dimensions.

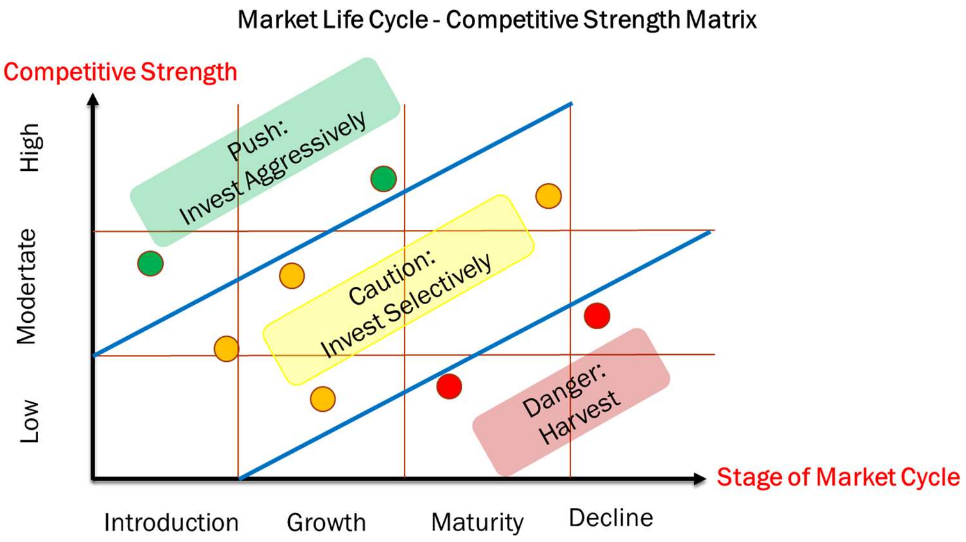
Figure 11 Boston Consulting Group Matrix



Source: Gite, P., & Roy, C. K. (2014). Export Markets ' Segmentation , Performance and Marketing of Indian Carpet Industry : A BCG Matrix Approach, 6(11): 28–33.

Market Life Cycle - Competitive strength matrix handles product portfolio from another perspective. Subjects, stage of the market life cycle and competitive strength are the angles at which it focuses. Due to its refined view, it can be a supportive tool for the BCG Matrix. Figure 12 shows a sample matrix. To extend the view, Industry Attractiveness and Business Position matrix can be used in the portfolio management process (Miller, 1997).

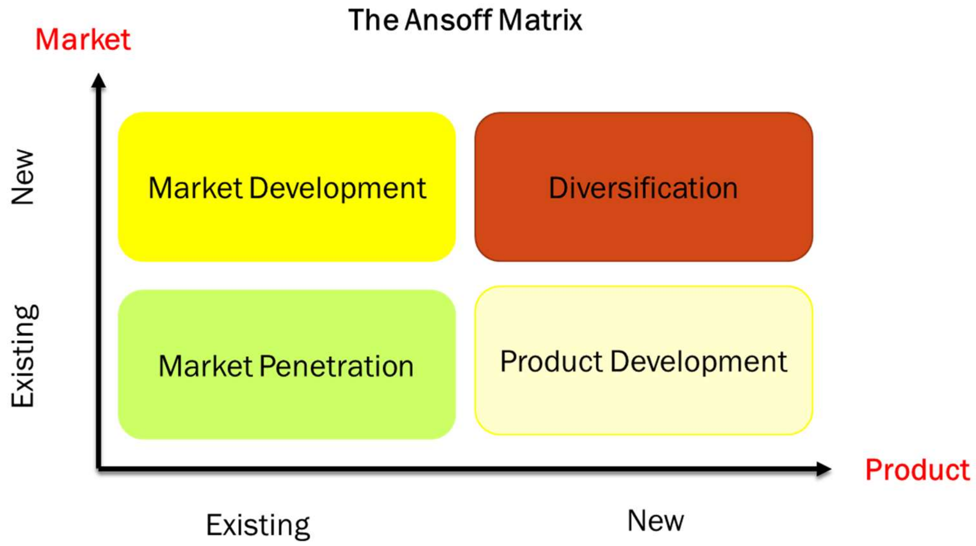
Figure 12. Market Life Cycle - Competitive Strength Matrix



Source: Miller, A. (1997). *Strategic Management: 3rd (Third) edition*. McGraw-Hill Companies.

Other than evaluating the current businesses, designing the business portfolio involves finding businesses and products the company should consider in the future. The Ansoff matrix (product/market expansion grid) is used to break down the market and products in two dimensions where the new market and existing market (Ansoff, 1957; Kotler & Armstrong, 2018). Ansoff matrix and its quadrants are shown in Figure 13. To break down a wide market into smaller similar bins, the market segmentation process is used. Aftermarket segmentation, the needs of small homogeneous groups of customers based on similar characteristics are identified.

Figure 13 The Ansoff Matrix



Source: Shaw, E. H. (2012). Marketing strategy: From the origin of the concept to the development of a conceptual framework. Journal of Historical Research in Marketing, 4(1): 30–55.

These mentioned methods here are the most important ones for positioning and portfolio analysis based marketing management.

There are three general strategies commonly used by businesses to reach and maintain a competitive advantage (Tanwar, 2013). Michael Porter has described these categories. These three generic strategies are defined in two dimensions: strategic market scope and strategic strength on competency. Strategic scope is a demand-side dimension where the size and composition of the market. Strategic strength is a supply-side dimension and looks at the strength or core competency of the firm. These three generic strategies are differentiation, cost leadership and focus strategies. Differentiation strategy is aimed at the broad market that involves the creation of a product or service that is perceived throughout its industry as unique. The cost leadership strategy emphasizes efficiency. In the focus strategy, the firm concentrates on a select few target markets. It is also called as a niche strategy. If the market and product are known then the marketing strategy should be built according to the evaluation of situation analysis.

There are also six main different strategy types in marketing (Getscher, 2017; Miller, 1997).

- Innovate: Disrupt the status quo
- Grow: Acquire new customers and increase market share
- Retain: Maintain market share and maintain customers
- Harvest: Try to minimize expenses and maximize profit
- Pause: Play out the current situation
- Exit: Terminate operation on that market and marketing support. Allow resources to be used somewhere else.

For this dissertation, among various strategic management models given above, BCG Matrix is the most suitable model for the data type obtained. Bilateral trade volumes, where export volumes from one country to another country is assumed as the market. And by using export volumes, market growth rate and relative market share can be calculated easily. While implementing the BCG Matrix, companies are advised to apply to grow strategy to Stars. For Cash Cows, a retain strategy would be implemented. Harvest or exit strategy can be applied in the Dogs quadrant. Grow or exit strategies can be applied according to the market response for Question Marks. According to the selected strategies, an action plan is made. Targets are set to realize the action plan and related marketing mix is prepared (Kotler & Armstrong, 2018). As the last step implementation of the plan and monitoring results by using selected methodologies, this cycle continues.

Next chapter categories in Machine Learning algorithms, problem types solved by these algorithms and typical workflow are given. Random Forest and Artificial Neural Network, the selected machine learning models for this dissertation, and their variations, principles and hyperparameters are discussed in detail.

1.5. Machine Learning (ML) Models

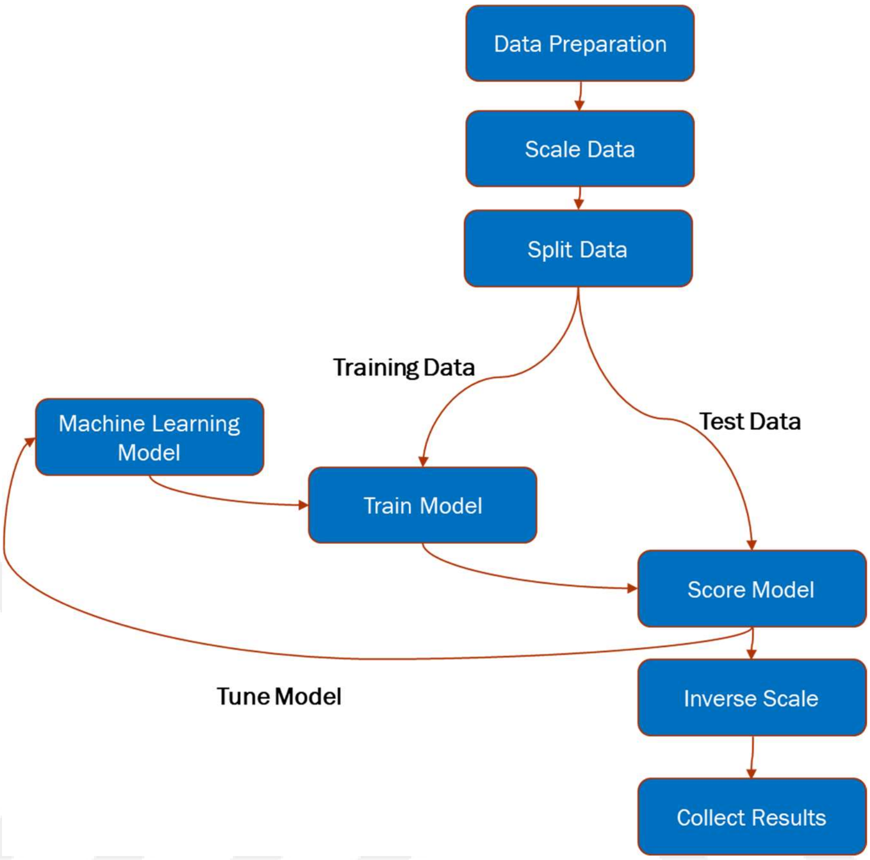
Machine learning is the science of getting computers to act without being explicitly programmed and the first definition of ML is made by Arthur Samuel. His formal definition stated that ML is programming computers to learn from experience, eventually eliminate the need for much of this detailed programming effort (Feigenbaum & Feldman, 1963). Arthur Samuel built the first self-learning computer program to play a game of checkers in 1954 (Weiss, 1992). His program became a better player after many games against itself. The program observed which moves were winning strategies and adapted its programming to incorporate those strategies.

Machine Learning is one of the most exciting technologies nowadays. As the name suggests, it gives the computer what makes it more similar to humans: the ability to learn. Both statistical and ML models are used for prediction and inference, however, statistical models have a long-standing focus on relationships between variables, ML models focus on the prediction by using general-purpose models to find in often rich and unwieldy data (Bzdok, Altman, & Krzywinski, 2018). ML algorithms may be classified into the following four main categories. Categories determined by mappings input data and anticipated output presented during the learning phase (Awad & Khanna, 2015; Mehmood & Graham, 2015; SAS, 2019):

- Supervised learning models are built according to observed data (independent data or input data) and a target variable (dependent data or output data) relationship. The learning mechanism based on this affects the relationship between the observed data and a target variable that is subject to prediction.
- Un-supervised learning models are designed to uncover hidden relationships in unlabeled datasets. In this models target variables are unknown. Clustering and dimension reduction are samples for un-supervised learning models (Awad & Khanna, 2015; C. P. Chen & Zhang, 2014).
- Semisupervised learning uses a combination of a small number of labeled and a large number of unlabeled datasets to generate a learning model. The labeling process of target data is expensive and impractical since it requires intensive skilled human labor(Awad & Khanna, 2015).
- Reinforcement learning (RL) uses three primary components: the agent, the environment and actions. The RL algorithms discover ways of an adaptive sequence of actions or behaviors done by the agent to maximize the cumulative reward. The objective is for the agent to choose actions that maximize the expected reward over a given amount of time. The agent can reach the goal much faster by following a good policy. Therefore, the goal of reinforcement learning is to learn the best policy (Awad & Khanna, 2015).

Figure 14 shows the basic flow for supervised learning (Awad & Khanna, 2015; C. P. Chen & Zhang, 2014). The first step in Machine Learning is the preparation of a dataset. Datasets have two components. The first component is the input dataset, in other words, features in ML world and independent variables in statistics language. The second component is output which is called a target variable in ML and dependent variable in statistics. According to the purpose of the ML algorithm transformation is sometimes needed. For example, time windowing for time-series problems or conversion of an image into pixel-wise data for image recognition problems. In case that each feature is in different ranges, magnitudes and units, the scaling process is applied. There are various scaling methods; minimum-maximum scaling, mean normalization and unit vector scaling are some of the common used scaling methods. After getting ML ready dataset, the next step is splitting the dataset into two-part. The first split is the training set and the second one is the test set. As a rule of thumb, between 70 and 80 percent of all dataset taken as training and the rest of the dataset taken as test dataset. To increase the success of ML algorithm training and test sets should be split carefully depending on the properties of data. For example, after splits presence of all types should distribute uniformly in test and training datasets in classification problems. In the training step, the ML algorithms look for the best model to obtain the outputs of the training dataset using the inputs of the training dataset. To get the score of the trained model, inputs of test dataset are put into the model and results are compared with test dataset outputs. To collect the real results from the model, inverse scaling should be done as the last step.

Figure 14. Typical Workflow for Supervised Learning



Source: Awad, M., & Khanna, R. (2015). Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. Apress.

ML divided into specific classes according to the type of problem to be solved. Some common problem types in ML are (Awad & Khanna, 2015; SAS, 2019):

- Regression: A supervised learning problem where the answer is a continuous value. This dissertation uses regression in its proposed big data analytics methodology.
- Classification: A supervised learning problem where the answer is one of the finitely many possible values.
- Segmentation: An unsupervised learning problem where the structure to be learned is a set of clusters of similar labels.
- Network analysis: An unsupervised learning problem where the structure to be learned is information about the importance and role of nodes in the network

Support vector machines, decision trees, association rule analysis or any type of artificial neural networks are some of the algorithms in order to derive knowledge from data reflecting conditions, processes and patterns (Stahlbock & Voß, 2010). The dataset of this dissertation contains monthly inputs and can be classified as a multi-variate time series. Because of their learning ability from complex relationships from these multi-variate data (S. Mishra, Mishra, & Santra, 2016), Artificial Neural Networks and Random Forests algorithms are selected as forecasting methods of this dissertation.

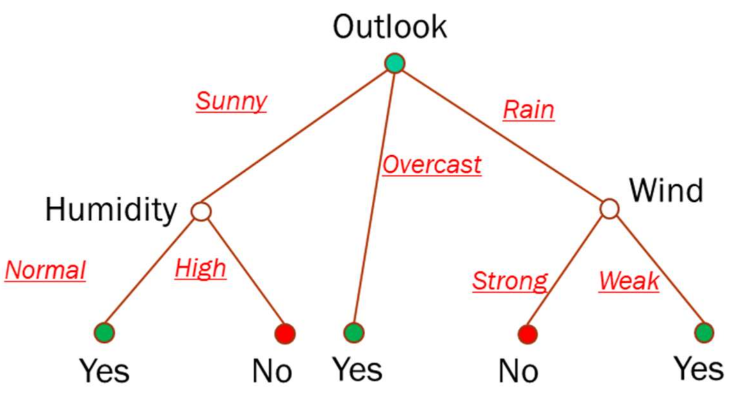
All machine learning algorithms depend on various parameters. There are two main groups. Some of the parameters are set before the training process. Parameters set before the training process are called hyperparameters of the ML model. Other parameters are calculated during the training process. Each ML algorithm has a different hyperparameter set depending on the structure of algorithm. To achieve more robust and successful results during training, a proper set of hyperparameters should be found. Instead of using the defaults of hyperparameters, a systemized tuning process should be applied During the hyperparameter tuning process, high R^2 scores and models robustness are aimed.

The two ML methods used in this dissertation are explained in Chapters 1.5.1 and 1.5.2

1.5.1. Random Forest (RF)

A decision tree is a decision support tool and also the basis for a tree predictor machine learning algorithms. It is a map of the possible outcomes of a series of related choices. It enables to give weight on possible actions against one another based on their costs, probabilities, and benefits. A decision tree usually begins with a single node branching out to possible consequences. Each of these results leads to additional nodes branching to other possibilities. This gives it a treelike shape. In Figure 15, a sample decision tree is given to give a decision on playing tennis.

Figure 15. Sample Decision Tree for Play Tennis

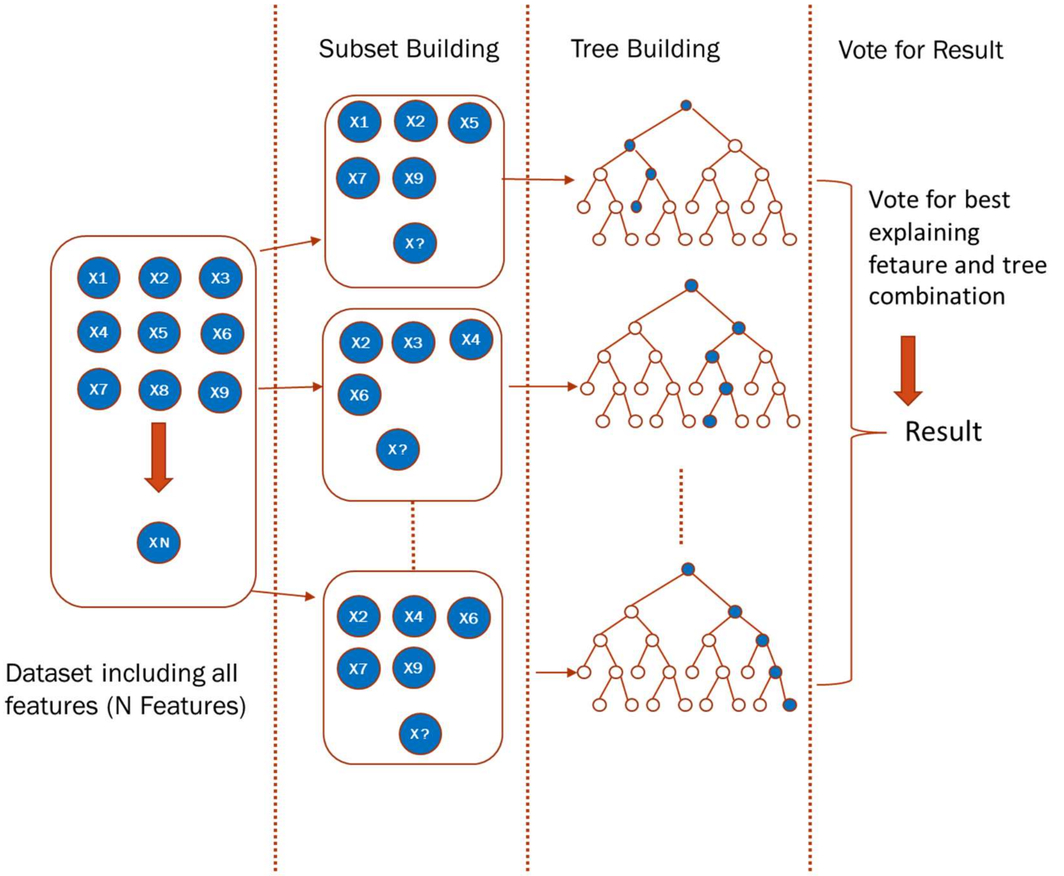


Source: Wagacha, P. W. (2003). Induction of decision trees. Nairobi. Retrieved from http://41.204.161.209/bitstream/handle/11295/44263/decision Trees.pdf ?sequence=1&isAllowed=y (date accessed: 20/07/2019).

Decision trees are simple to understand and visualize. Decision trees are supervised learning algorithms, that are widely used for both classification and regression. Moreover, they successfully work on multi-output problems. Non-linear relationships in features do not affect the performance of the algorithm. Despite the above advantages, the decision tree can create complex trees that do not generalize the structure of data. Prediction by using a single decision tree can cause an overfitting problem in the training set, which yields low-quality results in the test set. To avoid this overfitting due to the single dimension of randomness in a single decision tree, a random forest algorithm trains each tree with a random subset of the complete dataset.

Breiman (2001) first developed this ensemble learning approach by using different decision trees. With the inclusion of this second dimension of randomness by using random subsets, the random forest algorithm has the ability to reach high stability and robustness. This property allows using the best features among a random subset of candidate features (Figure 16). Building random subsets of features is the first step in the Random Forest algorithm. Forests use a random selection of features at each node to determine the split of trees. An important question is how many features to select at each node. Therefore, the RF algorithm is used not only in forecasting but also in the feature selection processes. In the random forest algorithm, after creating a large number of trees, the best descriptive combination selected within that trees (Breiman, 2001). Random forests have been used in many fields such as medical science, management science, and economics (Fang, Jiang, & Song, 2016). The RF applications are reported to yield successful results for small datasets with a relatively large number of features (Grömping, 2009). The general process of random forest algorithms explained in Figure 16.

Figure 16. The General Procedure of the Random Forest Algorithm



Source : Li, K., Yu, N., Li, P., Song, S., Wu, Y., Li, Y., & Liu, M. (2017). Multi-label spacecraft electrical signal classification method based on DBN and random forest. PLoS ONE, 12(5): 1–19.

For Random Forest, maximum features, minimum samples leaf, maximum leaf nodes, minimum weight fraction for leaf, minimum impurity decrease, number of estimators and random state are the hyperparameters used in this dissertation. The hyperparameters of Random Forest are explained and given in Table 1.

Table 1. The Tuned Hyperparameters of the Random Forest

Name of the Hyperparameter	**Hyperparameter Description**	**Search Space of the Hyperparameter**
Maximum Features	It represents the number of variables randomly sampled as candidates at each individual decision tree. The starting value for tuning is taken as “10”.	Four different values of Maximum Features parameters are selected for tuning. [10 ; "log2" ; "sqrt" ; "auto”]
Minimum Sample Leaf	It stands for the minimum number of samples in newly created leaves. The starting value for tuning for Minimum Sample Leaf is taken as “5”.	Five different values of the Minimum Sample Leaf parameter are are selected for tuning. [1 ; 2 ; 5 ; 20 ; 30]
Maximum Leaf Nodes	It denotes the maximum number of terminal node trees that a forest can have. The starting value for tuning for Maximum Leaf Node is taken as “100”.	Five different Maximum Leaf Nodes parameter are selected for tuning. [2 ;5 ; 10 ; 100 ; 200 ; 300]
Minimum Weight Fraction for Leaf	It refers to the threshold value for the minimum weighted fraction of the sum of all input weights in newly created leaves. The starting value for tuning for the Minimum Weight Fraction for the Leaf is taken as “0.00001”.	Five different values of Minimum Weight Fraction for Leaf parameter are selected for tuning. [0.00001 ; 0.0001 ; 0.001 ; 0.01 ; 0.1]
Minimum Impurity Decrease	It represents the threshold value for the impurity decrease achieved by node split. The starting value for tuning for Minimum Impurity Decrease is taken as “0.001”.	Three different values of Minimum Impurity Decrease parameter are selected for tuning. [0.000001 ; 0.00001 ; 0.001 ; 0.01]

Name of the Hyperparameter	Hyperparameter Description	Search Space of the Hyperparameter
Number of Estimators	It refers to the number of trees created in the forest by the algorithm. The starting value for tuning for estimator is taken as "200".	Seven different values of Number of Estimators parameter are selected for tuning. [100 ; 200 ; 500 ; 1000 ; 5000 ; 10000 ; 20000]
Random Seed	It refers to the Random Seed used during training sessions.	During the tuning process, each training session is held with a different random number. Therefore, the Random Seed is not included in the set of hyperparameters for tuning.

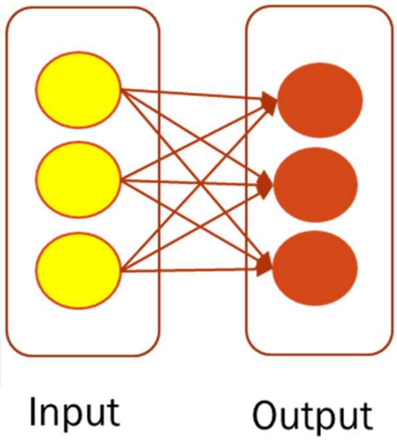
Source: Meinert, R. (2019). Towards Data Science - Optimizing Hyperparameters in Random Forest Classification. Retrieved July 20, 2019, from https://towardsdata science.com/optimizing-hyperparameters-in-random-forest-classification-ec7741f9d3f6 (date accessed: 20/07/2019). ; Scikit_Learn. (2019b). Random Forest Regressor. Retrieved July 20, 2019, from https://scikit-learn.org/stable /modules/generated/sklearn.ensemble.RandomForestRegressor.html (date accessed: 20/07/2019).

1.5.2. Artificial Neural Networks (ANN)

Artificial Neural Networks are based on the idea to mimic biological neural systems (Nummelin & Hänninen, 2016; Zhang, Patuwo, & Hu, 1998) by simple and connected processing nodes. Although the first application of ANN is dated back to 1964 with a weather forecast model (Zhang et al., 1998), it has applied in various areas for the last two decades including forecasting, credit scoring, financial analysis, and fraud analysis (C. P. Chen & Zhang, 2014; Tkáč & Verner, 2016). The core concept for ANN is based on a perceptron idea (Rosenblatt, 1958). Perceptron is the generic name given by the psychologist Frank Rosenblatt to a family of theoretical and experimental artificial neural net models which he proposed in the period 1957–1962

(Hunt, Minsky, & Papert, 2006). In basic perceptron topology, there are two layers as input and output. All neurons in the input layer connected to output layer neurons. Figure 17 shows the basic perceptron topology where the inputs are connected to output neurons.

Figure 17. Perceptron Topology

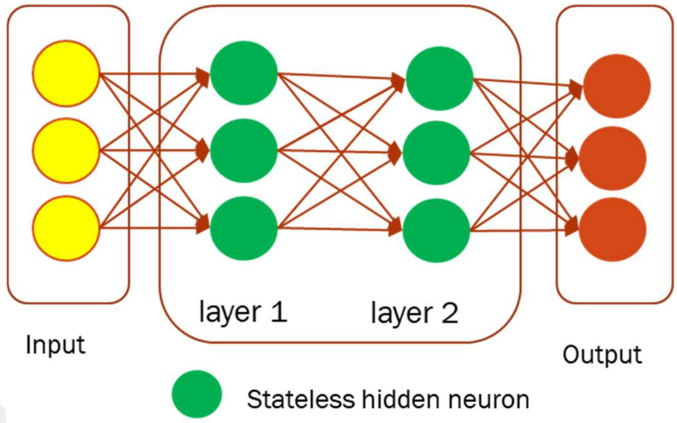


Source: Hunt, E., Minsky, M., & Papert, S. (2006). Perceptrons. In The American Journal of Psychology (Vol. 84, pp. 11218–11221).

ANN includes computational structures that are designed to mimic the biological central nervous system (Nummelin & Hänninen, 2016; Zhang et al., 1998). ANN is based on the accumulation of knowledge during training sessions. Due to the generalization ability coming from the knowledge accumulation attribute, the ANN can be used in any function approximation problem (Kaastra & Boyd, 1996). It is also a valuable tool for pattern recognition, classification and forecasting. There are various types of Neural Network topologies. Specifically, the multi-layer perceptron is used to refer to specific artificial neural network structures based on Rosenblatt's perceptron (Hunt et al., 2006). Multi-layer Perceptron (MLP), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Gated Recurring Units (GRU) are some of the well-known ANN topologies.

MLP topology (Figure 18) consists of three main layers: the input layer, hidden layer and output layer. The input layer transfers raw input data to the network. The number of nodes in the input layer depends on the number of features used in the model. The second layer in the network, hidden layer, consists of multiple layers and many nodes within them. After the hidden layer, the final solution is constructed in the output layer. They feed information from the front to the back (input and output, respectively) and the cells are stateless. They called Feed Forward Neural Networks as well.

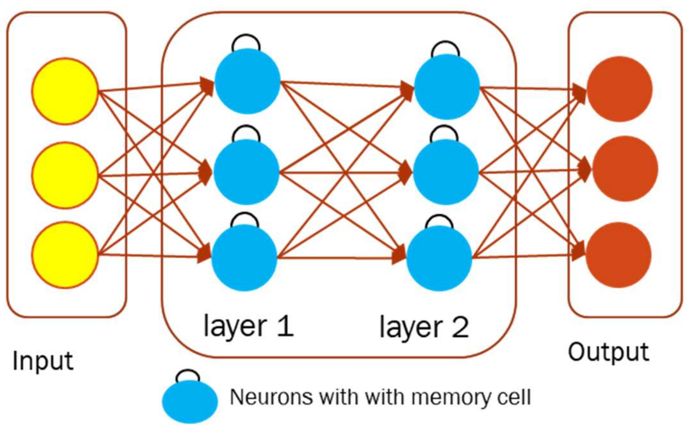
Figure 18. Multi-layer Perceptron Topology



Source: Allende, H., Moraga, C., & Salas, R. (2002). Artificial neural networks in time series forecasting: A comparative analysis. *Kybernetika*, *38*(6): 685–707.; Law, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management*, *21*(4): 331–340.

RNN is similar to MLP, with the addition of feedback information (Figure 19). The cells are not stateless; they have connections between passes, connections through time (Elman, 1990). Neurons are fed information not just from the previous layer but also from themselves from the previous passes.

Figure 19 Recurrent Neural Networks (RNN) Topology

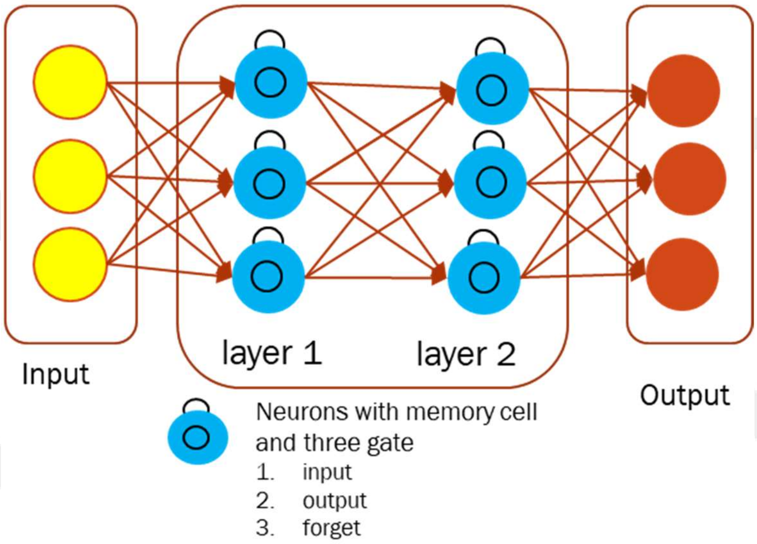


Source: Elman, J. L. (1990). Finding structure in time. *Cognitive Science*: *14*(2): 179–211.; The_Asimov_Institute. (2019). - The Neural Network Zoo.

Retrieved from http://www.asimovinstitute.org/neural-network-zoo/ (date accessed: 20/07/2019).

LSTM networks are inspired by circuitry, not from biology. Each neuron has a state and memory and three gates: input, output and forget (Sepp & Jurgen, 1997). The function of these gates is to control information flow. Figure 20 shows the LSTM topology.

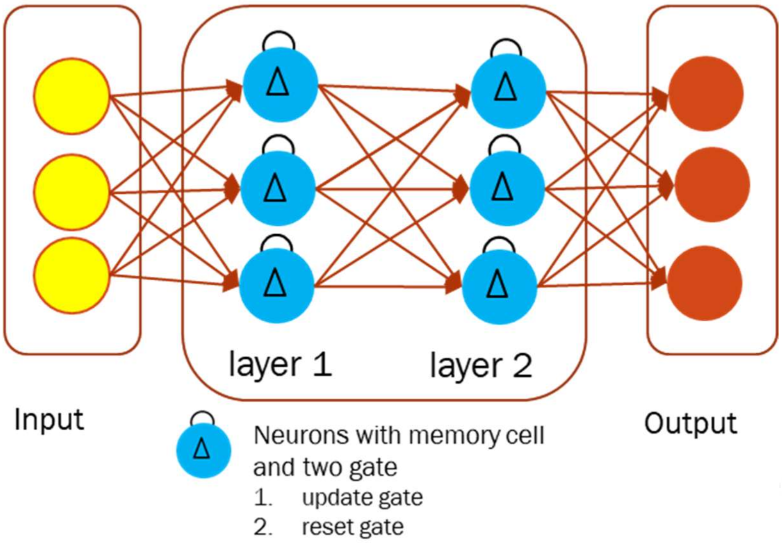
Figure 20 Long/short term Memory (LSTM) Topology



Source: The_Asimov_Institute. (2019). - The Neural Network Zoo. Retrieved from http://www.asimovinstitute.org/neural-network-zoo/ (date accessed: 20/07/2019).

GRU is another type of ANN where a slight variation of LSTM (Chung, Gulcehre, Cho, & Bengio, 2014). GRU has two gates instead of three gates in LSTM (Figure 21). They are the update and reset gates. Update gate decides the information to keep from the last state and how much information to let in from the previous layer. Reset gate behaves like LSTM's forget gate.

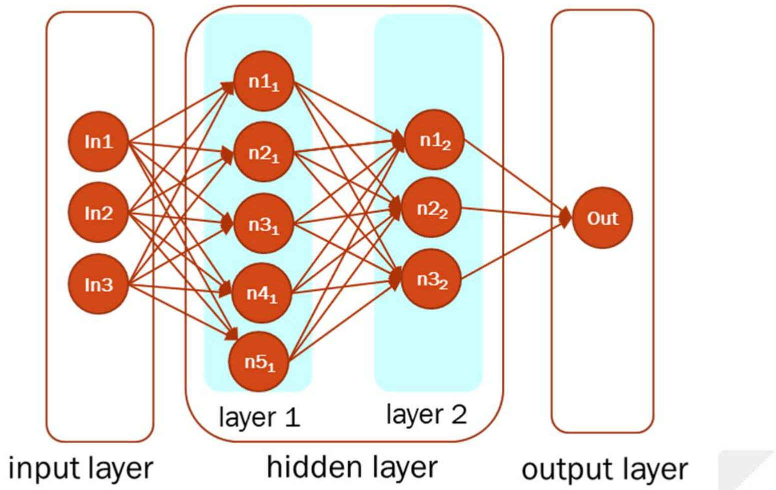
Figure 21. Gated Recurrent Units (GRU) Topology



Source: The_Asimov_Institute. (2019). - The Neural Network Zoo. Retrieved from http://www.asimovinstitute.org/neural-network-zoo/ (date accessed: 20/07/2019).

In this dissertation, MLP with the feedforward type selected because it is one of the most widely used topologies in forecasting and it has low resource requirements. Similar to the case study of this dissertation, Co & Boosarawongse (2007), Pakravan, Kelashemi, & Alipour (2011) and Nummelin & Hänninen (2016) used MLP to forecast various trade products. Figure 22 presents a Multi-Layer Perceptron topology with three inputs, hidden layers (with two layers) and a single output. Note that each node in MLP is fed from the nodes of the previous layers and they are fully connected.

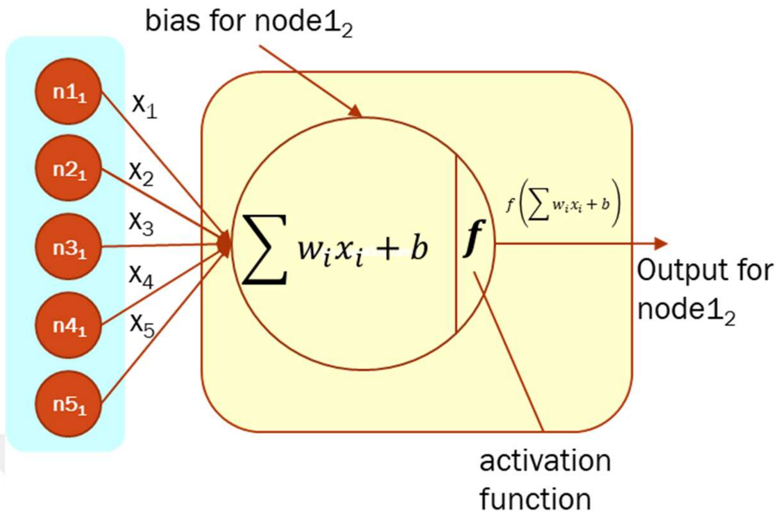
Figure 22. A Multi-Layer Perceptron Topology(Three Inputs, Five Neurons in First and three Neurons in Second Hidden Layer and a Single Output)



Source: Law, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management*, *21*(4): 331–340.

Each node in the ANN has four components: input, weight, bias, and activation functions. The input of each node is generated by the outputs of the previous layer nodes. Weight is the transformation factor of each input. Bias is a general factor in each node. The summation passes through an activation function, which decides how much of the information from this sum is the resulting output for that node. Non-linearity of ANN is provided by this activation function, which can be can be various types, such as logistics sigmoid or hyperbolic tangent. To demonstrate how the linear transformation is applied to the input values using weight and bias, Figure 23 shows the mathematical model of the first node of the second hidden layer.

Figure 23. Mathematical Model of the First Node of the Second Hidden Layer (n12) of Figure 22



Source: Chokmani, K., Khalil, B. M., Ouarda, T. B. M. J., & Bourdages, R. (2007). Estimation of River Ice Thickness Using Artificial Neural Networks. *Proceedings - 14th Workshop on the Hydraulics of Ice Covered Rivers, Quebec City, CGU-HS CRIPE*, (July 2014): 1–12.

While training (Training step in Figure 14) the ANN, the calculations are bi-directional. In the forward directional calculation move, the training dataset passes through the entire network and results in an output. This output is the network's prediction results. Then, the forecasts and real values of the training set are compared to make necessary adjustments in the model. Upon the calculation of the differences between the predicted and real values, the backpropagation step starts in the reverse direction. In this backward move, various types of solver methods are used to optimize weights and biases in each node. This process continues to optimize weights with these back and forth moves until the convergence achieved or a certain number of iterations is completed. Type of the solver used in the model, type of activation function used in the model, regularization parameter, the maximum number of iterations, topology of the network, learning rate and random state are the used hyperparameters of MLP in this dissertation. They are selected because they had more impact on the effects on the neural network model. These hyperparameters explained in Table 2 in detail.

Table 2. The Tuned Hyperparameters of the Artificial Neural Network model

Name of the Hyperparameter	Hyperparameter Description	Search Space of Hyperparameter
Solver Method Type	It denotes the Solver Method Type used for weight optimization in the system. The starting value for tuning for Solver Method Type is taken as "adam".	Three different values of Solver Type options are selected for tuning. ["lbfgs" ; "sgd" ; "adam"]
Activation Function Type	It represents the Activation Function Type used in hidden layers. The starting value for tuning for Activation Function Type is taken as "relu".	Four different values of Activation Type options are selected for tuning. ["identity" ; "logistic" ; "tanh" ; "relu"]
Alpha	It refers to the regularization parameter by giving a penalty to the network. The starting value for tuning for Alpha is taken as "0.001".	Three different values of Alpha parameter are selected for tuning.[0.001 ; 0.00001 ; 0.0000001]
Maximum Number of Iterations	It denotes the solver continues to optimize weights until to reach this number of iterations. The starting value for tuning for maximum iteration is taken as "200".	Four different values of Maximum Number of Iterations parameter are selected for tuning. [1000 ; 10000 ; 50000 ; 100000]
Hidden Layer Size	It represents the topology of the network in the hidden layer. It has two dimensions: (1) the number of layers and (2) the number of neurons in each layer. The starting value for tuning is taken as [30, 30] and it represents a topology with two hidden layers with 30 neurons on each layer.	Five different values of Hidden Layer topologies are selected for tuning. [(10, 10) ; (30, 30) ; (100, 100) ; (30, 30, 30) ; (30, 100, 30)]

Name of the Hyperparameter	Hyperparameter Description	Search Space of Hyperparameter
Learning Rate	This parameter changes the weight optimization learning rate. There are three different types of strategies in the SciKit Learn Library. The options in the library are 'constant', 'invscaling' and 'adaptive’.	This hyperparameter is valid only where the solver type is “sgd”. In our experiments, the “lbfgs” solver type performs better than the other solver types. The reason for this is that the used dataset in this dissertation has relatively a small number of observations. Therefore, the learning rate parameter is omitted from the hyperparameter tuning process.
Random Seed	It refers to the random seed used during training sessions.	During the tuning process, each training session is held with a different random number. Therefore, the random seed is not included in the set of hyperparameters for tuning.

Source: Scikit_Learn. (2019a). Multi Layer Perceptron. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLP Regressor.html (date accessed: 20/07/2019).

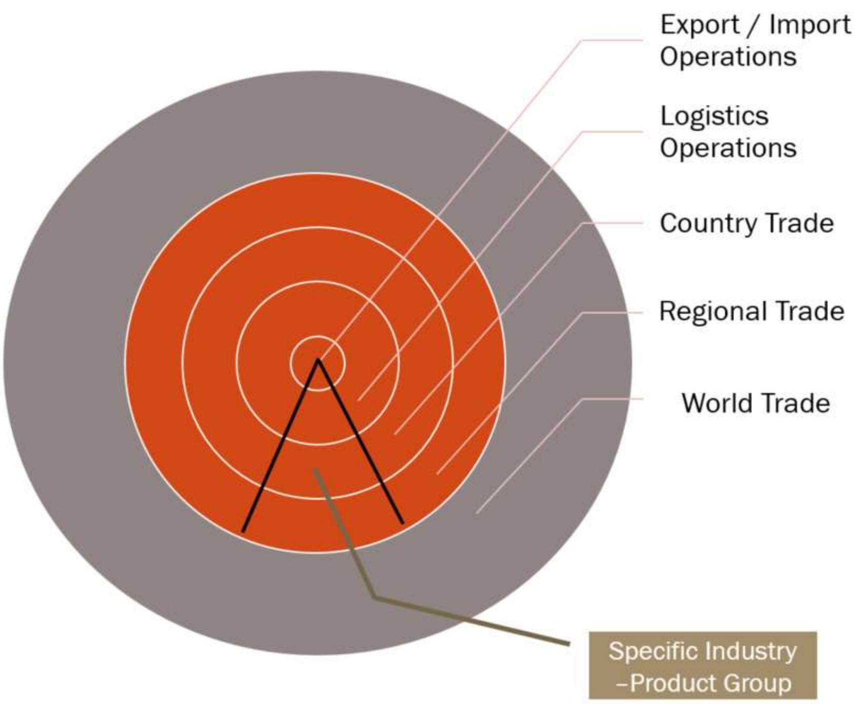
The following chapter gives the problem definition of this dissertation. It describes the process for the determination of the case, where Turkey’s trade data on the basis of the product group and exported country were presented. The selection process for product groups and a reference country other than Turkey is explained in detail.

2. PROBLEM DEFINITION

Today's competitive business environment, force business decision-makers to take the high volume of data into account for making their strategic decisions. The aim of this dissertation is to develop a Big Data Analytics approach to build a decision support system for the market analysis of export/import companies. The developed decision support system improves export/import companies' strategic market analysis by making accurate forecasts and creating foresight about future market situations. Both internal and external data are used by the proposed system.

World Trade is composed of different layers and components. The first layer is the regional level trade which is followed by country level trade level. Logistics operations, between countries and within a country, are included in the next level. Since exporters and importers are selected as the target group for this dissertation, export and import operations can be assumed as the last layer of the trading system. The scope of the dissertation is depicted in Figure 24. For the case studies of this dissertation, regional bilateral trade is applied for a specific industry/product group.

Figure 24. Scope of the Dissertation



With the recent global developments, Turkey becomes one of the main export partners of Europe. Since this dissertation is done in Turkey, Turkey is selected as the main exporter country. Chapter 84 (Machinery) exports are selected as the base for the case because it is closely related to the end-consumer behavior. Also, according to Turkey's export volumes (see Table 3), Chapter 84 (Machinery) is the second biggest export chapter after Chapter 87 (Vehicles). Within Chapter 84, the product group of refrigerators and freezers (HS=84.18) is the biggest product group according to the export volumes.

Table 3. Turkey's Export Volumes in Value According to Harmonized Codes (Thousand US Dollar)

HS Code			Product Label	2015	2016	2017	2018
87			Vehicles other than railway or tramway rolling stock, and parts and accessories thereof	17,463,564	19,801,974	23,940,852	26,759,684
84			Machinery, mechanical appliances, nuclear reactors, boilers; parts thereof	12,333,803	12,339,237	13,825,494	15,831,703
	8418		Refrigerators, freezers and other refrigerating or freezing equipment, electric or other; heat.	1,721,260	1,740,073	1,802,257	1,970,693
		841810	Refrigerators and freezers; combined refrigerator-freezers, fitted with separate external doors, electric or other	711,584	884,550	916,703	992,722
		841850	Furniture incorporating refrigerating or freezing equipment; for storage and display	231,983	241,678	259,520	295,578

HS Code			Product Label	2015	2016	2017	2018
		841840	Freezers; of the upright type, not exceeding 900l capacity	184,068	197,103	202,870	220,714

Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

As seen in Table 4, the United Kingdom is the biggest importer in HS=84.18 product group from Turkey in 2016. Therefore, for the bilateral trade forecasting case presented in this chapter, Turkey and the United Kingdom are selected as the source (exporting) and target (importing) countries, respectively. Another supportive reason to select the United Kingdom as the target country is that the UK is within the top ten importers list as given in Table 4. In this dissertation, the values for the year 2016 are taken as the reference year since 2016 is the last year that has full data when the research on this dissertation started. Therefore, product groups, importing country, reference country, top exporters to the importing country and other variables are selected by using 2016 as a reference.

Table 4. Turkey's Biggest Import Countries for Refrigerators and Freezers Product Group (HS 84.18) (Thousand USD)

Importer	Year			
	2015	**2016**	**2017**	**2018**
United Kingdom	222,308	212,341	200,071	224,844
Germany	188,426	202,314	211,439	220,487
United States of America	101,850	108,716	128,181	171,648
France	131,387	118,826	105,107	124,153
Italy	81,251	98,137	121,638	106,398

Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

The effectiveness of the proposed approach is validated in two ways. The first one is that it does not depend on a single exporting country. The second one is that it does not depend on a single product type. To build the case and validation conditions, three different product types in HS=84.18 group and one reference exporter country with similar products are investigated. Criteria for selecting products and reference country are given below:

- The reference country should be a stable exporter.
- Turkey and reference country should be the main exporter for selected products in common for the UK market
- Export volumes to the UK Market should be considerably high.

Thus, China is selected as a reference country according to satisfy the above criteria. As shown in Figure 2, China is also the largest exporter with 13 percent of total exports in the world. Therefore, competition between companies in China and Turkey is considerably high.

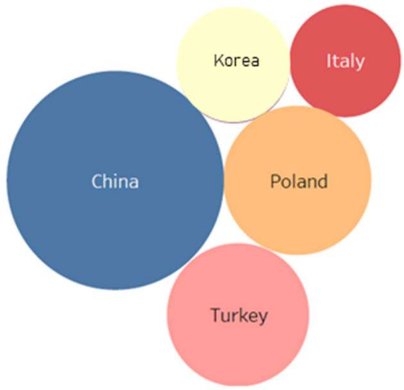
The three product groups selected in this dissertation are listed below.

- 84.18.10: Refrigerators and freezers; combined refrigerator-freezers, fitted with separate external doors, electric or other.
- 84.18.40: Freezers; of the upright type, not exceeding 900l capacity.
- 84.18.50: Furniture incorporating refrigerating or freezing equipment; for storage and display.

Table 3 shows Turkey's export volumes to the World for Chapter 87 (Vehicles) and Chapter 84 (Machinery) over the years. In addition to chapters, yearly trade values for the selected product groups in this dissertation are given. Figure 25, Figure 26 and Figure 27 show that Turkey and China are the main exporters for the selected product groups in the United Kingdom. For Product 84.18.10, there is a 31.1 percent increase in the total export market to the United Kingdom in five years. The exports of Turkey to the United Kingdom are increased by 25.2 percent in parallel to the increase of the market (31.1 percent). Meanwhile, the exports of China to the United Kingdom are increased by 70.9 percent, which is significantly more than the market increase. In 2016, the yearly change of the total market (-7.9 percent) and exports from China (-7.9 percent) are almost parallel. Turkey's exports are reduced slightly more than (-9.4 percent) the total market (-7.9 percent) in 2016. In addition to China and Turkey, Poland, Korea and Italy are the main exporters to the United Kingdom as of 2016.

Figure 25. Market Change and Bubble Graph of Top Five Exporters to the United Kingdom for Product 84.18.10

	2011-2016%	2015-2016 %
China	70,9%	-7,9%
Turkey	25,2%	-9,4%
Total Market	31,1%	-7,9%

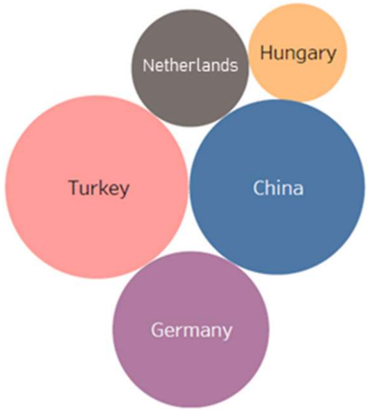


Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure 26 shows the export market changes in Product 84.18.40 to the United Kingdom. For product 84.18.40 there is a slight shrinkage (-2.1 percent) in the total export market to the United Kingdom in five years. The exports of Turkey to the United Kingdom market are decreased -25.4 percent and its faster than the total market decrease (-2.1 percent). Meanwhile, the exports of China to the United Kingdom are decreased -6.3 percent and it is more than market shrinkage, which means some other players fill this gap. In 2016, while the yearly change in the total market is -5.0 percent, exports from China are decreased to 14.6 percent and exports from Turkey are decreased -17.2 percent. The exports of both China and Turkey diminished more than the total market. In addition to Turkey and China, Netherlands, Germany and Hungary are the main exporters to the United Kingdom as of 2016.

Figure 26. Market Change and Bubble Graph of Top Five Exporters to the United Kingdom for Product 84.18.40

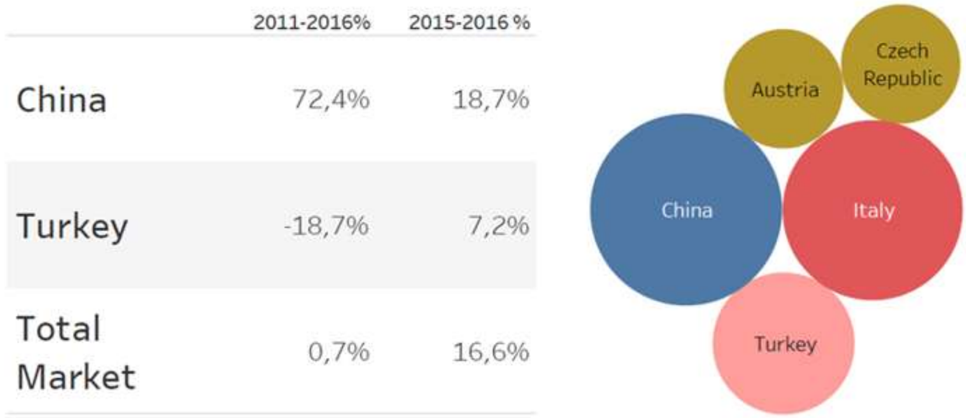
	2011-2016%	2015-2016 %
China	-6,3%	-14,6%
Turkey	-25,4%	-17,2%
Total Market	-2,1%	-5,0%



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

As shown in Figure 27, the market is in a steady position (0.7 percent) in the total export market of product 84.18.50 to the United Kingdom in five years. The exports of Turkey to the United Kingdom are decreased -18.7 percent in five years. Meanwhile, the exports of China to the United Kingdom are increased by 72.4 percent. This means that China took some market share from Turkey and from other players in this product group. In 2016, the total market grows 16.6 percent yearly. The exports from China are increased by 18.7 percent. Therefore, the total market change and China export change were almost parallel. Turkey's exports are increased by 7.2 percent in 2016, but it is not as large as the total market increase. China continued to obtain market share from Turkey and the other exporters. In addition to China and Turkey, Italy, Austria and the Czechzia are the main exporters to the United Kingdom as of 2016.

Figure 27. Market Change and Bubble Graph of Top Five Exporters to the United Kingdom for Product 84.18.50



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

The cases are established for products 84.18.10, 84.18.40 and 84.18.50 after the selection of main and reference exporting countries, Turkey and China, respectively. With two country pairs each represents one company and three product types, there are 6 different cases studied in this dissertation as given in Table 5.

Table 5. Implemented Cases

Case	**Country Pair**	**Product Groups**		
Main Case Turkish Company exports to the UK	Turkey – United Kingdom	84.18.10	84.18.40	84.18.50
Reference Case Chinese Company exports to the UK	China – United Kingdom	84.18.10	84.18.40	84.18.50

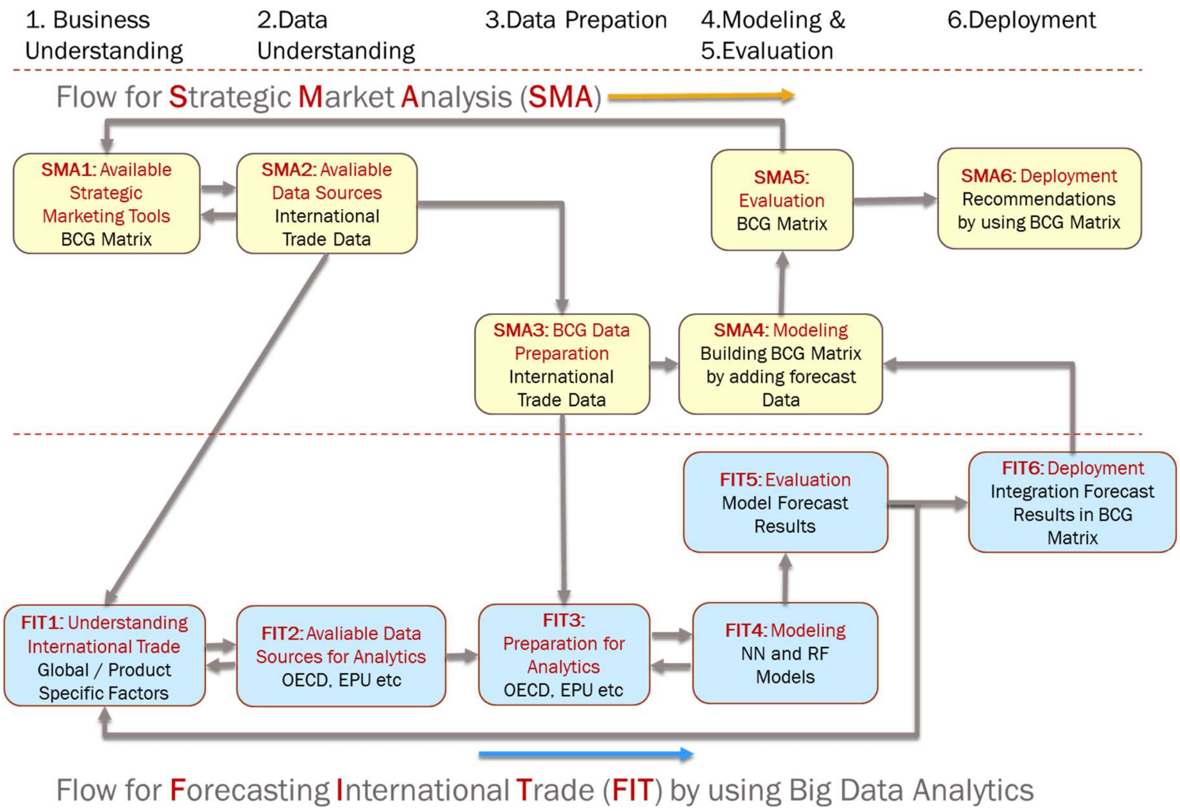
All steps of the proposed holistic approach for strategic market analysis approach are described in detail in the next chapter. The proposed approach is conducted in two levels: (1) strategic market analysis model, and (2) forecasting international trade. Cases given in Table 5 are applied to the methodology and resulting experiments and input dataset combinations are presented in the last part of this chapter.

3. METHODOLOGY

3.1. The Proposed Holistic Approach For Strategic Market Analysis

By using the steps of CRISP-DM (Wirth & Hipp, 2000) (i.e., business understanding, data understanding, data preparation, modeling and evaluation) as stages of the framework, the required steps to develop a holistic methodology for the market analysis using Big Data Analytics are added. As shown in Figure 28, the proposed methodology of this dissertation is conducted in two different levels in parallel: (1) Strategic Market Analysis (SMA), and (2) Forecasting International Trade (FIT). These two levels and their steps are explained in the following chapters in detail.

Figure 28. The Proposed Big Data Analytics Based Methodology for Strategic Market Analysis



3.1.1. Strategic Market Analysis Model

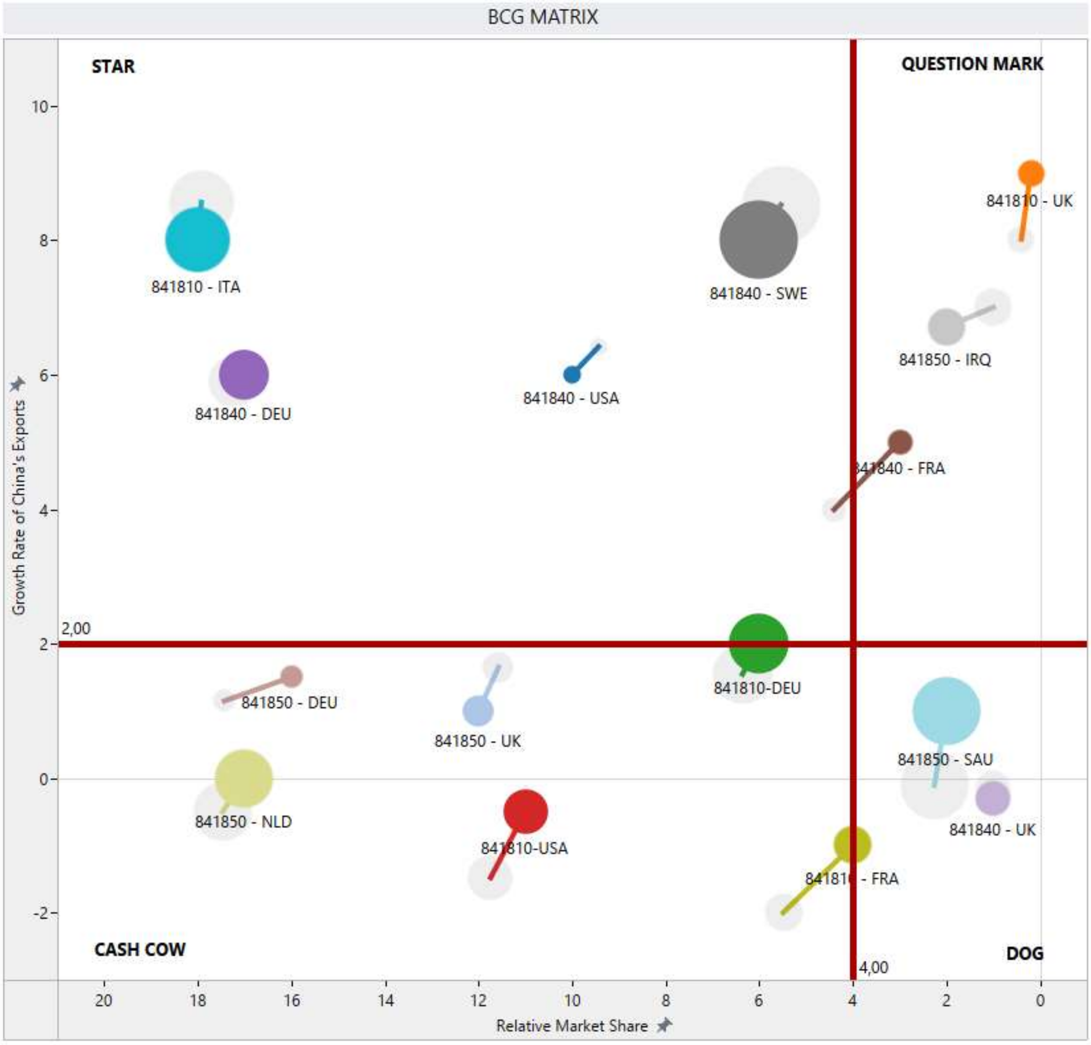
Exporters, importers and logistics service providers are considered as the potential users of the proposed methodology. These companies are exporting, importing or facilitating movements of products from one country to another. Their portfolios consist of product and country pairs. Successful portfolio management is important to be one step ahead in today's competitive business environment **(SMA1)**. International bilateral trade data between countries are publicly available and easily accessible by the companies. However, the key issue for the companies is to find a proper way of using these open international trade data in their decision-making process **(SMA2)**. International trade data are required for the case study in two ways. First, trade data give us the import market size of a country for a specific product at a certain time. Therefore, the market growth rate can be calculated using the data. Second, by combining the market size with the sales of the company, a company can derive its relative market share **(SMA3)**.

Among the various strategic management models in marketing, Boston Consulting Group Matrix is the most suitable model in this dissertation because it uses market growth rate and relative market share. The BCG Matrix allows managing the product portfolio in a company by showing opportunities in the market. Market growth rate and relative market share are the dimensions of the BCG Matrix (Kotler & Armstrong, 2018). Based on these two dimensions, product groups are classified into four quadrants: Stars, Cash Cows, Question Marks and Dogs **(SMA4)**. In Figure 29, an example of BCG Matrix is presented for a hypothetical company which exports three types of products to several countries. The values in Figure 29 are hypothetically generated in this dissertation to demonstrate the BCG matrix. Colored spots show the past data where these spot sizes are proportional to the sales volumes of companies. In the methodology, the BCG matrix is extended by adding forecast information about the market **(FIT1)**. The market forecasts of the products are depicted with the grey spots assuming that the company sales volumes remain the same as the previous period. Foreseeing three-month ahead has been set as the target to forecast **(SMA5)** **(FIT6)**. While building the BCG Matrix, a two percent growth rate of the market is taken as the separating value between stars and cash cows. For relative market share, four percent is taken as the threshold value between dogs and cash cows.

In the BCG Matrix, Stars are high market share and high market growth rate products. Therefore, companies advised investing in Stars. In the hypothetical example given in Figure 29, the export products of 84.18.40 to Sweden, the USA and Germany, and 84.18.10 to Italy are in the Stars quadrant. According to the forecasts (depicted as their grey shades), they would stay in the same quadrant in the near future (i.e., the next 3-6 months). Therefore, investment in those markets is crucial. High market share and low market growth rate are defined as Cash Cows. Some investment is recommended for Cash Cows to maintain a certain level of cash flow. This action could be applied to the export products of 84.18.50 to Germany, Netherlands and the UK, and 84.18.10 to the USA. Note that, the export product 84.18.10 to Germany has the potential to move from Cash Cows to Stars, which may require more investment like a Star than the regular Cash Cow. Dogs are located in the low market share and low market growth rate quadrant. These product groups are prime candidates for downsizing (or even exiting the market). In Figure 29, the export products of 84.18.50 to Saudi Arabia and 84.18.40 to the UK are in this quadrant. The export product of 84.18.10 to France is forecasted to move from Dogs to Cash Cows. Therefore, the market exit plan can be postponed until its relative market share decreases. Low market share and high market growth rate define Question Marks. The matrix recommends investing in "Question Marks" if the product has the potential to move into Stars (or divesting, otherwise). The export products of 84.18.10 to the UK and 84.18.50 to Iraq are Question Marks. The export product 84.18.40 to France has the potential for investment because of its movement to Stars quadrant.

Entering new markets to enrich the portfolio is another strategic decision. And, this corresponds to increasing the number of question marks in the matrix. As demonstrated by the example presented in Figure 29, the BCG Matrix covers all product types and countries, which makes it easy to identify the current market conditions of the newly entered market. In today's rapidly changing business world and with the constant flow of big data, product portfolios can be evaluated more frequently, even monthly, to identify current trends promptly and act accordingly **(SMA6)**. The proposed methodology helps to identify future trends and market situations by adding forecasted information to the BCG matrix.

Figure 29. A Sample BCG Matrix for a Hypothetical Company (Colored spots show the current information and the grey spots show the forecasted information)



Source: Data for this Sample BCG Matrix Randomly Generated.

The forecast horizon for BCG Matrix is chosen as three-month. The main issue to use the BCG matrix is to determine the growth rate because the increase (or decrease) from a specific month to the next one can be very large. Since the monthly fluctuations of the trade volume can be erratic, it can cause artificially high (or low) growth rate calculations, and therefore, mislead the BCG Matrix. To overcome this problem, the growth rate is calculated using the averages of the two consecutive quarters (the period of three-month) as explained in the Results Chapter in detail.

3.1.2. Forecasting International Trade

To accurately forecast international trade, the dynamics and potential factors affecting bilateral trade between countries must be identified first **(FIT1)**. This is the business-understanding step of the trade forecasting. After doing a literature survey and conducting the appropriate research, the candidate factors are identified. These factors come from various domains, such as trade, economy, business and politics. Since the aim of this dissertation is to forecast the bilateral trade volume, there are some constraints for determining factors. In this dissertation, to have enough training and test sets, more than 10 years of data are included.

These factors can be grouped into two main groups: (1) Product-specific trade information, and (2) Country or global conditions related features. The main components affecting bilateral trade are the supply and demand factors of the related countries (Ayankoya et al., 2016; Kangas & Baudin, 2003; Nummelin & Hänninen, 2016). To model the demand for a specific product in a target country (in this case study, the United Kingdom), the volumes of the top five exporters to the target country are considered. To model the supply from the Exporting Country (in this case study, Turkey and China), the trade volumes of the top five target countries that the source country exports are taken into account. As the last product-specific factor, the unit value of the traded product is included in the model. The country or global factors are divided into the subgroups of the political environment, business environment (Bovi & Cerqueti, 2016), economic environment (Keck, Raubold, & Truppia, 2010; Sokolov-Mladenović et al., 2016) and trade environment-related factors. The business environment is included by adding Business Confidence and Consumer Confidence Indicators. With the inclusion of the Economic Political Uncertainty Index, political factors are covered. Economic factors are Gross Domestic Product (GDP), Exchange Rates, Composite Leading Indicators, Consumer and Producer Price Indices. World Trade Volume and World Economic Political Uncertainty indices are included as global trade parameters. Table 6 summarizes all these factors used in the forecasting model **(FIT2)**. To obtain these data, four different open data sources are used and they are also given in the subgroup column of Table 6. The trade model notations used in this dissertation are based on the commonly acceptedworld trade model notations (Marwah, 1976) and they are listed as follows:

$X^t{}_{ijk}$ = Trade flow volume from source (exporting) country *i* and target (importing) country *j* for product *k* in period *t*

i_0 = Exporting Country where $i_0 \in$ {Turkey and China}

j_0 = Importing Country (United Kingdom)

K = Products, where $K \in \{84.18.10, 84.18.40, 84.18.50\}$

I_{j_0k} = The top five Exporting Countries for j_0 for product k $\forall$ K

J_{i_0k} = The top five Importing Countries for i_0 for product k $\forall$ K

There are two different trade data reporting the trade volumes. These are export data reported by the exporter country and import data reported by the importer country. These two data do not fully match with each other due to the differences in export and import data-keeping procedures of the countries. In this dissertation, to be consistent, the trade data reported by the exporter countries are used. All product-specific data used in this dissertation are given as line plots in Appendix 1. All global data used in this dissertation are given as line plots in Appendix 2. To have insight into the relation between target data and other export volumes the related pair plots are given in Appendix 3.

The dataset is examined according to intrinsic data dimensions (Wigan & Clarke, 2013). Intrinsic dimensions are accuracy, timeliness, consistency, completeness and frequency. The first concern is the accuracy and 10 years of data are reachable from reliable open data sources. The next one is timeliness, and the start and end dates of the data are around 2007 and 2017, respectively. Completeness is the next dimension, and it means that data do not have any missing points during that period. The consistency dimension is the expectancy to be on the same unit or at least convertible to the same unit. As the frequency dimension, the monthly period is selected in model and data served in monthly periods are preferred. After passing all these intrinsic data dimension constraints, the features used in the datasets are determined. Features given in Table 6 are the final list of features passing these intrinsic data constraints. Units of the used features are;

- Export Capacity, Total Export Capacity and Import Capacity are in US Dollars
- Unit Value change per kg. Given for each Product is US Dollar/kg

- Economic Policy Uncertainty, Business Confidence Indicator, Consumer Confidence Indicator, Composite Leading Indicator Producer Price Indicator, Consumer Price Indicator are adjusted indices
- GDP is US Dollars
- Currencies are in daily exchange rates with respect to US Dollars
- Total World Trade is in Volume

Table 6. Features (Factors) Used in the Forecasting Model

Main Group	Subgroup	The formula for the Feature	The Similar Studies Used the Same Feature
Product Specific	Trade Information (Source: International Trade Center)	Export Capacity of the Exporting Country for Product k: $X^{t}{}_{i\ jk}, \forall j \in J_{i_0k}, k \in K$	Predicting grain price (Ayankoya et al., 2016), Predicting bilateral trade (Nummelin & Hänninen, 2016)
		Total Export Capacity of the Exporting Country for Product k: $X^{t}{}_{i_0 k}, \forall k \in K$	
		Import Capacity of the Importing Country for Product k: $X^{t}{}_{i j_0 k}, \forall i \in I_{j_0 k}, k \in K$	Predicting grain price (Ayankoya et al., 2016)
		Unit Value of the Product from Exporting Country to Importing Country: $XUV^{t}{}_{i_0 j_0 k}, \forall k \in K$	
Global	Political Environment (Source: Economic Politic Uncertainty)	Economic Policy Uncertainties (EPU) of Exporting and Importing Countries, and the World: $EPU^{t}{}_{i}, \forall i \in \{i_0, j_0, World\}$	
	Business Environment (Source: OECD)	Business Confidence Indicators (BCI) of Exporting and Importing Countries: $BCI^{t}{}_{i}, \forall i \in \{i_0, j_0\}$	Predicting macroeconomic fundamentals (Bovi & Cerqueti, 2016)
		Consumer Confidence Indicators (CCI) of Exporting and Importing Countries: $CCI^{t}{}_{i}, \forall i \in \{i_0, j_0\}$	
	Economic Environment (Source: OECD)	Composite Leading Indicators (CLI) of Exporting and Importing Countries: $CLI^{t}{}_{i}, \forall i \in \{i_0, j_0\}$	
		Gross Domestic Products (GDP) of Exporting and Importing Countries: $GDP^{t}{}_{i}, \forall i \in \{i_0, j_0\}$	Predicting bilateral trade (Nummelin & Hänninen, 2016), Predicting GDP (Sokolov-Mladenović et al., 2016), Predicting

Main Group	Subgroup	The formula for the Feature	The Similar Studies Used the Same Feature
			import volumes (Keck et al., 2010)
		Producer Price Indices (PPI) of Exporting and Importing Countries: $PPI^t{}_i, \forall\, i \in \{i_0, j_0\}$	
		Consumer Price Indices (CCI) of Exporting and Importing Countries: $CPI^t{}_i, \forall\, i \in \{i_0, j_0\}$	Predicting import volumes (Keck et al., 2010)
		Exchange Rate of Currencies of Exporting and Importing Countries to USD $EXC^t{}_i, \forall\, i \in \{i_0, j_0\}$	As currency exchange rates (Ayankoya et al., 2016), Predicting bilateral trade (Nummelin & Hänninen, 2016)
	Trade Environment (Source: CPB World Trade Monitor)	Total World Trade Volume: $World\_Trade^t$	

Datasets from different sources are needed to be formatted into a single format. In addition, multiple entries are cleaned, and data are trimmed to exclude missing values. Time windowing is done by shifting the data points one through eight months. After time windowing, the dependent variable datasets and independent variable datasets are prepared for each product to determine the forecast horizon. All data combinations are created to find the best set of dependent variables from past trade data. Data labels standardized for easy understanding. For example, CHN_UK_841810 -3 stands for the export volume of China to the United Kingdom for 84.18.10 product where the values are shifted months back **(FIT3)**. These combinations are based on monthly data as listed in Table 7. For each forecast horizon, three input month combination is taken into account. The first one is single month input and it uses the shifted month data same as forecast horizon. The second one uses the data of two consecutive months, shifted the same as forecast horizon. The third one takes three consecutive month information into account starting with a forecast horizon shift. For example, the combination of the two-month forecast horizon and two input month uses the data of the previous two and three months.

Table 7. The Relation Between Forecast Horizon and Input Month Combinations

Input month combinations	Forecast Horizon					
	1 month forecast	2 months forecast	3 month forecast	4 month forecast	5 month forecast	6 month forecast
Only one	Only (-1)	Only (-2)	Only (-3)	Only (-4)	Only (-5)	Only (-6)
Double Months	(-1),(-2) months together	(-2), (-3) months together	(-3), (-4) months together	(-4), (-5) months together	(-5), (-6) months together	(-6), (-7) months together
Triple Months	(-1), (-2), (-3) months together	(-2), (-3), (-4) months together	(-3), (-4), (-5) months together	(-4), (-5), (-6) months together	(-5), (-6), (-7) months together	(-6), (-7), (-8) months together

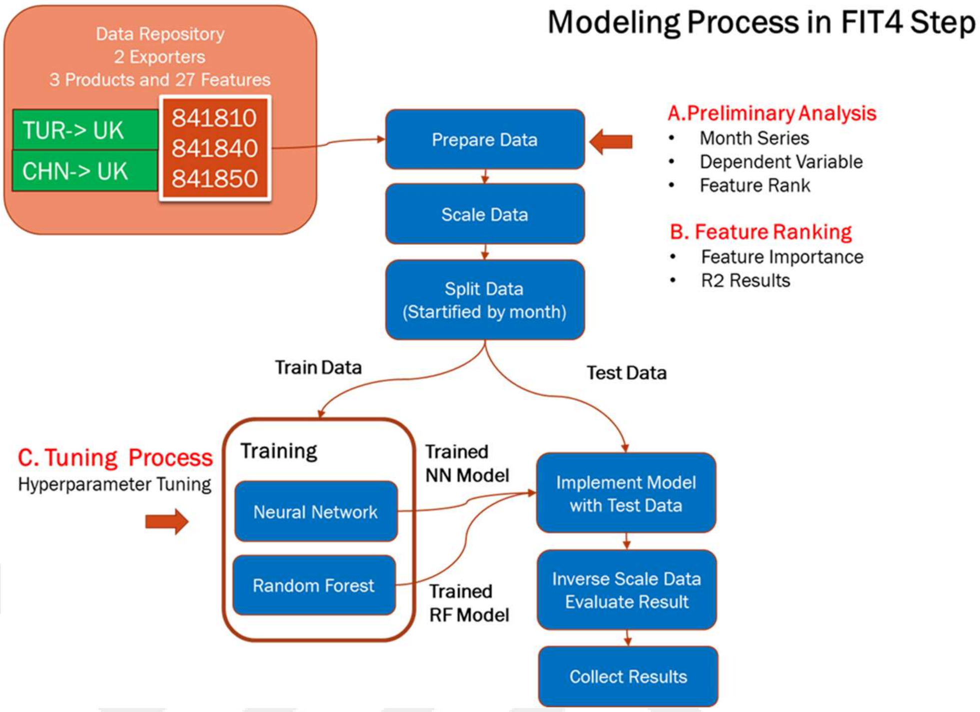
Table 8 represents the time windowing process for Product 84.18.10 from China to the UK (CHN_UK_841810). It is done for three different months (3-month shift, 4-month shift and 5-month shift) with two variables. One variable is the same as the target variable and labeled as CHN_UK_841810. The other one is China's total export to the World on that Product and labeled as CHN_World_841810. Target data belonging to the month 200703 is given as "5818" in Table 8. The input dataset consists of data from three-month, four-month and five-month lagged values. Input data for (-3) month shift comes from date 200612 with "9841" for CHN_UK_841810-3 label and "56805" for CHN_World_841810-3. Input data with (-4) month lagged month, comes from date 200611 for CHN_UK_841810-4 label as "7766" and for CHN_World_841810-4 as "66038". Input data with (-5) month shift comes from 200610 is "12617" for CHN_UK_841810-5 label and "82839" for CHN_World_841810-5. The same process is applied for all features used in the model to build input datasets. For example, the combination of three-month horizon forecast and with only three-month lagged input set uses "5818" as the target for month 200703. Inputs for CHN_UK_841810-3 (its value is 9841) and for CHN_World_841810-3 (its value is 56805) come from 3-month lagged information and they are used in the ML model.

Table 8. Sample for Time Windowing Process (Product 84.18.10 from China to the UK) with -3, -4 and -5 Shift on Two Variables

	Target Data	**Sample Input Data**					
		3 Month Shift		4 Month Shift		5 Month Shift	
Date	CHN_UK_ 841810	CHN_UK_ 841810 -3	CHN_World _841810 -3	CHN_UK_ 841810 -4	CHN_World _841810 -4	CHN_UK_ 841810 -5	CHN_World _841810 -5
200609	8070	7874	66944	5824	54651	2323	42475
200610	12617	7287	57512	7874	66944	5824	54651
200611	7766	14050	70024	7287	57512	7874	66944
200612	9841	8070	79249	14050	70024	7287	57512
200701	4094	12617	82839	8070	79249	14050	70024
200702	7378	7766	66038	12617	82839	8070	79249
200703	5818	9841	56805	7766	66038	12617	82839
200704	5785	4094	52470	9841	56805	7766	66038
200705	5587	7378	62917	4094	52470	9841	56805
200706	5617	5818	81204	7378	62917	4094	52470
200707	8591	5785	84891	5818	81204	7378	62917
200708	11770	5587	86921	5785	84891	5818	81204
200709	7401	5617	101190	5587	86921	5785	84891
200710	6729	8591	108648	5617	101190	5587	86921
200711	9328	11770	110799	8591	108648	5617	101190
200712	5935	7401	108539	11770	110799	8591	108648

Artificial Neural Networks and Random Forests algorithms are selected as forecasting models **(FIT4)**. The typical workflow for ML (Figure 14) is customized for this dissertation. The detailed process is given in Figure 30 for these ML models Preliminary analysis, feature ranking and tuning are the main steps in the applied workflow. Details of these steps are given in Chapter 4.1.

Figure 30. Machine Learning Process Followed in this Dissertation



After building the model, the forecasted demands of products are found by using the Random Forest and Artificial Neural Network algorithms and then are used as market data forecast **(FIT5)**. The sales forecasting of the company is simply assumed to be the same as the previous period to demonstrate the status of the portfolio without taking any action according to changing market conditions. The market forecast and company sales forecast for the next three-month are the required information to calculate grey spots in the BCG Matrix **(FIT6)**. The results of the case study are presented in Chapter 4.1 including a preliminary analysis of data preparation and the process of hyperparameter tuning.

In the next chapter, the combinations of cases, implemented experiments and combinations of input datasets used in this dissertation are given.

3.2. Implemented Experiments

This chapter describes the implemented experiments and studied datasets more clearly. With two country pairs each represents one company and three product types, there are 6 different cases studied in ML models. Table 9 represents the studied cases in the ML model.

Table 9. Implemented Cases in Machine Learning Model

Case	Country Pair	Product Groups		
Main Case Turkish Company exports to the UK	Turkey – United Kingdom	84.18.10	84.18.40	84.18.50
Reference Case Chinese Company exports to the UK	China – United Kingdom	84.18.10	84.18.40	84.18.50

In the preliminary analysis, 162 different input combinations are studied for each case, which means that a total of 972 (162 combinations * 6 country/product pairs) different datasets are created and passed through the preliminary analysis phase. The best combinations for each case are selected to use in the tuning process. Table 10 explains these cases and input dataset combinations.

Table 10. Implemented Cases and Input Dataset Combinations

Preliminary Analysis (Total:972 Different Input Combinations)	**Forecast Horizon**	1 month	2 month	3 month	4 month	5 month	6 month
	Input: Format of Target variable	Same		Log		Sqrt	
	Input: Used Percentage of Features	50 percent		75 percent		100 percent	
	Input: Used Month Series as Input	One lagged month		Two consecutive lagged month		Three consecutive lagged month	

In a dataset used in machine learning, each column in the dataset represents a feature and each row represents a member of the dataset. The model has 28 features as given in Table 6. Depending on the preliminary analysis the size of dataset changes. The size of a dataset depends on two factors, the first one is the length of the month series and the second one is the percentage of features used in the model. The relation between forecast horizon and input month combinations are given in Table 7. According to input month combinations and used percentage of features, Table 11 represents the number of features in datasets.

Table 11. The Number of Features in Datasets Depending on Preliminary Analysis Results

Used Percentage of Features	**50 percent**	**75 percent**	**100 percent**
Used Input month combinations			
Single - One lagged month	14 Features	21 Features	28 Features
Two - Two consecutive lagged months	28 Features	42 Features	56 Features
Three - Three consecutive lagged months	42 Features	63 Features	84 Features

Fluctuations in export volumes are so high. Also, one month ahead forecast horizon is too short and six months is too long to make decisions. Therefore, a three-month horizon forecast is selected to implement in the BCG matrix for the proposed methodology of this dissertation.

The best hyperparameter values for each case given in Table 9 are achieved by tuning step (see Chapter 4.1.3). For each case, 11 different hyperparameters (6 for RF and 5 for ANN) are tuned. In total 66 different tuning steps are done. Table 12 shows tuned hyperparameters for each implemented case in the machine learning model.

Table 12. Implemented Cases and Tuning Steps

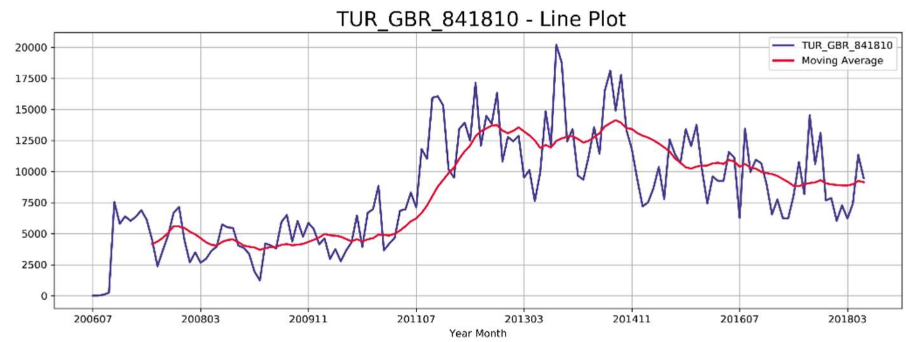
Machine Learning Method / Tuned Hyperparameters		
Tuning Process (Total:66 Tuning Step)	**Random Forest Parameters and Search Space**	**Multilayer Perceptron Parameters and Search Space**
	6 Hyperparameters: • maximum features [10 ; "log2" ; "sqrt" ; "auto"], • minimum samples leaf [1 ; 2 ;5 ;20 ; 30], • maximum leaf nodes [2 ;5 ;10 ; 100 ; 200 ; 300], • minimum weight fraction for leaf[0.00001 ; 0.0001 ; 0.001 ; 0.01 ; 0.1], • minimum impurity decrease [0.000001 ;0.00001 ; 0.001 ; 0.01], • number of estimators [100 ;200 ; 500 ;1000 ;5000 ; 10000 ; 20000]	5 Hyperparameters: • solver method [“lbfgs” ; “sgd” ; “adam”], • activation function type [“identity” ; “logistic” ; "tanh” ; “relu”], • regularization parameter [0.001 ; 0.00001 ; 0.0000001], • maximum number of iterations [1000 ; 10000 ; 50000 ; 100000], • the topology of the network [(10, 10) ; (30, 30) ; (100, 100) ; (30, 30, 30) ; (30, 100, 30)],

The next chapter gives results of the methodology given in this chapter. Since the proposed approach is conducted in two levels, results are given two subchapters. In first chapter, machine learning modelling process results of the experiments given above, explained in detail. The second chapter explains the results of the proposed methodology and generation of the BCG Matrix.

4. RESULTS

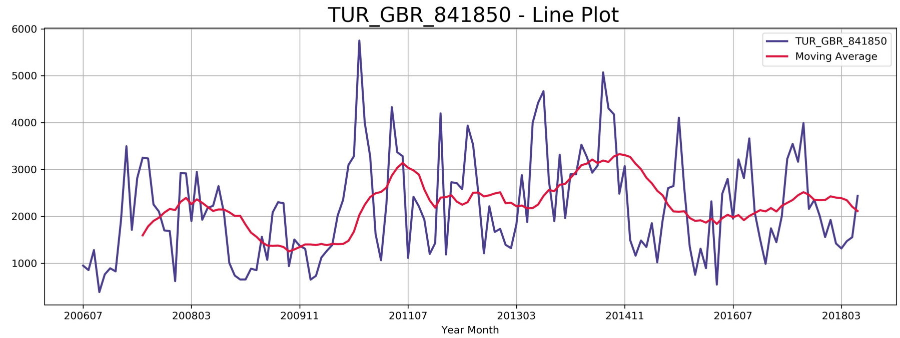
The hypothetical Turkish and Chinese case study companies are assumed to export refrigerators and freezers to several countries. They are mainly exporting three types of products, where the product selection process for the hypothetical companies is explained in detail in Chapter 2. One of the companies is assumed to be exporting from Turkey the other one from China. To demonstrate the proposed methodology, only the export volumes to the United Kingdom are forecasted. The dataset used in the forecasting model build from different open data sources. These data sources are given in the "Subgroup" column of Table 6. The dataset starts in March 2006 and ends in March 2018. It has 144 observation data. To give an insight about the dataset, time series plot from Turkey for products 84.18.10, 84.18.40 and 84.18.50 are given in Figure 31, Figure 32 and Figure 33, respectively. The moving average values through the months are indicated with red lines.

Figure 31. Export Volumes of the Product 84.18.10 from Turkey to the UK from July 2006 to March 2018



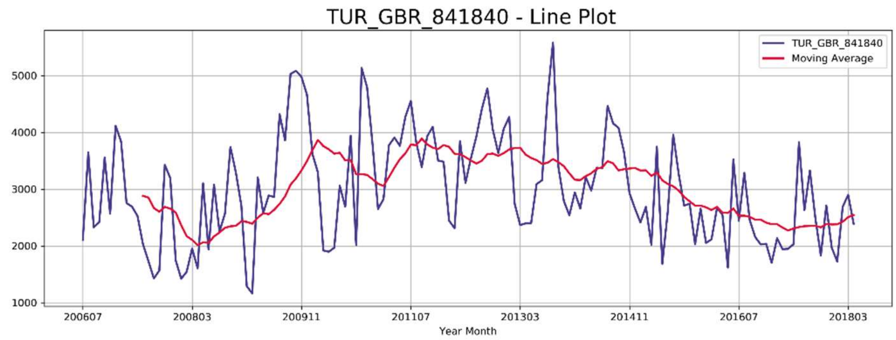
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure 32. Export Volumes of the Product 84.18.40 from Turkey to the UK from July 2006 to March 2018



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

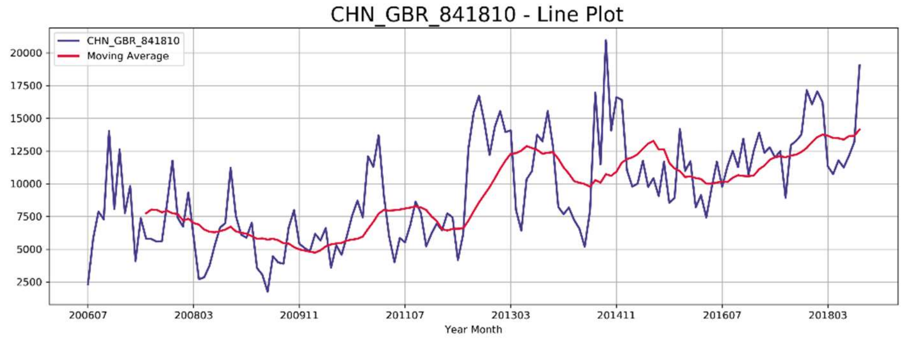
Figure 33. Export Volumes of the Product 84.18.50 from Turkey to the UK from July 2006 to March 2018



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

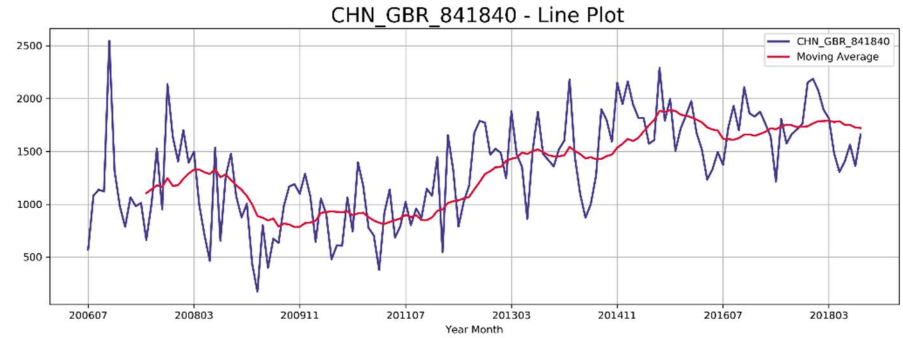
The time series plots of the exports from China to the UK for products 84.18.10, 84.18.40 and 84.18.50 are given in Figure 34, Figure 35 and Figure 36, respectively.

Figure 34. Export Volumes of the Product 84.18.10 from China to the UK from July 2006 to March 2018



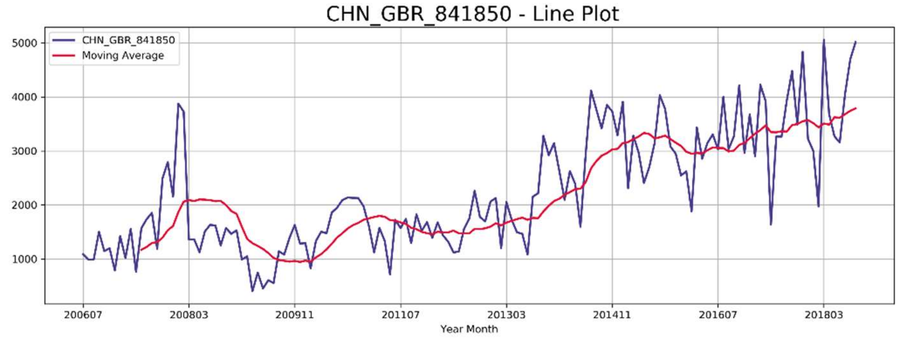
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure 35. Export Volumes of the Product 84.18.40 from China to the UK from July 2006 to March 2018



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure 36. Export Volumes of the Product 84.18.50 from China to the UK from July 2006 to March 2018



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

The top five import and export partners are determined based on Turkey's total exports and United Kingdom's total imports in 2016. The year 2016 is taken as the reference year in this dissertation as explained in Chapter 2. The top import and export partner countries are given in Table 13. Line plots for each trade couple are given in Appendix 1 - Line Plots Product Specific Features. Pair plots of each trade couple are presented in Appendix 3 - Pair Plots of Target variables versus Trade Couple.

Table 13. Turkey's Top Five Import Partners and the UK's Top Five Export Partners for Products in Case (in 2016).

Harmonized Code	Turkey's Top Five Import Partners (From largest trade value to the lowest)	The UK's Top Five Export Partners (From largest trade value to the lowest)
84.18.10	Germany, The United Kingdom, Italy, France, The United States	China, Turkey, Poland, Republic of Korea, Italy
84.18.40	The United Kingdom, The United States, France, Germany, Sweden	Turkey, China, Germany, Hungary Netherlands
84.18.50	The United Kingdom, Germany, Iraq, Saudi Arabia, Netherlands	Italy, China, Turkey, Austria, Czechia

Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Similar to Turkey, the top five importers and exporters are determined based on China's total exports and United Kingdom's total imports in 2016. The list of importer and exporter countries is presented in Table 14. Line plots for each trade couple are given in Appendix 1 - Line Plots Product Specific Features. Pair plots of each trade couple are presented in Appendix 3 - Pair Plots of Target variables versus Trade Couple.

Table 14. China's Top Five Import Partners and the UK's Top Five Export Partners for Products in Case (in 2016).

Harmonized Code	China's Top Five Import Partners (From the largest trade value to the lowest)	The UK's Top Five Export Partners (From the largest trade value to the lowest)
84.18.10	The United States, Japan, The United Kingdom, France, Australia	China, Turkey, Poland, Republic of Korea, Italy
84.18.40	The United States, The United Kingdom, France, Japan, Germany	Turkey, China, Germany, Hungary Netherlands
84.18.50	The United States, Australia, Italy, The United Kingdom, Indonesia	Italy, China, Turkey, Austria, Czechia

Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

In Chapter 4.1, the application steps (Figure 30) of machine learning algorithms are explained in detail. In the first step, the preliminary analysis approach is explained as given in Figure 30 as "A. Preliminary Analysis". According to the preliminary analysis, R^2 results of all applied combinations are given for each product and company case. Next, the applied steps in feature ranking are given in detail. It is labeled as "B. Feature Ranking" in Figure 30. In Chapters 4.1.3 and 4.1.4, the tuning process is applied for Turkey with a four-month forecast horizon and China with three-month and six-month horizon forecasts. This process is explained in detail in Figure 30 as the "C. Tuning Process".

4.1. Machine Learning Modeling Process

In this chapter, the followed process in this dissertation for Machine Learning Modeling is explained. The first section explains the preliminary analysis which is done before the tuning process. In the next section, the feature selection approach is explained. And, the following sections give detailed information about the tuning process to construct BCG Matrix with three-month and six-month horizon forecasts.

4.1.1. Preliminary Analysis

The process followed in the machine learning modeling step (FIT4) of the proposed methodology is presented in Figure 30 For all products, 28 features are employed in both random forest and artificial neural network models. The specific features to be used in the model depend on the preliminary study which is demonstrated in this chapter.

This preliminary study analyzes the effects of the input month combinations, feature selection and dependent variable transformation decisions on the forecast quality. Without the preliminary analysis model, the model would forecast trade volumes with the one-month horizon using the previous month’s data and all other features without transforming the dependent variable.

As the first step, the data of each feature is scaled according to min-max normalization since each feature in the dataset has different ranges. After the scaling step, the dataset is split into training and test sets. The training set is 80 percent of the entire dataset with 115 observations and the test set is the remaining 20 percent with 28 observations. Data stratification is applied to ensure the presence of each month to be included in training and test sets. Therefore, the accumulation of certain months in test or training sets is prevented. Training is performed by using training data. If no tuning process is applied to the models, both machine learning models take default values for hyperparameters. The hyperparameters of Random Forest and Artificial Neural Networks are specifically designed for the corresponding algorithm and significantly affect the quality of the results. Trained models are tested with the test data to see the performance of the model. Since all data are scaled (including dependent variable) inverse scaling should be applied to see the real behavior of the output. Both models were implemented by using “scikit-learn” open source libraries for Python on a Windows PC. Python libraries of “MLPRegressor” and “RandomForestRegressor” were used for Neural Network and Random Forest implementations, respectively. All source codes and data files are stored in a GitHub repository (https://github.com/ozemre/Thesis_Project).

During the data preparation, a preliminary search process is conducted with three aspects to achieve successful results in forecasting. After checking these three aspects, the tuning process is started. The first aspect is to decide on which month or month combinations of the past trade data should be taken as input factors in the forecasting model. The month combinations are given in the first column of each table (Table 15 through Table 20). Note that, Table 21 only reports the summary of the important results since extensive analyses were conducted with 18 different combinations. The second aspect is the feature selection threshold, which defines the percentage of the features to be included in the model according to their feature selection scores. The searched options are 50 percent, 75 percent, and 100 percent. To demonstrate, 50 percent means that only half of the features are included in the model according to their feature ranks and the remaining ones are omitted. To calculate each feature's rank, two values are combined. The third aspect is to determine the usage of the dependent variable. In this analysis, three options are searched for the Transformation of the Input Variable: (1) same (no transformation), (2) logarithm transformation and (3) square root transformation. With all these aspects and their combinations, 162 different datasets are trained. The results of this preliminary search with 162 combinations (with 10 random seed each) are presented in Table 15, Table 16 and Table 17 for products 84.18.10, 84.18.40 and 84.18.50 from Turkey to the UK. Similarly, in Table 18, Table 19 and Table 20, preliminary analysis results for products 84.18.10, 84.18.40 and 84.18.50 from China to the UK are given. As presented in Table 15, the dependent variable without any transformation, with 50 percent of the most important features and with the past trade data of 6 and 7 months yield the best result (R^2 of 0.859) for product 84.18.10 for exports of Turkey.

Table 15. R^2 Results for the Export Volume of Product 84.18.10 from Turkey to the UK after Doing Preliminary Analysis

Dependent Variables Transformation Type		Same (no transformation)			log transformation			sqrt transformation		
Selected Feature Percentage		**50**	**75**	**100**	**50**	**75**	**100**	**50**	**75**	**100**
Past Trade Data (Months)	Only (-1)	0.789	0.780	0.775	0.765	0.757	0.752	0.735	0.738	0.741
	(-1) and (-2)	0.780	0.767	0.775	0.767	0.754	0.751	0.745	0.742	0.740
	(-1), (-2), (-3)	0.773	0.770	0.767	0.776	0.762	0.761	0.767	0.760	0.754
	Only (-2)	0.772	0.765	0.768	0.759	0.756	0.753	0.739	0.743	0.738
	(-2) and (-3)	0.759	0.764	0.768	0.771	0.773	0.770	0.776	0.763	0.761
	(-2), (-3), (-4)	0.773	0.774	0.773	0.780	0.774	0.777	0.772	0.767	0.768
	Only (-3)	0.757	0.777	0.765	0.793	0.786	0.783	0.787	0.777	0.772
	(-3) and (-4)	0.783	0.780	0.776	0.786	0.789	0.787	0.788	0.786	0.780
	(-3), (-4), (-5)	0.817	0.810	0.803	0.811	0.800	0.803	0.793	0.786	0.782
	Only (-4)	0.781	0.789	0.789	0.772	0.788	0.786	0.767	0.763	0.771
	(-4) and (-5)	0.826	0.813	0.813	0.796	0.804	0.797	0.775	0.770	0.775
	(-4), (-5), (-6)	0.830	0.821	0.816	0.830	0.817	0.818	0.805	0.802	0.801
	Only (-5)	0.827	0.816	0.815	0.782	0.802	0.793	0.766	0.757	0.765
	(-5) and -6)	0.828	0.819	0.817	0.831	0.824	0.822	0.811	0.800	0.801
	(-5), (-6), (-7)	0.858	0.852	0.845	0.847	0.839	0.839	0.811	0.812	0.818
	Only (-6)	0.814	0.814	0.806	0.829	0.826	0.819	0.788	0.805	0.803
	(-6) and (-7)	0.859	0.843	0.835	0.841	0.833	0.835	0.806	0.810	0.807
	(-6), (-7), (-8)	0.823	0.816	0.810	0.831	0.814	0.807	0.792	0.785	0.783

According to these results in Table 16, the dependent variable with log transformation, with 50 percent of the most important features and with the past trade data of four months yield the best result (R^2 of 0.427) for product 84.18.40 for exports of Turkey.

Table 16. R^2 Results for the Export Volume of Product 84.18.40 from Turkey to the UK after Doing Preliminary Analysis

Dependent Variables Transformation Type		**Same (no transformation)**			**log transformation**			**sqrt transformation**		
Selected Feature Percentage		**50**	**75**	**100**	**50**	**75**	**100**	**50**	**75**	**100**
Past Trade Data (Months)	Only (-1)	0.141	0.143	0.133	0.143	0.144	0.146	0.157	0.105	0.109
	(-1) and (-2)	0.177	0.157	0.161	0.176	0.163	0.16	0.152	0.127	0.12
	(-1), (-2), (-3)	0.163	0.139	0.144	0.146	0.143	0.137	0.124	0.131	0.122
	Only (-2)	0.169	0.181	0.177	0.148	0.165	0.161	0.109	0.096	0.082
	(-2) and (-3)	0.224	0.193	0.199	0.203	0.183	0.176	0.115	0.111	0.096
	(-2), (-3), (-4)	0.316	0.292	0.294	0.28	0.245	0.252	0.189	0.164	0.158
	Only (-3)	0.191	0.194	0.191	0.212	0.184	0.195	0.135	0.127	0.122
	(-3) and (-4)	0.323	0.312	0.314	0.305	0.263	0.253	0.263	0.225	0.208
	(-3), (-4), (-5)	0.323	0.301	0.284	0.273	0.217	0.223	0.214	0.191	0.184
	Only (-4)	0.419	0.393	0.37	0.427	0.37	0.314	0.358	0.305	0.295
	(-4) and (-5)	0.386	0.311	0.311	0.332	0.251	0.241	0.311	0.249	0.243
	(-4), (-5), (-6)	0.371	0.36	0.372	0.313	0.298	0.274	0.277	0.253	0.237
	Only (-5)	0.176	0.182	0.204	0.176	0.143	0.145	0.156	0.139	0.145
	(-5) and -6)	0.245	0.27	0.267	0.207	0.193	0.206	0.186	0.181	0.177
	(-5), (-6), (-7)	0.264	0.258	0.247	0.224	0.2	0.188	0.183	0.166	0.155
	Only (-6)	0.247	0.249	0.277	0.2	0.198	0.239	0.166	0.183	0.171
	(-6) and (-7)	0.244	0.233	0.243	0.193	0.196	0.192	0.148	0.139	0.145
	(-6), (-7), (-8)	0.294	0.284	0.286	0.259	0.217	0.236	0.15	0.156	0.153

Table 17 shows that the dependent variable without any transformation, with 50 percent of the most important features and with the past trade data of four and five months yield the best result (R^2 of 0.833) for product 84.18.50 for exports of Turkey.

Table 17. R^2 Results for the Export Volume of Product 84.18.50 from Turkey to the UK after Doing Preliminary Analysis

Dependent Variables Transformation Type		**Same (no transformation)**			**log transformation**			**sqrt transformation**		
Selected Feature Percentage		**50**	**75**	**100**	**50**	**75**	**100**	**50**	**75**	**100**
Past Trade Data (Months)	Only (-1)	0.644	0.635	0.63	0.63	0.637	0.638	0.646	0.641	0.626
	(-1) and (-2)	0.668	0.659	0.664	0.671	0.662	0.656	0.668	0.663	0.67
	(-1), (-2), (-3)	0.703	0.685	0.686	0.721	0.71	0.713	0.693	0.693	0.695
	Only (-2)	0.598	0.593	0.585	0.633	0.615	0.608	0.586	0.598	0.603
	(-2) and (-3)	0.676	0.686	0.677	0.706	0.709	0.698	0.702	0.705	0.699
	(-2), (-3), (-4)	0.731	0.715	0.717	0.746	0.745	0.742	0.752	0.732	0.738
	Only (-3)	0.734	0.726	0.717	0.682	0.672	0.669	0.688	0.696	0.699
	(-3) and (-4)	0.806	0.769	0.757	0.734	0.734	0.74	0.78	0.762	0.746
	(-3), (-4), (-5)	0.786	0.761	0.761	0.756	0.755	0.742	0.787	0.768	0.758
	Only (-4)	0.813	0.753	0.753	0.73	0.707	0.723	0.806	0.761	0.751
	(-4) and (-5)	0.833	0.794	0.791	0.765	0.748	0.75	0.812	0.79	0.785
	(-4), (-5), (-6)	0.781	0.75	0.749	0.719	0.713	0.725	0.759	0.749	0.763
	Only (-5)	0.674	0.664	0.62	0.699	0.672	0.65	0.713	0.689	0.655
	(-5) and -6)	0.649	0.636	0.644	0.62	0.619	0.599	0.596	0.605	0.626
	(-5), (-6), (-7)	0.654	0.619	0.641	0.626	0.615	0.644	0.646	0.639	0.653
	Only (-6)	0.413	0.433	0.464	0.398	0.442	0.414	0.417	0.454	0.471
	(-6) and (-7)	0.473	0.534	0.558	0.532	0.524	0.526	0.517	0.564	0.554
	(-6), (-7), (-8)	0.517	0.516	0.536	0.618	0.55	0.562	0.54	0.55	0.559

The same analysis conducted for Turkey are repeated for China as well. According to the results given in Table 18, the dependent variable without any transformation, with 50 percent of the most important features and with the past trade data of three and four months yield the best result (R^2 of 0.679) for product 84.18.10 for exports of China.

Table 18. R^2 Results for the Export Volume of Product 84.18.10 from China to the UK after Doing Preliminary Analysis

Dependent Variables Transformation Type		**Same (no transformation)**			**log transformation**			**sqrt transformation**		
Selected Feature Percentage		**50**	**75**	**100**	**50**	**75**	**100**	**50**	**75**	**100**
Past Trade Data (Months)	Only (-1)	0.563	0.568	0.567	0.56	0.573	0.579	0.568	0.568	0.569
	(-1) and (-2)	0.604	0.583	0.587	0.6	0.596	0.589	0.603	0.591	0.596
	(-1), (-2), (-3)	0.609	0.608	0.603	0.628	0.596	0.604	0.627	0.605	0.602
	Only (-2)	0.643	0.614	0.608	0.614	0.617	0.591	0.623	0.603	0.597
	(-2) and (-3)	0.627	0.627	0.596	0.618	0.605	0.594	0.647	0.617	0.601
	(-2), (-3), (-4)	0.663	0.63	0.624	0.64	0.633	0.614	0.645	0.629	0.623
	Only (-3)	0.657	0.648	0.628	0.626	0.6	0.601	0.64	0.619	0.608
	(-3) and (-4)	0.679	0.634	0.614	0.62	0.627	0.616	0.641	0.628	0.617
	(-3), (-4), (-5)	0.661	0.632	0.611	0.627	0.598	0.6	0.67	0.622	0.601
	Only (-4)	0.625	0.594	0.596	0.542	0.602	0.6	0.602	0.589	0.586
	(-4) and (-5)	0.62	0.618	0.605	0.566	0.598	0.595	0.632	0.62	0.594
	(-4), (-5), (-6)	0.641	0.616	0.599	0.625	0.614	0.589	0.656	0.613	0.599
	Only (-5)	0.632	0.577	0.593	0.537	0.538	0.555	0.587	0.577	0.574
	(-5) and -6)	0.671	0.625	0.575	0.58	0.581	0.524	0.627	0.621	0.561
	(-5), (-6), (-7)	0.544	0.521	0.491	0.514	0.495	0.468	0.532	0.505	0.493
	Only (-6)	0.651	0.641	0.6	0.576	0.576	0.544	0.626	0.614	0.569
	(-6) and (-7)	0.554	0.516	0.507	0.558	0.513	0.487	0.549	0.525	0.497
	(-6), (-7), (-8)	0.616	0.571	0.55	0.591	0.549	0.541	0.601	0.568	0.543

Results in Table 19 show that the dependent variable with log transformation, with 50 percent of the most important features and with the past trade data of 4 and 5 months yield the best result (R^2 of 0.584) for product 84.18.40 for exports of China.

Table 19. R^2 Results for the Export Volume of Product 84.18.40 from China to the UK after Doing Preliminary Analysis

Dependent Variables Transformation Type		**Same (no transformation)**			**log transformation**			**sqrt transformation**		
Selected Feature Percentage		**50**	**75**	**100**	**50**	**75**	**100**	**50**	**75**	**100**
Past Trade Data (Months)	Only (-1)	0.538	0.52	0.513	0.551	0.549	0.536	0.539	0.529	0.517
	(-1) and (-2)	0.514	0.492	0.493	0.527	0.527	0.523	0.523	0.521	0.506
	(-1), (-2), (-3)	0.515	0.484	0.493	0.547	0.536	0.532	0.528	0.515	0.517
	Only (-2)	0.415	0.358	0.339	0.418	0.379	0.375	0.408	0.373	0.361
	(-2) and (-3)	0.432	0.399	0.379	0.457	0.449	0.433	0.442	0.425	0.41
	(-2), (-3), (-4)	0.456	0.429	0.429	0.503	0.479	0.477	0.467	0.459	0.45
	Only (-3)	0.447	0.429	0.423	0.46	0.479	0.469	0.448	0.457	0.442
	(-3) and (-4)	0.478	0.472	0.471	0.533	0.52	0.511	0.519	0.507	0.482
	(-3), (-4), (-5)	0.505	0.481	0.462	0.551	0.527	0.5	0.533	0.499	0.484
	Only (-4)	0.517	0.501	0.477	0.556	0.515	0.501	0.557	0.506	0.495
	(-4) and (-5)	0.544	0.486	0.485	0.584	0.529	0.5	0.553	0.512	0.501
	(-4), (-5), (-6)	0.465	0.424	0.415	0.482	0.441	0.422	0.47	0.427	0.417
	Only (-5)	0.527	0.477	0.465	0.528	0.504	0.488	0.544	0.492	0.476
	(-5) and -6)	0.421	0.39	0.376	0.422	0.397	0.39	0.44	0.387	0.38
	(-5), (-6), (-7)	0.366	0.352	0.329	0.363	0.353	0.343	0.355	0.35	0.327
	Only (-6)	0.357	0.339	0.312	0.392	0.343	0.338	0.348	0.333	0.328
	(-6) and (-7)	0.333	0.298	0.287	0.32	0.311	0.308	0.306	0.303	0.289
	(-6), (-7), (-8)	0.374	0.348	0.333	0.361	0.367	0.363	0.357	0.36	0.34

According to the results in Table 20, the dependent variable with log transformation, with all features and with the past trade data of 3 and 4 months yields the best result (R^2 of 0.925) for product 84.18.50 for exports of China.

Table 20. R^2 Results for the Export Volume of Product 84.18.50 from China to the UK after Doing Preliminary Analysis

Dependent Variables Transformation Type		**Same (no transformation)**			**log transformation**			**sqrt transformation**		
Selected Feature Percentage		**50**	**75**	**100**	**50**	**75**	**100**	**50**	**75**	**100**
Past Trade Data (Months)	Only (-1)	0.806	0.809	0.805	0.848	0.843	0.855	0.810	0.833	0.831
	(-1) and (-2)	0.822	0.821	0.826	0.846	0.851	0.862	0.837	0.837	0.836
	(-1), (-2), (-3)	0.835	0.835	0.837	0.870	0.866	0.878	0.858	0.851	0.852
	Only (-2)	0.852	0.83	0.835	0.844	0.832	0.837	0.846	0.84	0.834
	(-2) and (-3)	0.853	0.841	0.829	0.847	0.836	0.833	0.856	0.846	0.848
	(-2), (-3), (-4)	0.865	0.868	0.87	0.865	0.867	0.874	0.87	0.871	0.865
	Only (-3)	0.887	0.915	0.883	0.895	0.905	0.905	0.915	0.918	0.891
	(-3) and (-4)	0.904	0.902	0.903	0.913	0.921	0.925	0.916	0.913	0.917
	(-3), (-4), (-5)	0.896	0.891	0.885	0.915	0.922	0.92	0.912	0.905	0.906
	Only (-4)	0.855	0.851	0.855	0.868	0.857	0.871	0.863	0.875	0.86
	(-4) and (-5)	0.867	0.864	0.852	0.865	0.88	0.875	0.864	0.873	0.859
	(-4), (-5), (-6)	0.881	0.881	0.873	0.879	0.887	0.888	0.873	0.876	0.88
	Only (-5)	0.869	0.86	0.853	0.835	0.862	0.863	0.855	0.849	0.858
	(-5) and -6)	0.878	0.868	0.865	0.87	0.872	0.879	0.873	0.867	0.874
	(-5), (-6), (-7)	0.901	0.891	0.887	0.911	0.916	0.912	0.899	0.899	0.9
	Only (-6)	0.86	0.878	0.884	0.86	0.881	0.89	0.86	0.887	0.894
	(-6) and (-7)	0.9	0.899	0.896	0.921	0.921	0.919	0.905	0.915	0.916
	(-6), (-7), (-8)	0.87	0.869	0.869	0.887	0.897	0.885	0.881	0.878	0.873

According to the results of the preliminary analysis (a total of 162 test combinations and 10 random number seed for each combination), the best combinations for tuning are used to answer two different questions. Firstly, which input combinations give the best result independent of the forecast horizon? Secondly, which forecast horizon gives the best result from the country perspective? The answer to the first question is given in Table 21.

Table 21. The Best Input Combination of each Product According to the Preliminary Analyses for Turkey and China

Product Harmonized Code	Included Past Trade Data (Months) (all month combinations)	Transformation of the Dependent Variable (no transformation, log, sqrt)	Percentage of Selected Features (50%, 75%, 100%)	Achieved R^2
Turkey to the United Kingdom				
84.18.10	-6 and 7[a]	same	50	0.859
84.18.40	-4	log	50	0.427
84.18.50	-4 and -5	same	50	0.833
China to the United Kingdom				
84.18.10	-3 and -4	same	50	0.679
84.18.40	-4 and -5	log	50	0.584
84.18.50	-3 and -4	log	100	0.925

[a] *"-7" refers to the trade volume seven months ago*

For the second question, finding the right forecast horizon for all products across the country, R^2 loss in values is taken into account. R^2 loss is the percentage reduction information with respect to achieving the highest R^2 value. For example, for Product 84.18.10 from Turkey to the UK highest achieved R^2 value in preliminary analysis independent of the forecast horizon is 0.859. For the three-month horizon forecast, the achieved R^2 is 0.817 and with respect to the highest R^2 (0.859), the loss is 5 percent. According to the results of Table 22, the best forecasts for China are obtained with the forecast horizon of three-month. Therefore, the forecast horizon for China is selected as three-month.

Table 22. The Best Input Combination of Each Product According to the Preliminary Analyses for Turkey and China for the Three-Month Forecast Horizon

Product Harmonized Code	Included Past Trade Data (Months) (-3, -3 and -4, -3, -4 and -5)	Transformation of the Dependent Variable (no transformation, log, sqrt)	Percentage of Selected Features (50%, 75%, 100%)	Achieved R^2	R^2 loss %
Turkey to the United Kingdom					
84.18.10	-3, -4 and -5	same	50	0.817	5
84.18.40	-3 and -4	same	50	0.323	24
84.18.50	-3 and -4	same	50	0.806	3
China to the United Kingdom					
84.18.10	-3 and -4	same	50	0.679	0
84.18.40	-3, -4 and -5	log	50	0.551	6
84.18.50	-3 and -4	log	100	0.925	0

According to the results of Table 23, the best forecasts for Turkey are obtained with the forecast horizon of four-month. Therefore, the four-month forecast horizon is selected for Turkey.

Table 23. The Best Input Combination of Each Product According to the Preliminary Analyses for Turkey and China for the Four-Month Forecast Horizon

Product Harmonized Code	Included Past Trade Data (Months) (-4, -4 and -5, -4, -5 and -6)	Transformation of the Dependent Variable (no transformation, log, sqrt)	Percentage of Selected Features (50%, 75%, 100%)	Achieved R^2	R^2 loss %
Turkey to the United Kingdom					
84.18.10	-4, -5 and -6	same	50	0.830	3
84.18.40	-4	log	50	0.427	0
84.18.50	-4 and -5	same	50	0.833	0
China to the United Kingdom					
84.18.10	-4, -5 and -6	sqrt	50	0.656	3
84.18.40	-4 and -5	log	50	0.584	0
84.18.50	-4, -5 and -6	log	100	0.888	4

To generalize the forecasting approach, the same input combinations across products can be used. However, using a single best input combination across all products reduces R^2 values. Therefore, each product and country pair should be handled separately because of their independent dynamics. For the input combinations of Turkey, the results of the preliminary study given in Table 23 are used. For the forecast horizon, four-month is selected according to Table 23. For China, the input combinations are selected as given in Table 22 and three-month is preferred.

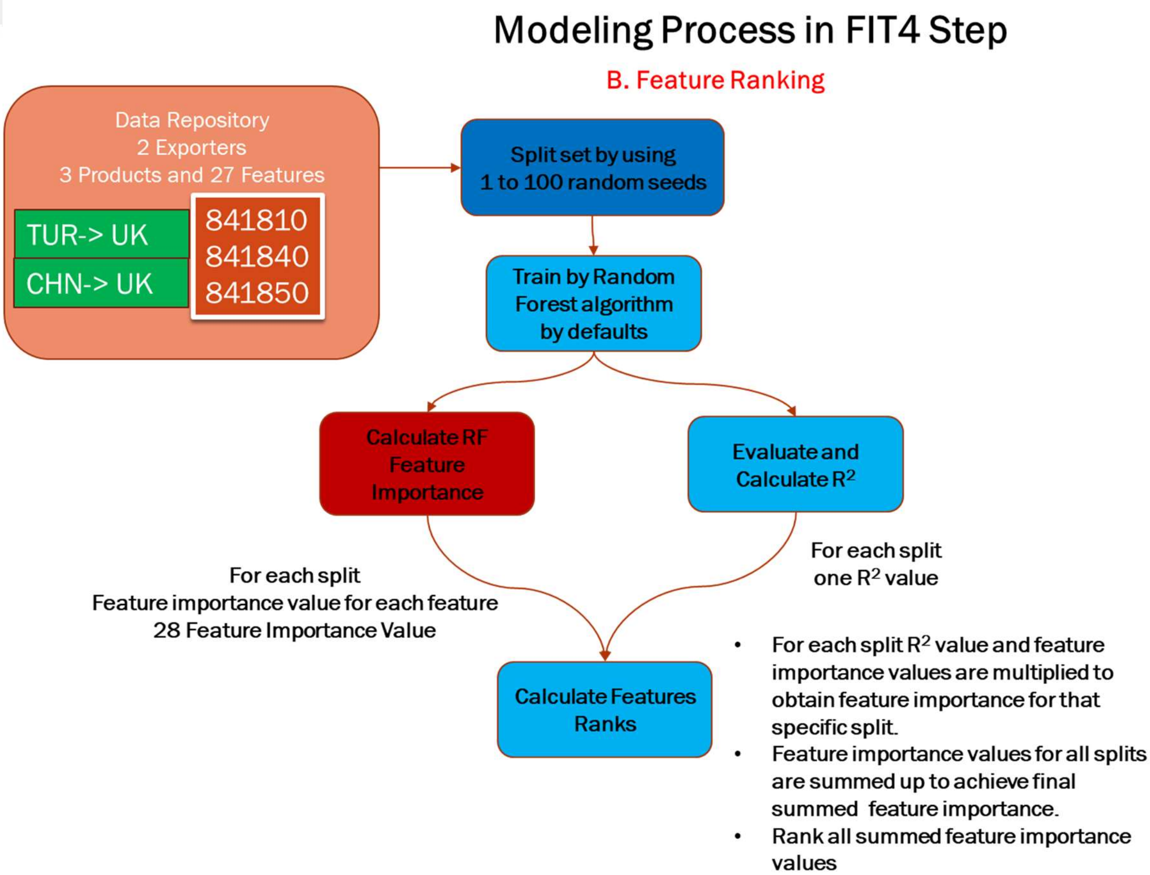
In Chapter 4.1.2, the Feature ranking approach of this dissertation is presented.

4.1.2. Feature Ranking

Feature selection is a very common process in machine learning and statistics. It is the process of synthesizing a subset of all input variables for the ML model by eliminating unnecessary or irrelevant features (Awad & Khanna, 2015). The main reason for using feature selection is to improve the accuracy of a model by reducing its complexity. It also helps to accelerate the training process. The feature selection process is mainly focused on selecting a subset of features that are more relevant to the model and have a higher impact on accuracy. The amount of reduction of unrelated features is another issue for feature selection. The dataset is mainly based on the factors given in Table 6. In standard settings, it has 28 features. The number of features entering the dataset varies depending on the number of lagged months used. Number of features doubles when two consecutive previous months are used. Similarly, it triples when three consecutive previous months are used. Table 11 presents the number of features depending on different input combinations.

To decide which features are more relevant and have higher impacts in the model, feature importance function of Random Forest is used. By using this function, the rank of the features is found for each dataset. Figure 37 gives the applied process for feature ranking. Each dataset split into 100 different test and training sets. Multiple test and training sets are used to increase the accuracy of feature selection and not to rely on the results of a small number of tests and training sets. By using Random Forest feature importance function, for each split importance values for each feature is calculated. For one specific split set, 28 feature importance values for all features are obtained. At the same time, the R^2 result of that specific split set is calculated. By combining each split's feature importance and R^2 result, feature ranks for that specific split are found. R^2 result is multiplied by each feature's importance value, which yields the feature importance of a specific split. After adding feature importance values of all splits, the final sum of feature importance for each feature is achieved. The summed final feature importance's for each feature is sorted to get the ranked list of all features. With the sorted list of features, the rank of each feature for that dataset is achieved. This process favors the features with high importance, where the features yielding higher R^2 values take higher points than the ones with lower R^2 values. After combining all values from 100 different sets, the rank of the features are found. The results of this process are applied both for RF and ANN models.

Figure 37. Calculation of Feature Rank Process



The selected month series of the input variables changes the number of features in each dataset. In Table 11 the number of features in each input combination is given. This makes it difficult to compare results of feature ranking. Three-month horizon forecast using only single lagged month (-3 month input as in dataset)is selected to compare feature rank results as given in Table 24 and 25.

Table 24 and Table 25 shows that the total export volume of an exporter for a product is the most common feature in all product and country pairs. The second group of common features is dependent variables with three-month shifted values, world total trade, Consumer Price Index of the exporter country, Consumer Price Index of the importer country and Composite Leading Indicator of the exporter country. These results indicate that the proposed feature selection method is correct since they reveal consumer behavior. The last common group from the top ten list are Business Confidence Indicator of the importer country, Consumer Confidence Indicator of the importer country, Composite Leading Indicator of the importer and import values of the top five importers from Turkey and China. On the other hand, Economic Politic Uncertainty both for the importer and exporter countries, and the Producer Price Index does not appear once in the top ten features list.

Table 24. Top-ranked Features of All Products Types for Turkey for the Three-Month Horizon Forecast

	Product Harmonized Code		
Feature Rank	**84.18.10**	**84.18.40**	**84.18.50**
1	TUR_FRA_841810-3	CCI_UK-3	TUR_World_841850-3
2	CPI_UK-3	TUR_FRA_841840-3	CCI_TUR-3
3	CPI_TUR-3	NLD_UK_841840-3	TUR_IRQ_841850-3
4	TUR_World_841810-3	TUR_UK_841840_UV-3	BCI_UK-3
5	TUR_ITA_841810-3	GDP_UK-3	CLI_TUR-3
6	CCI_TUR-3	CHN_UK_841840-3	GDP_TUR-3
7	World-3	CCI_TUR-3	TUR_UK_841850_UV-3
8	TUR_UK_841810-3	TUR_UK_841840-3	CHZ_UK_841850-3
9	TUR_DEU_841810-3	GDP_TUR-3	TUR_DEU_841850-3
10	TRY-3	DEU_UK_841840-3	CLI_UK-3

Table 25. Top-ranked Features of All Products Types for China for the Three-month Horizon forecast

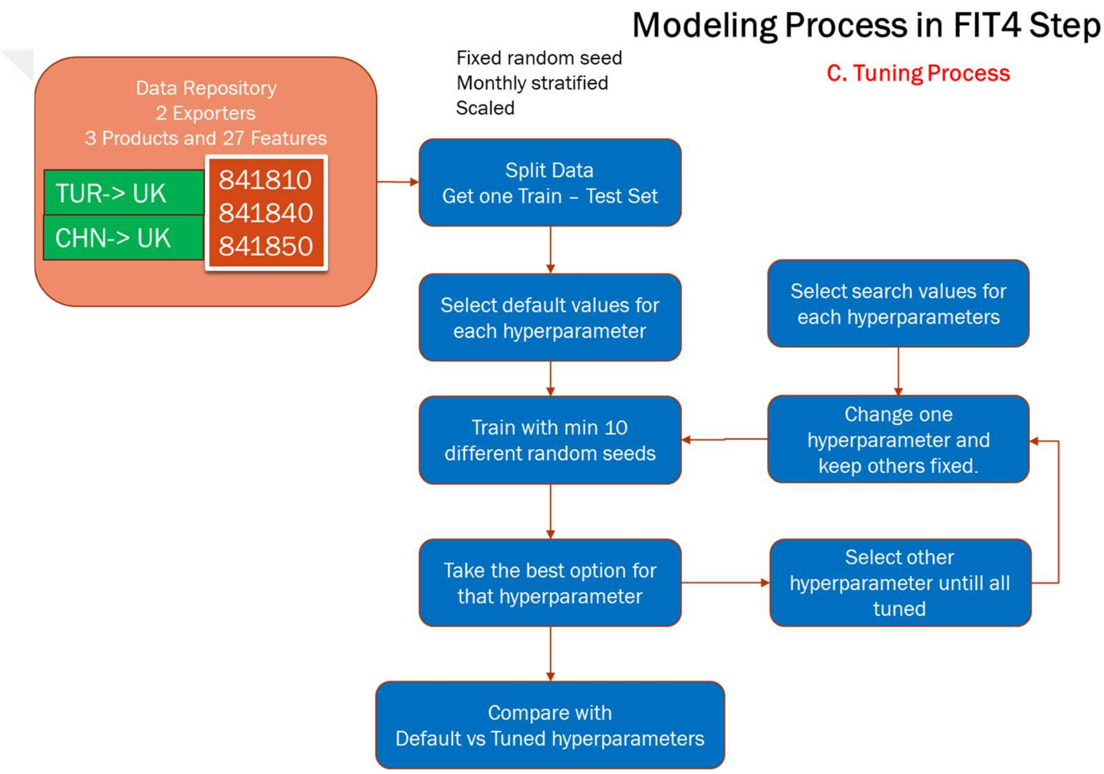
	Product Harmonized Code		
Feature Rank	**84.18.10**	**84.18.40**	**84.18.50**
1	CPI_CHN-3	CPI_CHN-3	CPI_UK-3
2	CPI_UK-3	CPI_UK-3	CPI_CHN-3
3	CHN_World_841810-3	World-3	CHN_USA_841850-3
4	CLI_UK-3	CHN_World_841840-3	World-3
5	World-3	CLI_CHN-3	CHN_UK_841850-3
6	BCI_UK-3	CHN_DEU_841840-3	CHN_World_841850-3
7	CLI_CHN-3	CHN_USA_841840-3	CLI_CHN-3
8	CHN_USA_841810-3	BCI_UK-3	CLI_UK-3
9	ITA_UK_841810-3	DEU_UK_841840-3	CCI_UK-3
10	BCI_CHN-3	BCI_CHN-3	AUT_UK_841850-3

In Chapter 4.1.4, the tuning process applied for Turkey and China dataset is explained. The tuning process is applied to both RF and ANN algorithms. Results are used to build a BCG matrix with a three-month horizon forecast for hypothetical Turkish and a Chinese company.

4.1.3. Hyperparameter Tuning to Construct BCG Matrix with Three-Month Horizon Forecast

After preliminary analysis by using feature ranking, the input conditions of each dataset are determined. Before starting hyperparameter tuning, datasets should be ready for training. Therefore, data scaling and data splitting by using month stratification are repeated. Each feature is normalized according to min-max normalization. After normalization, data split into training and test sets, where the training set is 80 percent of the entire dataset with 115 observations and test set is the remaining 20 percent with 28 observations. Data stratification is applied to ensure the presence of each month to be included in the training and test sets. The next step is to train the ML model according to the default values of hyperparameters. All hyperparameters, their default values and search space used in tuning for the RF algorithm are given in Table 1 and for the ANN algorithm are given in Table 2. Then, the first hyperparameter, according to its search space, is selected to train the ML model while keeping other hyperparameters fixed. After training with ten different random seeds, the hyperparameter value corresponding to the best performing R^2 score is fixed. Then, the same process is repeated for the next hyperparameter in the search space by using the best values of the previously fixed hyperparameters. This process continues until the search space of all hyperparameters is exhausted and their best performing values are found. This tuning process for both algorithms is summarized in Figure 38.

Figure 38. Hyperparameter Tuning Steps used in RF and ANN Models



Both methods are implemented by using "scikit_learn" open source libraries for Python 3.0 on a Windows running PC. Python libraries of "MLP-Regressor" and "Random-Forest-Regressor" are used for Artificial Neural Network and Random Forest implementations, respectively.

Upon completing the tuning of the Random Forest algorithm for each product, the hyperparameter values are obtained as in Table 26.

Table 26. Tuned Hyperparameter Values of the Random Forest Model for All Products and Countries

Product Harmonized Code	Maximum Features	Minimum Sample Leaf	Maximum Leaf Nodes	Minimum Weight Fraction For Leaf	Minimum Impurity Decrease	Number of Estimators
Turkey to the United Kingdom						
84.18.10	Auto	2	200	0.00001	0.000001	200
84.18.40	10	5	10	0.1	0.000001	1000
84.18.50	10	2	100	0.01	0.000001	20000
China to the United Kingdom						
84.18.10	Auto	1	100	0.00001	0.000001	1000
84.18.40	Auto	2	100	0.001	0.00001	200
84.18.50	Auto	1	200	0.01	0.00001	200

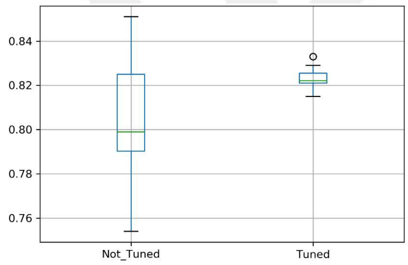
Similar to the tuning process of the Random Forest model, the hyperparameter values of the Artificial Neural Network model are found in Table 27. Note that applying the tuning process and using the appropriate hyperparameters are very important to achieve good quality results since the hyperparameter values are slightly different for each product in both models.

Table 27. Tuned Hyperparameters Values of the ANN Model for All Products and Countries

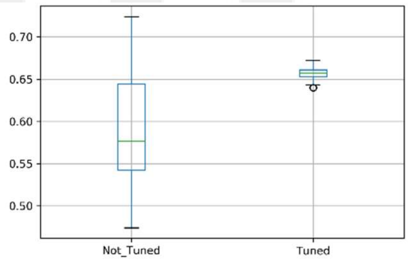
Product Harmonized Code	Solver	Activation	Alpha	Maximum Number Of Iterations	Hidden Layer Size
Turkey to the United Kingdom					
84.18.10	lbfgs	identity	0,001	100000	(100, 100)
84.18.40	lbfgs	identity	0,00001	100000	(100, 100)
84.18.50	lbfgs	identity	0,00001	50000	(30, 100, 30)
China to the United Kingdom					
84.18.10	lbfgs	logistic	1E-07	1000	(30, 100, 30)
84.18.40	lbfgs	relu	0,00001	100000	(30, 30, 30)
84.18.50	lbfgs	identity	0,00001	50000	(10, 10)

To demonstrate the efficiency of the tuning process, Figure 39 compares the R^2 values of the Random Forest model between tuned hyperparameter values and default (not tuned) hyperparameter values for Product 84.18.10. As seen from the boxplots of ten different training and test sets with different random number seeds, the tuning process yields significantly higher R^2 values and more robust results. Similarly, Figure 40 and Figure 41 compare the tuning results for Product 84.18.40 and Product 84.18.50, respectively. Note that Figure 39, Figure 40 and Figure 41 show only the results of the Random Forest model, however, the results of the Neural Network model are similar.

Figure 39. Tuning Results (Default and Tuned Hyperparameters) of the RF Model for 84.18.10 for Turkey and China Data

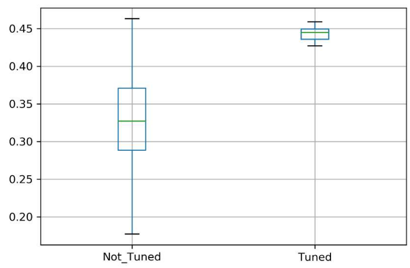


a) Turkey to the UK

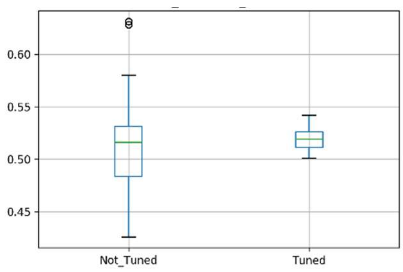


b) China to the UK

Figure 40. Tuning Results (Default and Tuned Hyperparameters) of the RF Model for 84.18.40 for Turkey and China Data

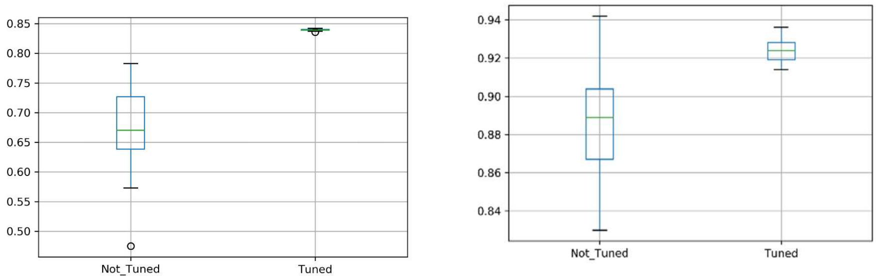


a) Turkey to the UK



b) China to the UK

Figure 41. Tuning Results (Default and Tuned Hyperparameters) of the RF Model for 84.18.50 for Turkey and China Data



a) Turkey to the UK b) China to the UK

Figure 42 compares the Random Forest and Artificial Neural Network models in terms of median R^2 values for Turkey data. Both the RF and ANN models are compared with 10 random number seeds. To demonstrate the robustness of the tuning process, the results of the same tuning training/test split (with 10 random seeds) and different random train/test splits (10 different splits) are compared for each model. As seen from the similar results of the same and different train/test datasets in Figure 42, the tuning yields robust results for both models and all product types. Similar observations are found for China data as shown in Figure 43.

According to Figure 42, the RF algorithm yields higher R^2 values than the ANN algorithm for all product types. Among three product types, the highest R^2 value (0.840 with RF) is observed in Product 84.18.50, and the lowest R^2 value (0.338 with ANN) is observed in Product 84.18.40. However, for Product 84.18.40 the median R^2 value is 0.449 when RF is used, which can be considered as a moderate effect size. The highest median R^2 value of the Neural Network model is 0.8781 for product 84.18.10, but the Random Forest model achieves 0.826 median R^2 value for the same product. Therefore, according to these results, the Random Forest model achieves higher R^2 values and performs better than the Artificial Neural Network model.

Figure 42. The Median R^2 Values of the Tuned RF and ANN Models with Using the Same Training Set and Different Training Sets for Turkey

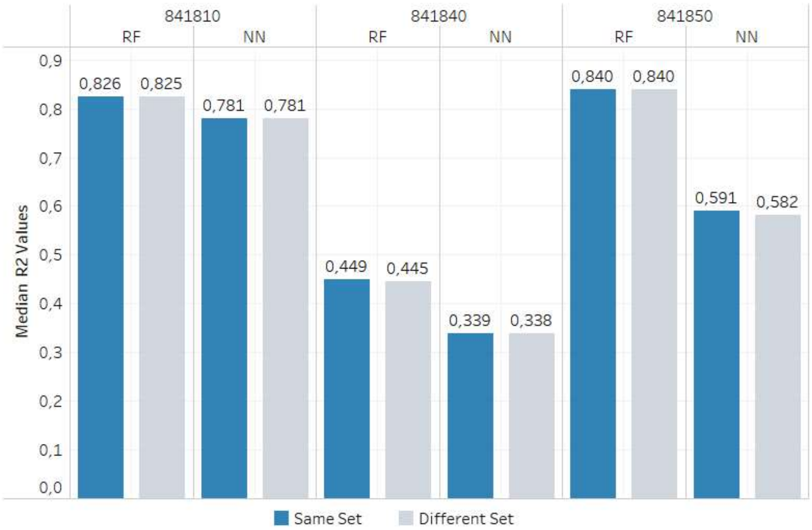
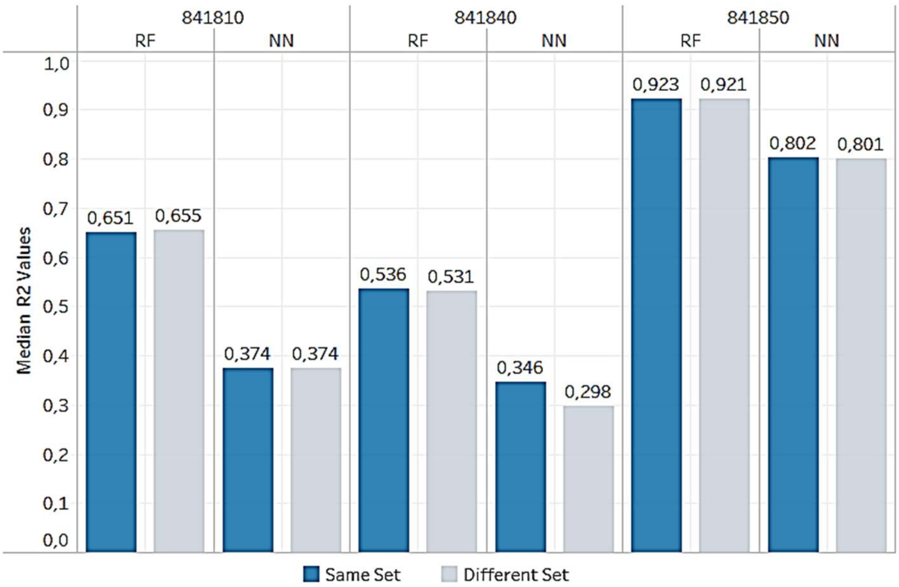


Figure 43 shows that the RF algorithm yields higher R^2 values than the ANN algorithm for all product types. Among three product types, the highest R^2 value (0.923 with RF) is observed in Product 84.18.50, and the lowest R^2 value (0.298 with ANN) is observed in Product 84.18.40. However, for Product 84.18.40 the median R^2 value is 0.531 when RF is used. It can be considered a medium effect on results. Similar to Product 84.18.40, the Artificial Neural Network Model yields a smaller R^2 value (0.374) for 84.18.10. And, the highest median R^2 value of the Neural Network model is 0.802 for product 84.18.50, but the Random Forest model achieves 0.921 median R^2 value. Therefore, according to these results, the Random Forest model achieves higher R^2 values and performs better than the Artificial Neural Network model.

Figure 43. The Median R^2 Values of the Tuned RF and ANN Models with Using the Same Training Set and Different Training Sets for China

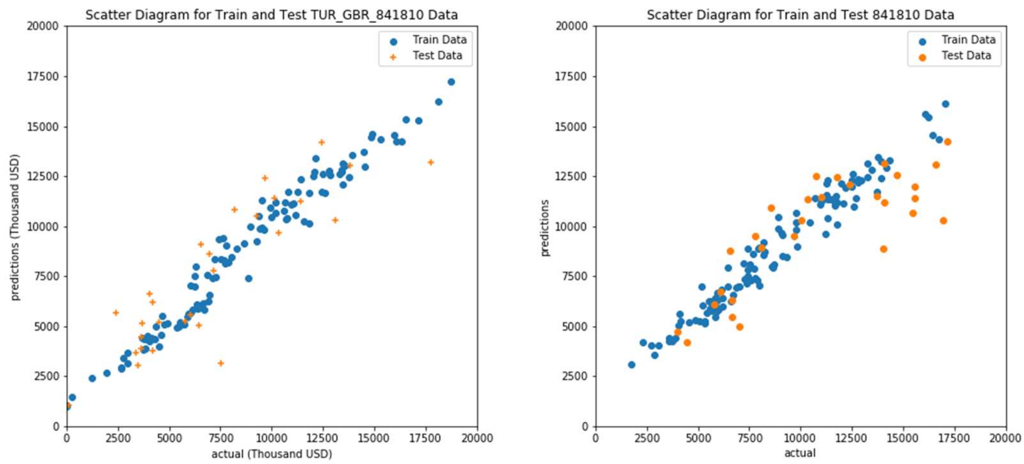


When Figure 42 and Figure 43 are examined together, the Random Forest model achieves higher scores than Artificial Neural Network Models. Product 84.18.40achieves the lowest results for both Turkey and China data. Another conclusion is that the achieved results are valid not only for the training and test set where tuning is applied but also for other random training and test splits.

The scatter plots presented in Figure 44, Figure 45 and Figure 46 show the actual and forecasted values of training and test data in the RF model. The blue circles denote the training data and the orange ones denote the test data. Note that, R^2 values are calculated by only using the test data. Scatter plots show the actual versus forecasted values of all datasets. The expected result is that a diagonal line having small deviations from that line.

Turkey's data distribution with 0.826 R^2 value for 84.18.10 product is given in Figure 44. Data for China is more can be easily recognized where deviations from the diagonal line are more than Turkey's data. The deviations of red dots from the diagonal line are supported with 0.651 R^2 value.

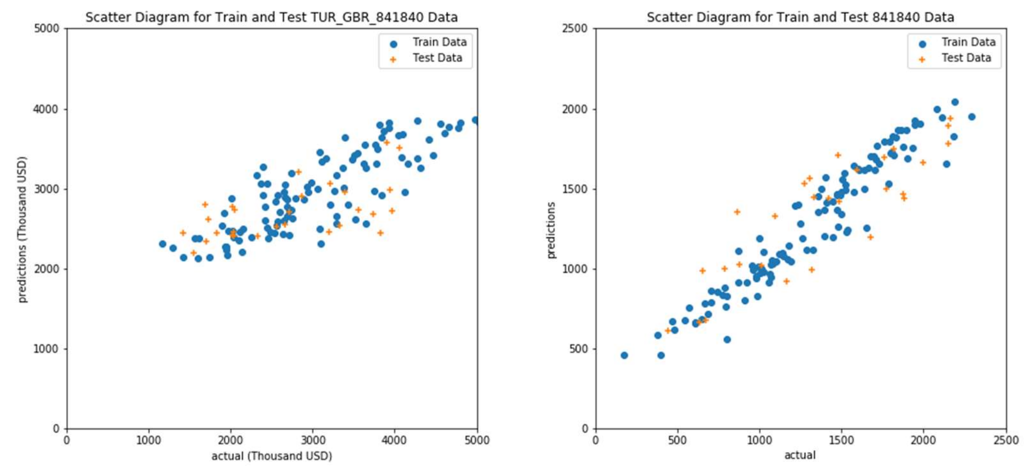
Figure 44. Actual vs. Forecasted Values (Training and Test Data) of the RF Model for Product 84.18.10.



a) Turkey to the UK b) China to the UK

Turkey's data distribution with 0.449 R^2 value for 84.18.40 product, where there is no diagonal line is apparent in Figure 45. Data for China results in a higher R^2 score with 0.536 and shows more tight distribution over the diagonal line than Turkey's data.

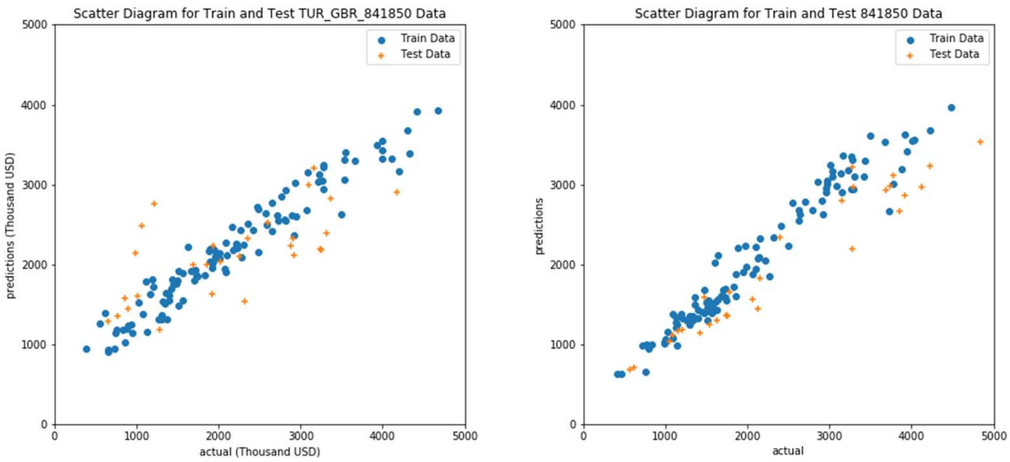
Figure 45. Actual vs. Forecasted Values (Training and Test Data) of the RF Model for Product 84.18.40.



a) Turkey to the UK b) China to the UK

Turkey and China's data show diagonal line distributions for 84.18.50 product as given in Figure 46. Turkey's R^2 value is 0.84 and China results in 0.923 R^2 where the distributions in the figures support these high R^2 values.

Figure 46. Actual vs. Forecasted Values (Training and Test Data) of the RF Model for Product 84.18.50.



a) Turkey to the UK b) China to the UK

Figure 47, Figure 48 and Figure 49 show the forecast results of all products from Turkey to the UK with the Random Forest model. In Figure 47 and Figure 49, high R^2 scores can be seen from figures where R^2 is 0.826 for 84.18.10 and 0.84 for 851850. Figure 48 shows the R^2 value of 0.449 for product 84.18.40.

Figure 47. Export Volume Forecast and Real Values of Product 84.18.10 from Turkey to the UK

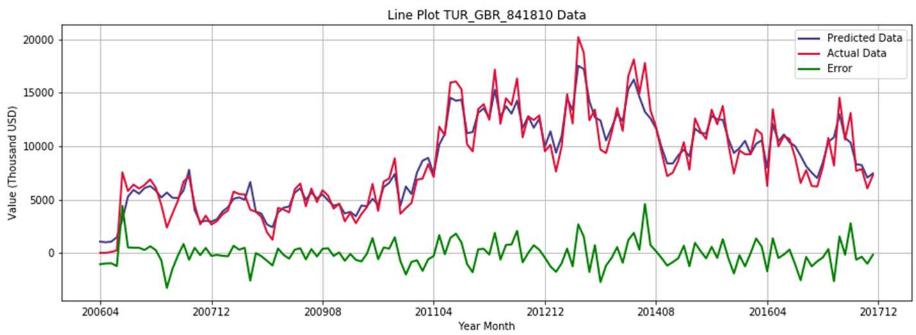


Figure 48. Export Volume Forecast and Real Values of Product 84.18.40 from Turkey to the UK

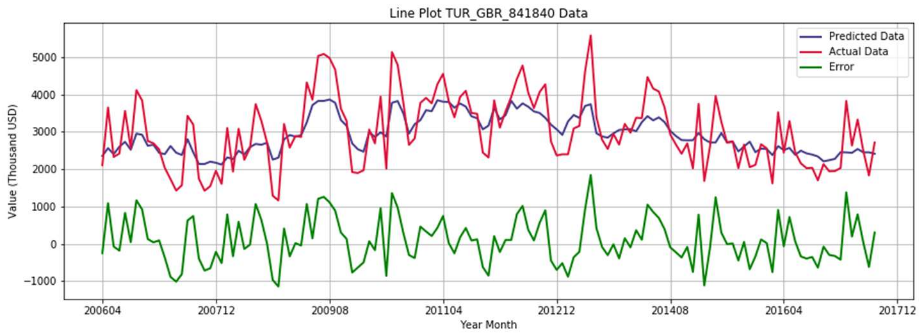


Figure 49. Export Volume Forecast and Real Values of Product 84.18.50 from Turkey to the UK

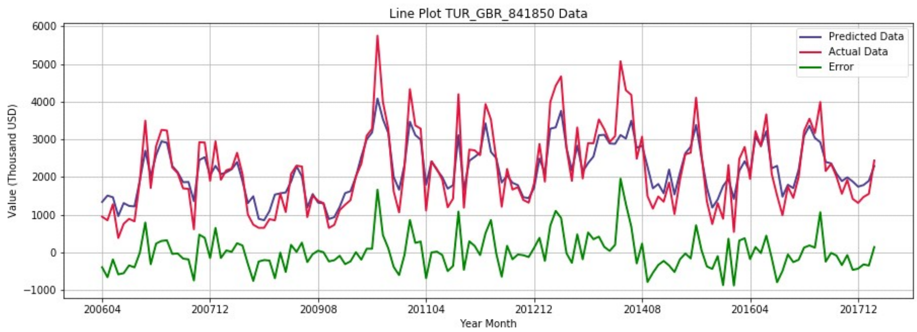


Figure 50, Figure 51 and Figure 52 show the forecast results of all products from China to the UK with the Random Forest model. In all figures, high R^2 values can be observed for all product types. The highest R^2 value is for 0.923 for Product 84.18.50 and the lowest R^2 value is 0.536 for Product 84.18.40.

Figure 50. Export Volume Forecast and Real Values of Product 84.18.10 from China to the UK

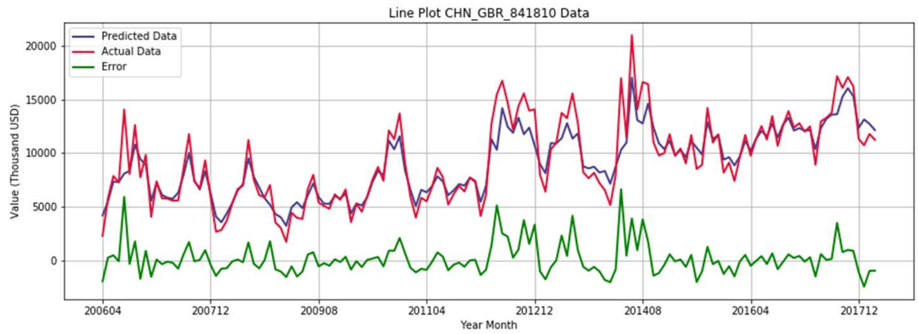


Figure 51. Export Volume Forecast and Real Values of Product 84.18.40 from China to the UK

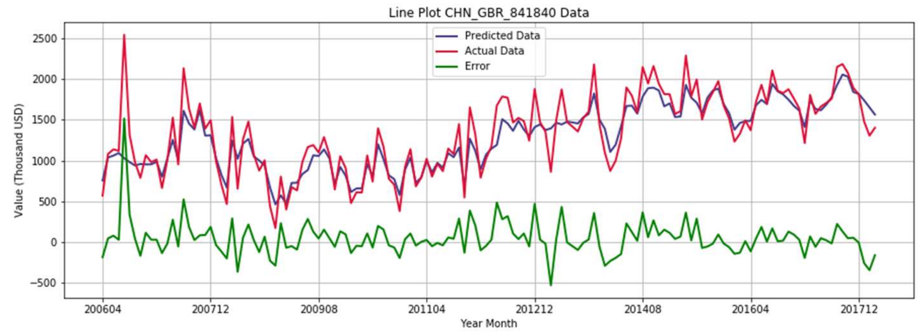


Figure 52. Export Volume Forecast and Real Values of Product 84.18.50 from China to the UK

The next chapter explains the hyperparameter tuning process for a six-month horizon forecast. The results of Chapter 4.1.4 are used in Chapter 4.2.4 to compare BCG Matrix with a three-month horizon forecast and six-month horizon forecast.

4.1.4. Hyperparameter Tuning to Construct BCG Matrix with Six-Month Horizon Forecast

The three-month horizon forecast is selected for BCG matrix implementation. There are two main reasons. First, R^2 values are mainly better for three and four-month forecast horizons. For China's case with a three-month horizon forecast and for Turkey's case with a four-month forecast horizon give the best results. The second one is that the fluctuations in export volumes are significantly large and one month ahead forecast horizon is too short and six months is too long to make decisions. To justify this approach for China products six-month horizon forecast is implemented with all tuning processes explained in Chapter 4.1.3. According to the preliminary study, the selected parameters and the resulting loss of R^2 values are given in Table 28.

Table 28. The Best Input Combination of Each Product According to the Preliminary Analyses for China for Six-Month Horizon Forecast

Product Harmonized Code	Included Past Trade Data (Months) (-6, -6 and -7, -6, -7 and -8)	Transformation of the Dependent Variable (no transformation, log, sqrt)	Percentage of Selected Features (50%, 75%, 100%)	Achieved R^2	R^2 loss %
China to the United Kingdom					
84.18.10	-6	same	50	0.651	4
84.18.40	-6	log	50	0.392	33
84.18.50	-6 and -7	log	50	0.921	0

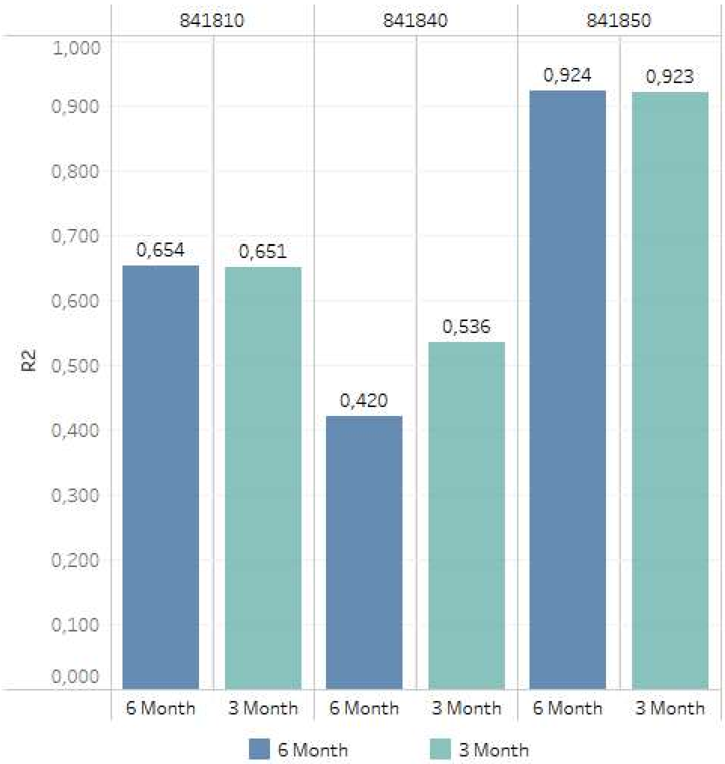
According to the developed tuning process for Random Forest which is given in Figure 38, achieved hyperparameters are given in Table 29. In this step, only Random Forest is tuned, since the Random Forest model gives better results than the Artificial Neural Network model.

Table 29. Hyperparameter Values after Tuning of the RF Model for All Products Exported from China to the UK

Product Harmonized Code	Maximum Features	Minimum Sample Leaf	Maximum Leaf Nodes	Minimum Weight Fraction For Leaf	Minimum Impurity Decrease	Number of Estimators
84.18.10	Auto	1	200	0,001	0,00001	200
84.18.40	10	2	200	0,00001	0,000001	500
84.18.50	Sqrt	2	200	0,01	0,00001	10000

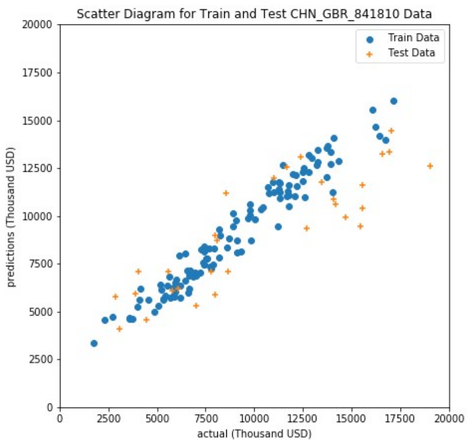
For a six-month horizon forecast for China's products, Figure 53 shows that R^2 values of the tuned algorithm are almost the same as the ones of the three-month horizon forecast. Only for Product 84.18.40, there is a slight decrease in R^2. However, the R^2 values of other products remain comparable.

Figure 53. The Median R^2 Values of the Tuned RF Model for Six-Month Horizon Forecast Compared with the Three-month horizon forecast for China Products.

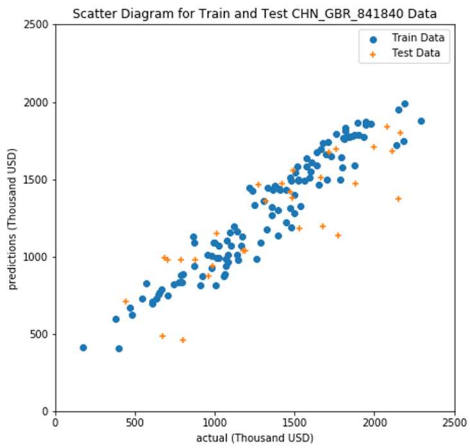


The scatter plots for each product presented in Figure 54 shows the actual and forecasted values of training and test data in the RF model. The blue circles denote the training data and the orange ones denote the test data. Note that, R^2 values are calculated by only using test data. R^2 score of 0.924 for product 84.18.50(Figure 54b) can be easily recognized which is centered well along as a diagonal line. Also, R^2 scores for other products (0.654 for 84.18.10 and 0.420 for 84.18.50) can be seen by the distribution of test data in Figure 54a and Figure 54c, respectively.

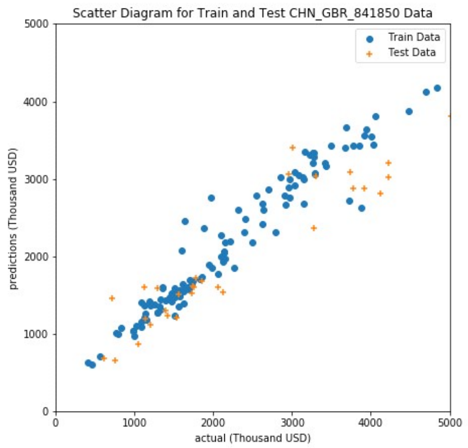
Figure 54. Actual vs. Forecasted Values (Training and Test Data) of the RF Model for All Products for The Six-Month Horizon Forecast.



a) 84.18.10



b) 84.18.40



c) 84.18.50

Chapter 4.2 explains the implementation of the BCG Matrix for the hypothetical Turkish and Chinese companies are given.

4.2. Implementing the BCG Matrix

This chapter explains the implementation of the BCG matrix by using tuning results in ML models. Chapter 4.2.1, Chapter 4.2.2 and Chapter 4.2.3 use the tuning parameters found in Chapter 4.1.3 to construct a three-month horizon forecast BCG Matrix. The last section (Chapter 4.2.4) uses the tuning parameters resulting in Chapter 4.1.4. to construct the six-month horizon forecast BCG Matrix.

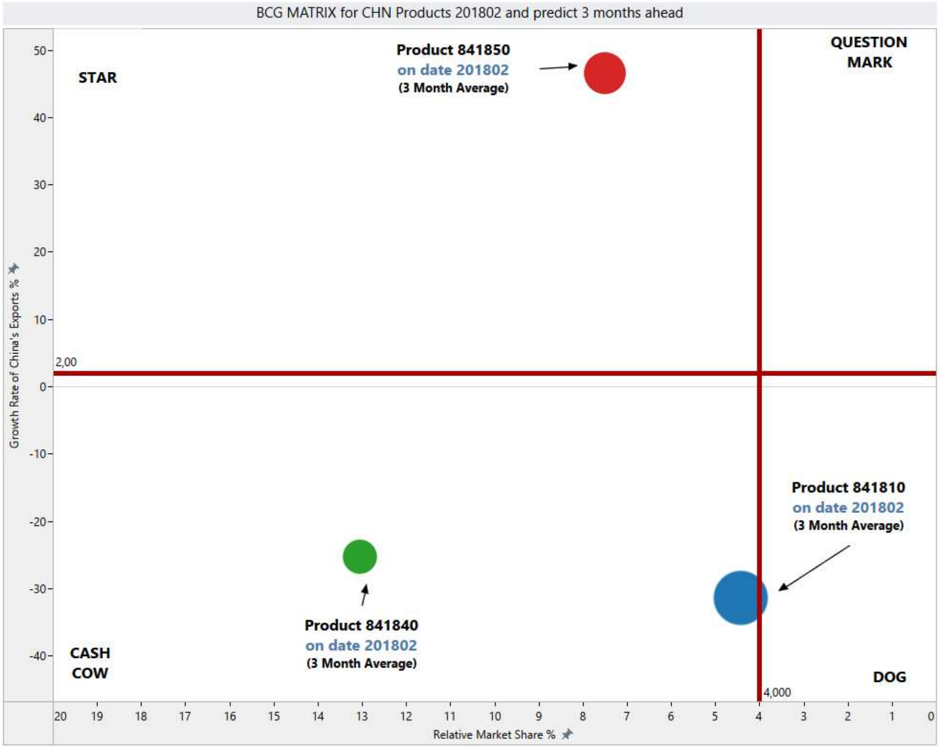
4.2.1. Illustration of the Construction of the Proposed BCG Matrix and its Components

After tuning the hyperparameters and selecting the models for all products, the next step is to apply these models to generate the BCG Matrix view. The case study considers three export products, and therefore, the BCG Matrix has three components. To demonstrate the proposed methodology, the BCG matrix analysis is conducted for all months between October 2017 and March 2018 for both Turkish and Chinese companies. To explain the components of the implemented BCG matrix, its steps are given in detail in this chapter.

The first step is to calculate the position of the company as of a given month. To illustrate the methodology, February 2018 is used as an example in this chapter. Note that all numbers and figures showed in this chapter are calculated with real values and forecasted values of the proposed methodology. In these calculations, the sales of the Chinese company to the UK are assumed to be \$500,000, \$200,000 and \$300,000 for the products 84.18.10, 84.18.40 and 84.18.50, respectively.

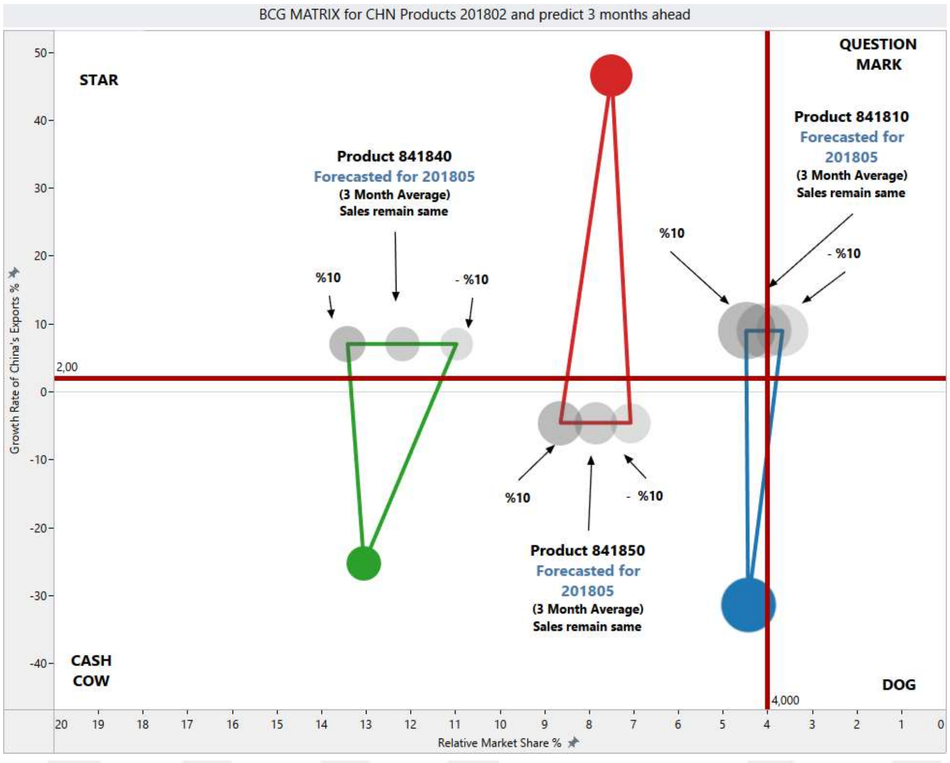
Colored spots in the BCG matrix show the current situation as of February 2018 in Figure 55. It shows that Product 84.18.10 is on the border between Cash Cows and Dogs. Product 84.18.40 stays in the center of the Cash Cows quadrant. Product 84.18.50 is in the Stars quadrant.

Figure 55. The Generated BCG Matrix for China Products in February 2018



In Figure 56, the grey spots represent the forecasted market situations of May 2018 (three-month forecasting horizon from February to May 2018). It shows that Product 84.18.10 is expected to move at the border between Stars and Question Marks. Product 84.18.40 moves from the center of Cash Cows to the bottom of Stars quadrant. Product 84.18.50 moves toward upper parts of the Cash Cows quadrant. Company sales to the UK are calculated in three scenarios. Grey dots with +10 and -10 annotations represent the sensitivity analysis in sales. The best scenario is 10 percent increase in sales, the worst scenario is 10 percentage of decrease in sales and medium scenario assumes the sales to remain the same.

Figure 56. The Generated BCG Matrix for China Products in February 2018 with Three-month horizon forecast



In Figure 57, grey spots represent the real market situations of May 2018. It shows similar moves as the predictions presented in Figure 56. According to real data, Product 84.18.10 moves at the border between Stars and Question Marks. Product 84.18.40 moves from the center of Cash Cows towards Stars quadrant, but stays in Cash Cows quadrant. Product 84.18.50 moves at the border between Cash Cows and Stars quadrant. Similar to Figure 56, grey dots with +10 and -10 annotations represent the sensitivity analysis in sales.

Figure 57. The Generated BCG Matrix for China Products on February 2018 with the Three-Month Horizon Real Values

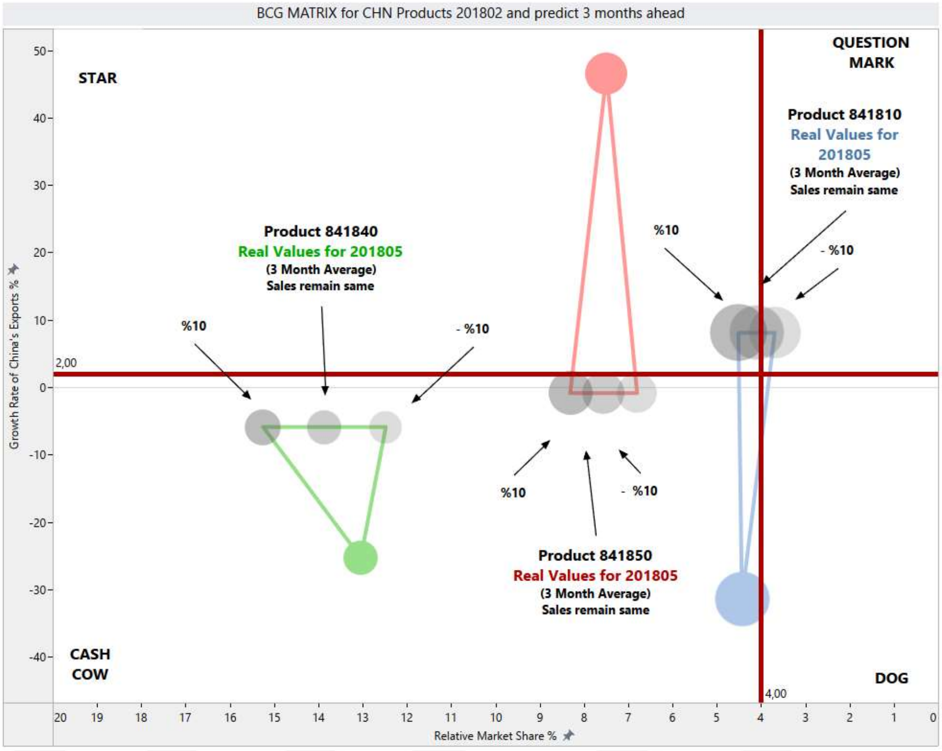
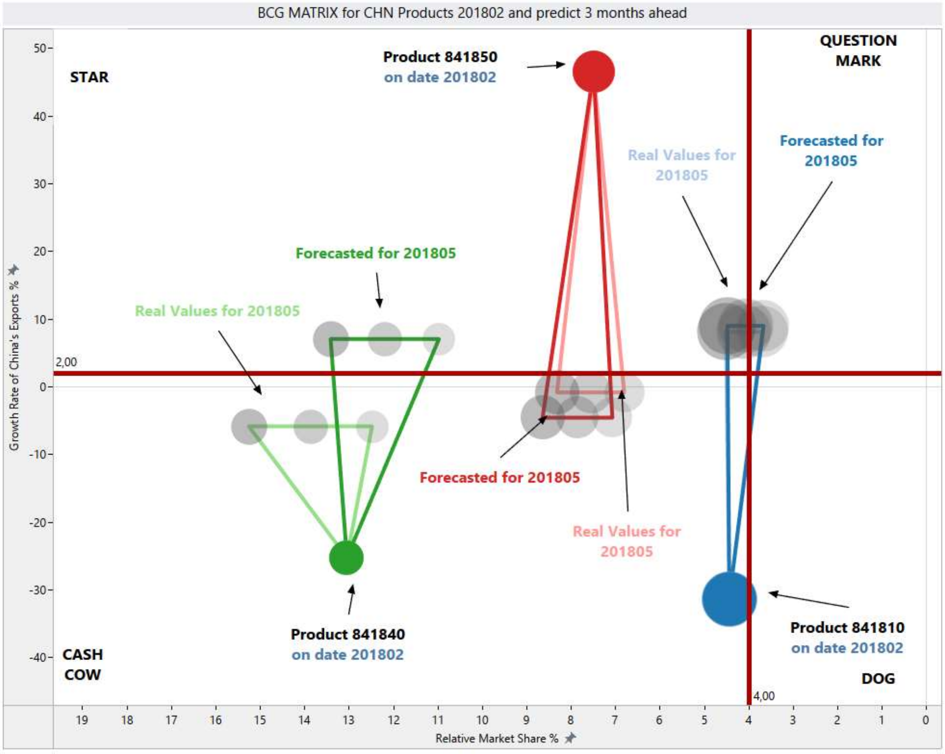


Figure 58 combines both forecasted and real market information obtained in Figure 56 and Figure 57. By using these real and forecasted values, a visual evaluation of the results can be performed. For Products 84.18.10 and 84.18.50, the forecasted and real values almost overlap. For Product 84.18.40, there is a small gap between real and forecasted values. Although forecasted results do not precisely show the real values, it gives an insight into the move of the market.

Figure 58. The Generated BCG Matrix for China Products in February 2018 with Three-month horizon forecast and Real Values



The BCG matrix figures shown in the next sections are constructed similar to the example given in this section. The next chapter explains the calculations of the numbers used in the BCG matrix.

4.2.2. Calculation of BCG Matrix Components for Turkish and Chinese Hypothetical Companies

In this chapter, the calculations to generate the BCG matrix are given. Two sample BCG matrices are calculated with the formulas in detail. The first BCG matrix is for a Turkish Company as of November 2017 and the second one is for a Chinese Company as of January 2018. Note that these dates are chosen arbitrarily to demonstrate the calculations of the methods.

4.2.2.1. Calculation of BCG Matrix Components for Hypothetical Turkish Company

The forecasted trade volumes are calculated by using hyperparameters presented in Table 26. BCG matrix for Turkish Company is generated by using Table 30, Table 31, Table 32 and Table 33. Table 30 shows the data and calculations of the colored spots where the position of the company as of a given month. Table 31 gives the calculation and data of the grey spot in the middle, where the forecasted position of the company with the same sales performance. The forecasted position of the company with a 10 percentage of decrease in sales is given in Table 32. Table 33 gives the calculation and data of the grey spot on the left-hand side, where the forecasted position of the company with a 10 percent increase. Table 30 uses the relative market share of the company and the growth rate of trade flow, where they are calculated by using Equations 1 and 2. Table 31, Table 32 and Table 33 use the forecasted market share of the company and forecasted growth rate of trade flow where they are calculated by using Equations 3 and 4. Equations use company sales to the importing country variable. All calculations assume the sales of the case company to the UK are$500,000, $200,000 and $300,000 for the products 84.18.10, 84.18.40 and 84.18.50, respectively. The next variable used in these equations is the real trade flow. And, the forecasted trade flow variable represents the forecast results of the ML model.

$Y^t{}_{j\,k}$= Company Sales to Import Country *j* for Product *k* in Period *t*

$X^t{}_{i\,j\,k}$= Real Trade Flow Volume between Source (Exporting) Country *i* and Target (Importing) Country *j* for Product *k* in Period *t*

$\hat{X}^t{}_{i\,j\,k}$= Forecasted Trade Flow Volume between Source (Exporting) Country *i* and Target (Importing) Country *j* for Product *k* in Period *t*

Equation 1 calculates the Relative Market Share (RMS) of the company. The sales are assumed to be the same in the forecast period. Therefore, only a single value is taken in the equation. As the market value, the average of the previous three months until the calculated month is taken. For November 2017 market value, market values of September, October, and November are averaged.

$$RMS^{\,t}{}_{i_0\,j_0\,k} = \frac{Y^t{}_{i_0\,j_0\,k}}{(X^t{}_{i_0\,j_0\,k}+X^{t-1}{}_{i_0\,j_0\,k}+X^{t-2}{}_{i_0\,j_0\,k})/3} \tag{1}$$

Equation 2 calculates the Growth Rate of Trade Flow (GRTF). To calculate growth, the market value of the calculated month and previous market value are taken.

The market value of the calculated month is the average of the previous three months until the calculated month. Previous market value is the average value of the previous three months of the calculated month. For November 2017 previous market value, market values of June, July and August are averaged.

$$GRTF^{t}_{i_0 j_0 k} = \frac{(X^{t}_{i_0 j_0 k} + X^{t-1}_{i_0 j_0 k} + X^{t-2}_{i_0 j_0 k}) - (X^{t-3}_{i_0 j_0 k} + X^{t-4}_{i_0 j_0 k} + X^{t-5}_{i_0 j_0 k})}{(X^{t-3}_{i_0 j_0 k} + X^{t-4}_{i_0 j_0 k} + X^{t-5}_{i_0 j_0 k})} \tag{2}$$

Equation 3 calculates Forecasted Relative Market Share (FRMS). Sales are considered to be the same throughout the forecast period. Therefore, only a single value is taken in the equation. As a forecasted market value, an average of forecasted market values of three months forward starting from the calculated month is taken. For November 2017 forecasted market value, forecasted market values of December 2017, January 2018 and February 2018 are averaged.

$$FRMS^{\mathbf{t}}_{i_0 j_0 k} = \frac{Y^{t}_{i_0 j_0 k}}{(\hat{X}^{t+3}_{i_0 j_0 k} + \hat{X}^{t+2}_{i_0 j_0 k} + \hat{X}^{t+1}_{i_0 j_0 k})/3} \tag{3}$$

Equation 4 is used to calculate the Forecasted Growth Rate of Trade Flow (FGRTF). To calculate forecasted growth, the forecasted market value of the calculated month and market value of the calculated month is taken. As Forecasted Market Value, an average of forecasted market values of the next three months starting from the calculated month is taken. The market value of the calculated month is the average of the previous three months until the calculated month.

$$FGRTF^{t}_{i_0 j_0 k} = \frac{(\hat{X}^{t+3}_{i_0 j_0 k} + \hat{X}^{t+2}_{i_0 j_0 k} + \hat{X}^{t+1}_{i_0 j_0 k}) - (X^{t}_{i_0 j_0 k} + X^{t-1}_{i_0 j_0 k} + X^{t-2}_{i_0 j_0 k})}{(X^{t}_{i_0 j_0 k} + X^{t-1}_{i_0 j_0 k} + X^{t-2}_{i_0 j_0 k})} \tag{4}$$

Table 30 shows the position of the company as of a given month using real values. For Product 84.18.10, GRTF is found as -6 percent and RMS is found as 4.8 percent. For Product 84.18.40, GRTF and RMS are found as 8 percent and 7.1 percent, respectively. For Product 84.18.50, GRTF found as -39 percent and RMS found as 13.8 percent. These calculations based on the averages in three months.

Table 30. The Calculations of the Colored Spots (Real Data) of the BCG Matrix Components for Turkish Company as of November 2017

Product Code (*k*)	Year-Month (*t*)	Company Sales to UK ($ 1,000) (*Y*)	Bilateral Trade Flow Volume from Turkey to the UK ($ 1,000) (*X*) Actual	Quarterly Average of Trade Volume ($ 1,000) ($\bar{X}$) Actual	Growth Rate of Trade Flow (*GRTF*)	Relative Market Share (*RMS*)
84.18.10	2017-06	500	10,765	11,161	$\frac{10{,}465 - 11{,}161}{11{,}161} =$ -6%	$\frac{500}{10{,}465} =$ 4.8%
	2017-07	500	8,186			
	2017-08	500	14,531			
	2017-09	500	10,606	10,465		
	2017-10	500	13,107			
	2017-11	500	7,683			
84.18.40	2017-06	200	1,948	2,631	$\frac{2.816 - 2{,}631}{2{,}631} =$ 8%	$\frac{200}{2.816} =$ 7.1%
	2017-07	200	2,030			
	2017-08	200	3,826			
	2017-09	200	2,631	2,816		
	2017-10	200	3,328			
	2017-11	200	2,489			
84.18.50	2017-06	300	3,547	3,567	$\frac{2.174 - 3{,}567}{3{,}567} =$ -39%	$\frac{300}{2{,}174} =$ 13.8%
	2017-07	300	3,165			
	2017-08	300	3,990			
	2017-09	300	2,159	2,174		
	2017-10	300	2,352			
	2017-11	300	2,010			

Table 31 gives the forecasted position of the company assuming the same sales performance. For Product 84.18.10, GRTF moves from -6 percent to -20 percent and RMS moves from 4.8 percent to 6 percent. For Product 84.18.40, GRTF changes from 8 percent to -24 percent and RMS moves from 7.1 percent to 9.3 percent. For Product 84.18.50, GRTF changes from -39 percent to -12 percent and RMS moves from 13.8 percent to 15.6 percent.

Table 31. The Calculations of the Grey Spots (Forecasted Data) of the BCG Matrix Components for Turkish Company as of November 2017 with the Same Sales Volume.

Product Code (*k*)	**Year-Month** (*t*)	**Company Sales to UK ($ 1,000)** (*Y*)	**Bilateral Trade Flow Volume from Turkey to the UK ($ 1,000)** (*X*) **Actual and** (*X*) **Forecasted**	**Quarterly Average of Trade Volume ($ 1,000)** $(\bar{X})$ **and** $(\bar{\bar{X}})$	**Growth Rate of Trade Flow** (*GRTF*)	**Relative Market Share** (*RMS*)
84.18.10	2017-09	500	10,606	10,465	$\frac{8,328 - 10,465}{10,465} =$	$\frac{500}{8,328} =$
	2017-10	500	13,107			
	2017-11	500	7,683			
	2017-12	500	8,329	8,328	-20%	6%
	2018-01	500	8,475			
	2018-02	500	8,181			
84.18.40	2017-09	200	2,631	2,816	$\frac{2,144 - 2,816}{2,816} =$	$\frac{200}{2,144} =$
	2017-10	200	3,328			
	2017-11	200	2,489			
	2017-12	200	2,151	2,144	-24%	9.3%
	2018-01	200	2,119			
	2018-02	200	2,161			
84.18.50	2017-09	300	2,159	2,174	$\frac{1,922 - 2,174}{2,174} =$	$\frac{300}{1,922} =$
	2017-10	300	2,352			
	2017-11	300	2,010			
	2017-12	300	1,892	1,922	-12%	15.6%
	2018-01	300	1,992			
	2018-02	300	1,882			

Table 32 gives the forecasted position of the company according to a 10 percent decrease in sales performance. The sales of the company to the UK are assumed to be $450,000, $180,000 and $270,000 for the products 84.18.10, 84.18.40 and 84.18.50, respectively. GRTF for all products remain the same. A decrease in sales effecting the RMS of the company. RMS of Product 84.18.10 is 5.4 percent where it was 6 percent with the same amount of sales. RMS of Product 84.18.40 is 8.4 percent where it was 9.3 percent with the same amount of sales. RMS of Product 84.18.50 found as 14 percent where it was 15.6 percent with no change in sales. In parallel to the decrease in sales, a decrease in RMS can be seen in each product.

Table 32. The Calculations of the Grey Spots (Forecasted Data) of the BCG Matrix Components for Turkish Company as of November 2017 with a 10 Percent Decrease in Sales Volume.

Product Code (k)	**Year-Month** (t)	**Company Sales to UK ($ 1,000)** (Y)	**Bilateral Trade Flow Volume from Turkey to the UK ($ 1,000) (X) Actual and (X) Forecasted**	**Quarterly Average of Trade Volume ($ 1,000) ($\bar{X}$) and ($\bar{\bar{X}}$)**	**Growth Rate of Trade Flow** ($GRTF$)	**Relative Market Share** (RMS)
84.18.10	2017-09	500	10,606	10,465	$\frac{8,328 - 10,465}{10,465} =$	$\frac{500}{8,328} =$
	2017-10	500	13,107			
	2017-11	500	7,683			
	2017-12	450	8,329	8,328	-20%	5.4%
	2018-01	450	8,475			
	2018-02	450	8,181			
84.18.40	2017-09	200	2,631	2,816	$\frac{2,144 - 2,816}{2,816} =$	$\frac{200}{2,144} =$
	2017-10	200	3,328			
	2017-11	200	2,489			
	2017-12	180	2,151	2,144	-24%	8.4 %
	2018-01	180	2,119			
	2018-02	180	2,161			
84.18.50	2017-09	300	2,159	2,174	$\frac{1,922 - 2,174}{2,174} =$	$\frac{300}{1,922} =$
	2017-10	300	2,352			
	2017-11	300	2,010			
	2017-12	270	1,892	1,922	-12%	14%
	2018-01	270	1,992			
	2018-02	270	1,882			

Table 33 gives the forecasted position of the company according to a 10 percent increase in sales performance. The sales of the company to the UK are assumed to be $550,000, $220,000 and $330,000 for the products. GRTF for all products remain the same as Table 32. An increase in sales also increases the RMS of the company. RMS of Product 84.18.10 is 6.6 percent where 6 percent with the same amount of sales. RMS of Product 84.18.40 found as 10.3 percent where it was 9.3 percent with the same amount of sales. RMS of Product 84.18.50 is 17.2 percent where it was 15.6 percent with the same amount of sales. In parallel to the increase in sales, an increase in RMS can be seen in each product.

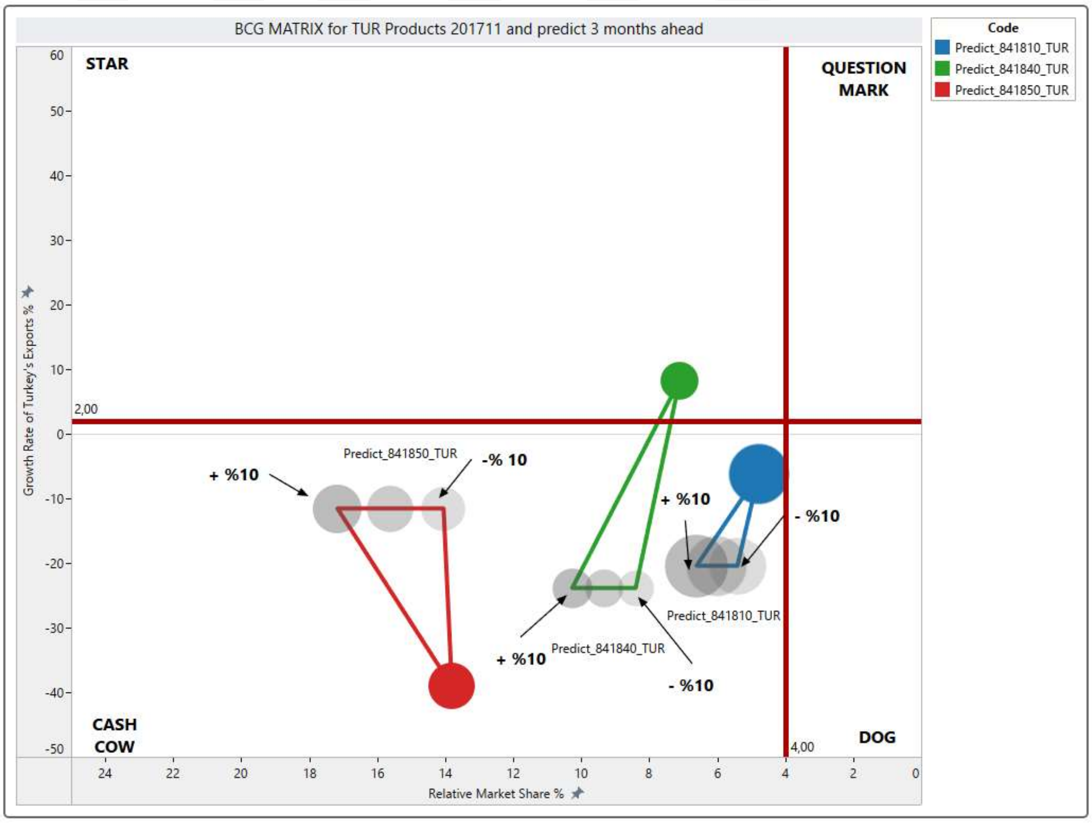
Table 33. The Calculations of the Grey Spots (Forecasted Data) of the BCG Matrix Components for Turkish Company as of November 2017 with a 10 Percent Increase in Sales Volume.

Product Code (*k*)	Year-Month (*t*)	Company Sales to UK ($ 1,000) (*Y*)	Bilateral Trade Flow Volume from Turkey to the UK ($ 1,000) (*X*) Actual and (*X*) Forecasted	Quarterly Average of Trade Volume ($ 1,000) ($\bar{X}$) and ($\bar{\bar{X}}$)	Growth Rate of Trade Flow (*GRTF*)	Relative Market Share (*RMS*)
84.18.10	2017-09	500	10,606	10,465	$\frac{8,328 - 10,465}{10,465} =$ -20%	$\frac{500}{8,328} =$ 6.6%
	2017-10	500	13,107			
	2017-11	500	7,683			
	2017-12	550	8,329	8,328		
	2018-01	550	8,475			
	2018-02	550	8,181			
84.18.40	2017-09	200	2,631	2,816	$\frac{2,144 - 2,816}{2,816} =$ -24%	$\frac{200}{2,144} =$ 10.3%
	2017-10	200	3,328			
	2017-11	200	2,489			
	2017-12	220	2,151	2,144		
	2018-01	220	2,119			
	2018-02	220	2,161			
84.18.50	2017-09	300	2,159	2,174	$\frac{1,922 - 2,174}{2,174} =$ -12%	$\frac{300}{1,922} =$ 17.2%
	2017-10	300	2,352			
	2017-11	300	2,010			
	2017-12	330	1,892	1,922		
	2018-01	330	1,992			
	2018-02	330	1,882			

Using the values that are found in Table 30 (Current Situation), Table 31 (Three-month horizon forecast with same sales), Table 32 (Three-month forecast horizon with 10 percent decrease in sales) and Table 33 (Three-month forecast horizon with 10 percent increase in sales), the next step is to construct the BCG matrix as presented in Figure 59.

According to Figure 59, the products 84.18.10 and 84.18.50 are in Cash Cows quadrant and product 84.18.40 is in Stars quadrant. The forecasted moves of 84.18.10 and 84.18.50 remain to stay in Cash Cows Quadrant. The main recommended strategy for Cash Cows is to maintain a certain investment level to keep them in Cash Cows Quadrant and provide a certain level of cash flow for supporting the company and the other business units (Kotler & Armstrong, 2018). Product 84.18.40 forecast shows a movement to the Cash Cows quadrant. For all three products, the market continues to shrink (negative growth rate). The forecasted move of Product 84.18.50 is different than the others. It shows that market shrinkage slows down for Product 84.18.50 while it speeds up for other products. Product 84.18.50 market shrinkage is from -39 percent to -12 percent. Product 84.18.40's forecasted move to Cash Cows quadrant shows the market shrinkage would start in the following months from 8 percent to -24 percent. Market shrinkage for Product 84.18.10 speeds up from -6 percent to -20 percent. If the amount of sales remains as in the current situation relative market share increases for Product 84.18.10.

Figure 59. The generated BCG Matrix for Turkish Company in November 2017 with Three-Month Horizon Forecast



In Chapter 4.2.2.2, the calculations to generate the BCG matrix for a hypothetical Chinese company are given. The BCG matrix is generated for a Chinese Company as of January 2018 using the three-month horizon forecast.

4.2.2.2. Calculation of BCG Matrix Components for Hypothetical Chinese Company

Forecasted trade volumes are calculated by using hyperparameters presented in Table 26. BCG matrix for Chinese Company is generated by using Table 34, Table 35, Table 36 and Table 37. Data and calculations of the colored spot are given in Table 34 where the Chinese company's position in a given month is located. Table 35 gives the calculations and data of the grey spot in the middle, where the forecasted position of the company with the same sales performance. The forecasted position of the company with a 10 percent decrease in sales is given in Table 36. Table 37 presents the calculations and data of the grey spot, where the forecasted position of the company with a 10 percent increase. In Table 34, Equation 1 and Equation 2 are used to calculate the relative market share of the company and the growth rate of trade flow. Table 35, Table 36 and Table 37 use the forecasted market share of the company and forecasted growth rate of trade flow where they are calculated by using Equation 3 and Equation 4. Table 34 shows the position of the Chinese company as of January. For Product 84.18.10 GRTF found as -23.9 percent and RMS found as 3.9 percent. For Product 84.18.40 GRTF calculated as -19.1 percent and RMS found as 11.6 percent. For Product 84.18.50 GRTF found as -3.1 percent and RMS calculated as 8.4 percent.

Table 34. The Calculations of the Colored Spots (Real Data) of the BCG Matrix Components for Chinese Company as of January 2018

Product Code (k)	Year-Month (t)	Company Sales to UK ($ 1,000) (Y)	Bilateral Trade Flow Volume from China to the UK ($ 1,000) (X) Actual	Quarterly Average of Trade Volume ($ 1,000) ($\bar{X}$) Actual	Growth Rate of Trade Flow ($GRTF$)	Relative Market Share (RMS)
84.18.10	2017-08	500	17,149			
	2017-09	500	16,072	16,761	$\frac{12{,}761 - 16{,}761}{16{,}761} =$	$\frac{500}{12{,}761} =$
	2017-10	500	17,063			
	2017-11	500	16,218		-23.9%	3.9%
	2017-12	500	11,329	12,761		
	2018-01	500	10,737			
84.18.40	2017-08	200	2,152			
	2017-09	200	2,187	2,139	$\frac{1{,}730 - 2{,}139}{2{,}139} =$	$\frac{200}{1{,}730} =$
	2017-10	200	2,077			
	2017-11	200	1,895		-19.1%	11.6%
	2017-12	200	1,813	1,730		
	2018-01	200	1,481			
84.18.50	2017-08	300	4,837			
	2017-09	300	3,226	3,689	$\frac{3{,}573 - 3{,}689}{3{,}689} =$	$\frac{300}{3{,}573} =$
	2017-10	300	3,003			
	2017-11	300	1,973		-3.1%	8.4%
	2017-12	300	5,063	3,573		
	2018-01	300	3,683			

Table 35 gives the forecasted position of the Chinese company assuming the same sales performance. Note that, similar to the Turkish company, the sales of the company are assumed to be $500,000, $200,000 and $300,000 for the products 84.18.10, 84.18.40 and 84.18.50, respectively. As given in the calculations of Table 35, for Product 84.18.10 GRTF moves from -23.9 percent to -1.4 percent and RMS changes from 3.9 percent to 4 percent. For Product 84.18.40, GRTF moves from -19.1 percent to -5.1 percent and RMS changes from 11.6 percent to 12.2 percent. For Product 84.18.50 GRTF moves from -3.1 percent to -7.5 percent and RMS changes from 8.4 percent to 9.1 percent.

Table 35. The Calculations of the Grey Spots (Forecasted Data) of the BCG Matrix Components for Chinese Company as of January 2018 with the Same Sales Volume.

Product Code (k)	**Year-Month** (t)	**Company Sales to UK ($ 1,000)** (Y)	**Bilateral Trade Flow Volume from China to the UK ($ 1,000)** (X) Actual and (X) Forecasted	**Quarterly Average of Trade Volume ($ 1,000)** ($\bar{X}$) and ($\bar{\bar{X}}$)	**Growth Rate of Trade Flow** ($GRTF$)	**Relative Market Share** (RMS)
84.18.10	2017-11	500	16,218	12,761	$\frac{12{,}581 - 12{,}761}{12{,}761} =$ -1.4%	$\frac{500}{12{,}581} =$ 4%
	2017-12	500	11,329			
	2018-01	500	10,737			
	2018-02	500	12,722	12,581		
	2018-03	500	12,152			
	2018-04	500	12,869			
84.18.40	2017-11	200	1,895	1,730	$\frac{1{,}642 - 1{,}730}{1{,}730} =$ -5.1%	$\frac{200}{1{,}642} =$ 12.2%
	2017-12	200	1,813			
	2018-01	200	1,481			
	2018-02	200	1,653	1,642		
	2018-03	200	1,566			
	2018-04	200	1,707			
84.18.50	2017-11	300	1,973	3,573	$\frac{3{,}304 - 3{,}573}{3{,}573} =$ -7.5%	$\frac{300}{3{,}304} =$ 9.1%
	2017-12	300	5,063			
	2018-01	300	3,683			
	2018-02	300	3,222	3,304		
	2018-03	300	3,361			
	2018-04	300	3,329			

Table 36 presents the forecasted position of the company according to 10 percent decrease in sales performance. The values of GRTF for all products remain the same. A decrease in sales also affects the RMS of the company. RMS of Product 84.18.10 is 3.6 percent where it was 4 percent with the same amount of sales. RMS of Product 84.18.40 is 11 percent where it was 12.2 percent in case that the sales were the same as the previous year. RMS of Product 84.18.50 is 4.2 percent where it was 9.1 percent with the same amount of sales. In parallel to the decrease in sales, a decrease in RMS can be seen in each product.

Table 36. The Calculations of the Grey Spots (Forecasted Data) of the BCG Matrix Components for Chinese Company as of January 2018 with a 10 Percent Decrease in Sales Volume.

Product Code (*k*)	**Year-Month** (*t*)	**Company Sales to UK ($ 1,000)** (*Y*)	**Bilateral Trade Flow Volume from China to the UK ($ 1,000) (*X*) Actual and (*X*) Forecasted**	**Quarterly Average of Trade Volume ($ 1,000)** $(\bar{X})$ **and** $(\bar{\bar{X}})$	**Growth Rate of Trade Flow** (*GRTF*)	**Relative Market Share** (*RMS*)
	2017-11	500	16,218			
	2017-12	500	11,329	12,761		
84.18.10	2018-01	500	10,737		$\frac{12,581 - 12,761}{12,761} =$	$\frac{450}{12,581} =$
	2018-02	450	12,722			
	2018-03	450	12,152	12,581	-1.4%	3.6%
	2018-04	450	12,869			
	2017-11	200	1,895			
	2017-12	200	1,813	1,730		
84.18.40	2018-01	200	1,481		$\frac{1,642 - 1,730}{1,730} =$	$\frac{180}{1,642} =$
	2018-02	180	1,653			
	2018-03	180	1,566	1,642	-5.1%	11%
	2018-04	180	1,707			
	2017-11	300	1,973			
	2017-12	300	5,063	3,573		
84.18.50	2018-01	300	3,683		$\frac{3,304 - 3,573}{3,573} =$	$\frac{300}{3,304} =$
	2018-02	270	3,222			
	2018-03	270	3,361	3,304	-7.5%	8.2%
	2018-04	270	3,329			

Table 37 summarizes the forecasted position of the company according to 10 percent of the increase in sales performance. GRTF for all products remain the same as Table 36. An increase in sales affects the RMS of the company. RMS of Product 84.18.10 is 4.4 percent where it was 4 percent with the same amount of sales. RMS of Product 84.18.40 is 13.4 percent where it was 12.2 percent in case the sales were the same as the previous year. RMS of Product 84.18.50 is 10 percent where it was 9.1 percent with the same amount of sales. As sales increase, RMS values also increase in all products.

Table 37. The Calculations of the Grey Spots (Forecasted Data) of the BCG Matrix Components for Chinese Company as of January 2018 with a 10 Percent Increase in Sales Volume.

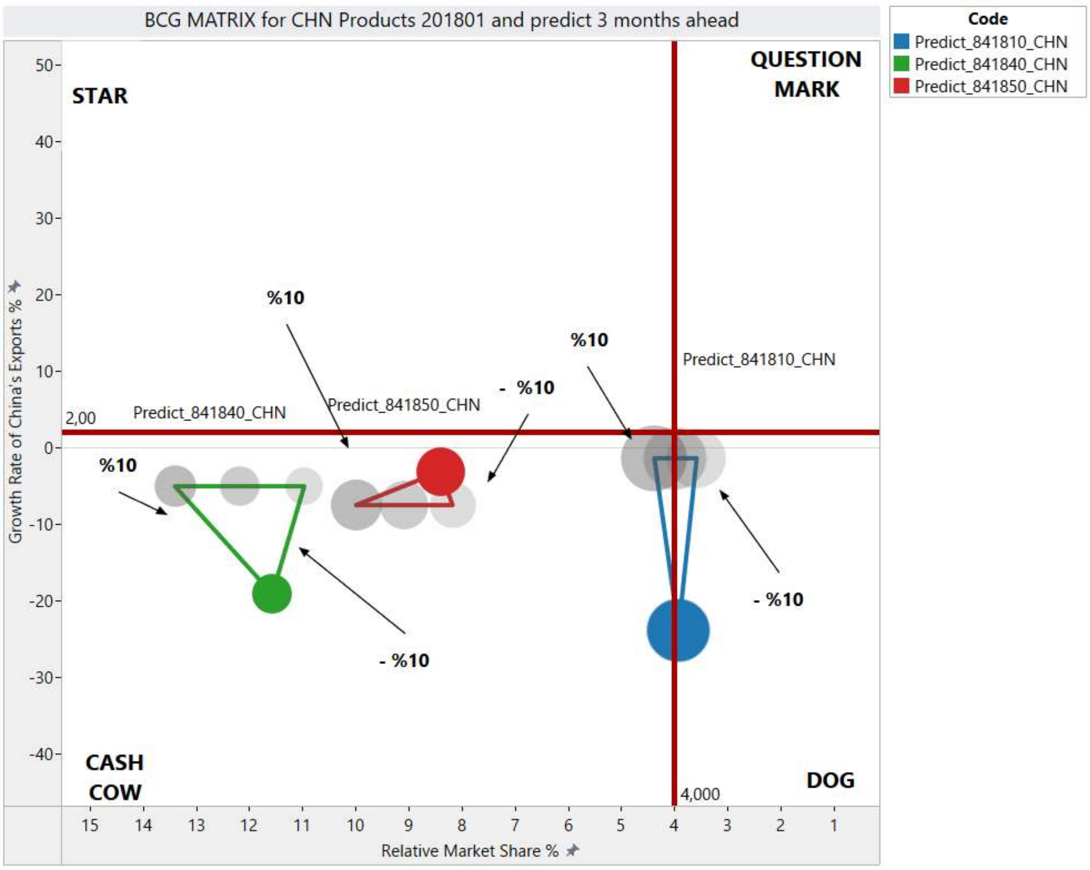
Product Code (*k*)	Year-Month (*t*)	Company Sales to UK ($ 1,000) (*Y*)	Bilateral Trade Flow Volume from China to the UK ($ 1,000) (*X*) Actual and (*X*) Forecasted	Quarterly Average of Trade Volume ($ 1,000) $(\bar{X})$ and $(\bar{\bar{X}})$	Growth Rate of Trade Flow (*GRTF*)	Relative Market Share (*RMS*)
84.18.10	2017-11	500	16,218			
	2017-12	500	11,329	12,761	$\frac{12,581-12,761}{12,761}=$	$\frac{550}{12,581}=$
	2018-01	500	10,737			
	2018-02	550	12,722		-1.4%	4.4%
	2018-03	550	12,152	12,581		
	2018-04	550	12,869			
84.18.40	2017-11	200	1,895			
	2017-12	200	1,813	1,730	$\frac{1,642-1,730}{1,730}=$	$\frac{220}{1,642}=$
	2018-01	200	1,481			
	2018-02	220	1,653		-5.1%	13.4%
	2018-03	220	1,566	1,642		
	2018-04	220	1,707			
84.18.50	2017-11	300	1,973			
	2017-12	300	5,063	3,573	$\frac{3,304-3,573}{3,573}=$	$\frac{330}{3,304}=$
	2018-01	300	3,683			
	2018-02	330	3,222		-7.5%	10%
	2018-03	330	3,361	3,304		
	2018-04	330	3,329			

Using the values that found in Table 34 (Current Situation), Table 35 (Three-month forecast horizon with same sales), Table 36 (Three-month forecast horizon with 10 percent decrease in sales) and Table 37 (Three-month forecast horizon with 10 percent increase in sales), the next step is to construct the BCG matrix as presented in Figure 60. The colored spots in the BCG matrix show the current situation as of January 2018 and grey spots show the forecasted market situations with a 10 percent increase and decrease in sales.

According to Figure 60, the products 84.18.40 and 84.18.50 are in Cash Cows quadrant and product 84.18.10 is on the border between Cash Cows and Dogs. The forecasted moves of 84.18.40 and 84.18.50 remain to stay in Cash Cows Quadrant.

The main recommended strategy for Cash Cows is to maintain a certain investment level to keep them in Cash Cows Quadrant and provide a certain level of cash flow for supporting the company and the other business units (Kotler & Armstrong, 2018). Product 84.18.10's forecast stays on the borderline. The next moves of product 84.18.10 should be carefully analyzed. If it becomes Cash Cows, then some investment should be made. If it turns into a Dog, a downsizing plan can be applied because this product has a large share in sales. For all three products, the market continues to get smaller (negative growth rate). For product 84.18.40, the figure indicates that the market shrinkage slows down from -19.1 percent to -5.1 percent. For product 84.18.10, a similar move is expected. On the other hand, market shrinkage slightly accelerates for product 84.18.50 (from -3.1 percent to 7.5 percent). The case study is limited to three products that are mainly located in Cash Cows Quadrant.

Figure 60. The Generated BCG Matrix for Chinese Company in 2018 January with Three-Month Horizon Forecast



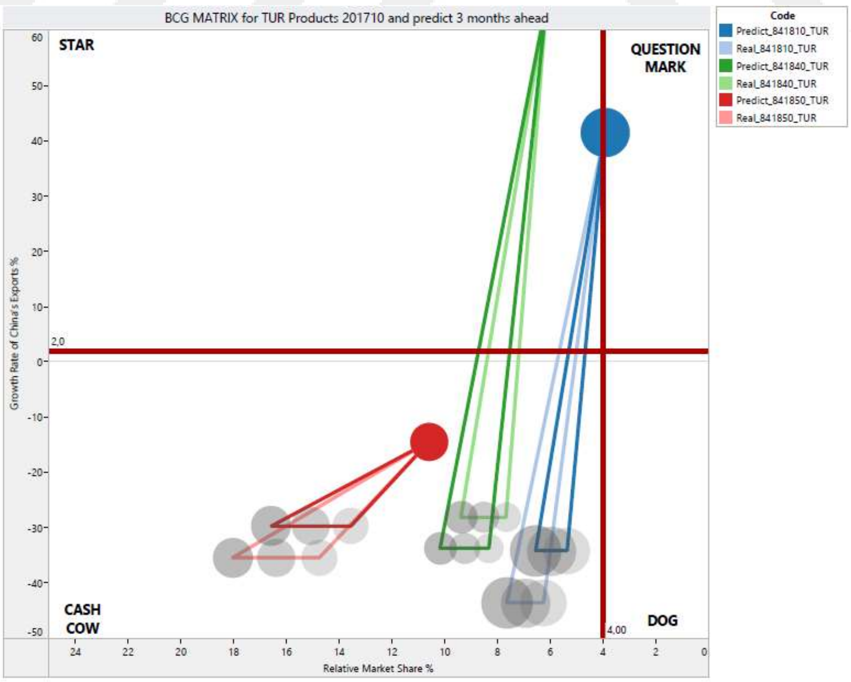
In Chapter 4.2.3, comparisons of BCG matrices for the hypothetical Turkish company case between October 2017 and March 2018 are given.

4.2.3. Comparison of BCG Matrix Results with Real Values between October 2017 and March 2018 for Turkish Company.

This chapter demonstrates the accuracy of the proposed model. To this end, the generated BCG Matrix includes both forecasted and real values for the consecutive months between October 2017 and March 2018.

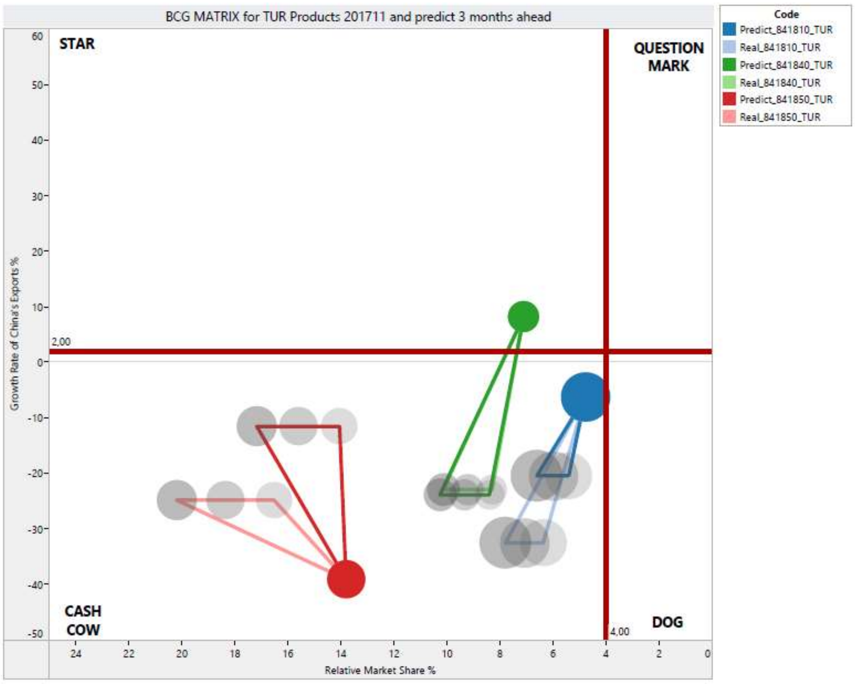
The first analysis is conducted as of October 2017. In Figure 61, the forecasted and real values overlap each other for all product types of the Turkish company. There is a small difference between the forecasted and real values, however, the results mainly give a correct insight into the direction of the market.

Figure 61. The Generated BCG Matrix for Turkish Company in October 2017 with Three-Month Horizon Forecasts



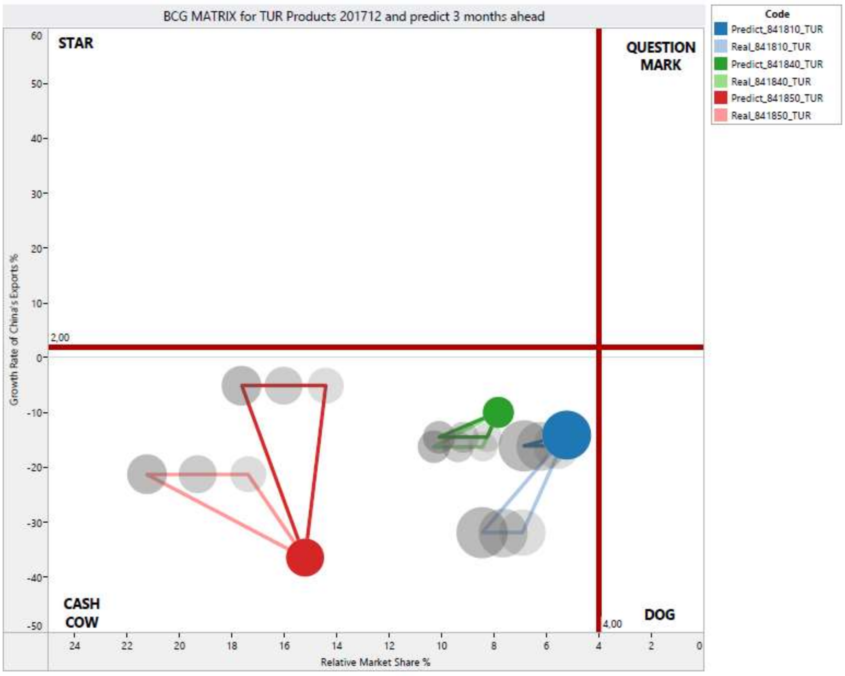
In Figure 62, the BCG matrix for Turkish Company in November 2017 is given. The forecasted and real values overlap for Product 84.18.40. For the other two products, there are slight differences between forecasted and real values but the matrix mainly gives insight in the correct direction.

Figure 62. The Generated BCG Matrix for Turkish Company in November 2017 with Three-Month Horizon Forecasts



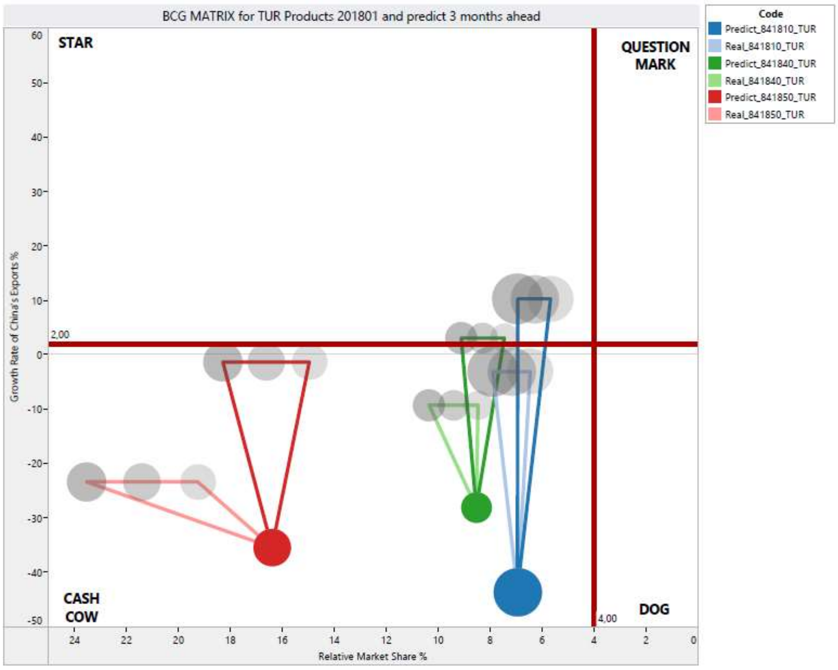
BCG matrix for Turkish Company as of December 2017 is given in Figure 63. Similar to the BCG matrix of November (Figure 62), the forecasted and real values overlap for Product 84.18.40. For Product 84.18.10, the BCG matrix shows a steady position in the following months as forecast where market shrinkage speeds up in real values. For Product 84.18.50, the forecasted slowdown in market shrinkage is not as low as the real values. The forecasted results expect a sharper slowdown in market shrinkage than the real values.

Figure 63. The Generated BCG Matrix for Turkish Company in December 2017 with Three-Month Horizon Forecasts



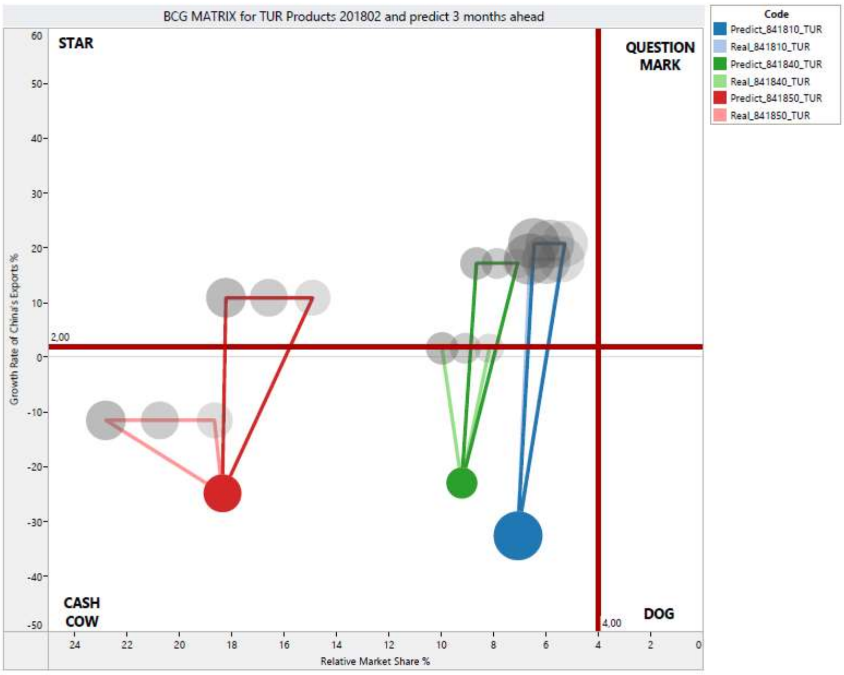
BCG matrix for Turkish Company as of January 2018 given in Figure 64. For Product 84.18.10 and Product 84.18.40, slowdown trends in the market shrinkage are forecasted and real values verify with this insight. For Product 84.18.50 forecasted slowdown in the market, shrinkage is not as high as real values. Nevertheless, the BCG Matrix gives more or less the right idea about the expected market behavior.

Figure 64. The Generated BCG Matrix for Turkish Company in January 2018 with Three-Month Horizon Forecasts



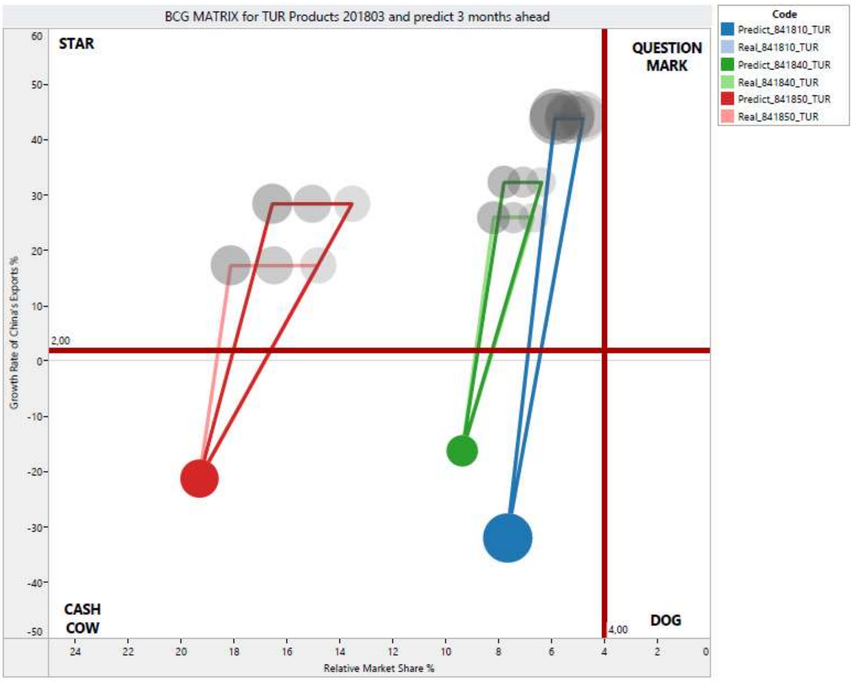
In Figure 65, the BCG matrix for Turkish Company in February 2018 is given. For Product 84.18.10 and Product 84.18.40 market, the moves from Cash Cows to Stars quadrant. Both the forecasted and real values indicate market growth would start. For Product 84.18.50 forecasted slowdown in market shrinkage is not as high as real values. But the direction gives the right idea about the market. This BCG matrix has similar properties with the BCG matrix of January 2018.

Figure 65. The Generated BCG Matrix for Turkish Company in February 2018 with Three-Month Horizon Forecasts



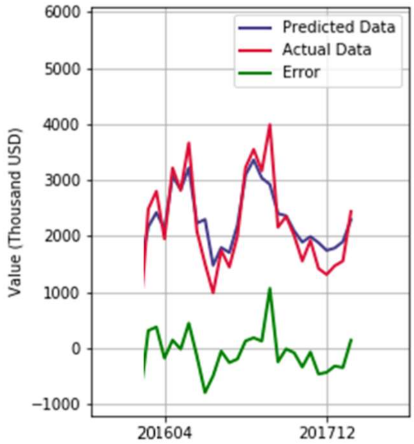
BCG matrix for Turkish Company in March 2018 is given in Figure 66. All products move from Cash Cows to Stars both in forecasted and real values. Real and forecasted values overlap each other. All products start the market growth phase in the forecasted and in real values.

Figure 66. The Generated BCG Matrix for Turkish Company in 2018 March and Three-Month Horizon Forecasts



In the BCG Matrices between November 2017 and February 2018, there are deviations between forecasted and real values for Product 84.18.50. The main reason for these deviations of the forecasts for Product 84.18.50 can be seen in Figure 49 in which, real values staring from October 2017 are “1924”, “1422”, “1315”, “1471”, “1555” and “2428”. Forecasted values for the same months are “1992”, “1882”, “1745”, “1788”, “1901” and “2294”. Figure 67 is the zoomed version of Figure 49 for the last months. It can easily be seen for the last months, there is a gap between the forecasted and real values. This gap is reflected in the BCG Matrices between November 2017 and February 2018 as well.

Figure 67. Zoomed to Last Months of Figure 49



BCG Matrices of the consecutive months show the proposed methodology to generate BCG Matrix gives the right insight. Even though the forecasts and real values do not fully overlap, the proposed methodology seems to give the correct prediction for the movement of the market in the future. This provides valuable insight for the companies in their strategic market decisions.

The next chapter explains the comparison results of the six-month horizon forecast and three-month forecast horizon for Chinese hypothetical company.

4.2.4. Comparison of Six-Month Horizon Forecast with Three-Month Horizon Forecast for the Chinese Company.

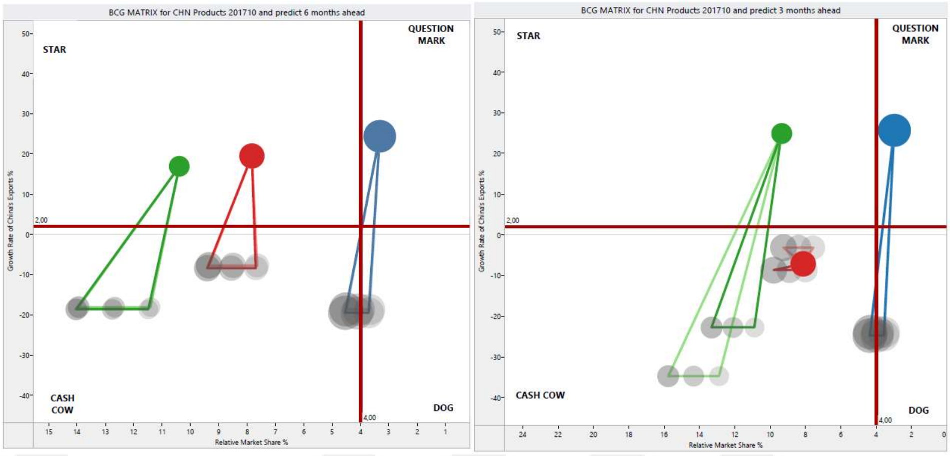
Generating a BCG matrix with the six-month forecast is an alternative for a three-month horizon forecast for decision-makers. This dissertation mainly focuses on a three-month horizon forecast as explained in Chapter 4.2.2.1. The first reason to select three-month horizon forecast is the achieved high R^2 scores on that horizon. The second reason is to prevent overlooking some sudden ups and downs in the forecasted values because six-month forecasting horizon can smooth some of the spikes in the real data as explained later in this chapter. Therefore, a comparison between three-month and six-month horizon forecasts are provided in this chapter to demonstrate the behavior of the model. In this comparison, the BCG Matrices of the Chinese company are generated between October 2017 and December 2017.

The forecasted trade volumes are calculated by using RF and its hyperparameters presented in Table 29. To build a six-month horizon forecast BCG matrix, the calculations of the colored and grey spots should be made with slight changes in Equations 1, 2, 3 and 4. To calculate the market size in current and future situations, the values of the successive six months are used instead of three months. In the calculation of the relative market share of the company, market value in Equation 1 takes the average six months of market values starting from the calculated month. Equation 2 is modified as market value takes six months backward starting from the calculated month and previous market value averages the previous six months of the calculated month. To calculate the forecasted market share, Equation 3 is modified as forecasted market value averages of forecasted market values in the next six months starting from the calculated month. Equation 4 is modified as Forecasted Market Value is average of forecasted market values six months starting from the calculated month and Market value of the calculated month is the average of previous six months starting from the calculated month.

BCG Matrix with six-month forecast results overlaps between the forecasted and real values. This can be observed in Figure 68a, Figure 69a and Figure 70a, where they show the BCG matrices of October 2017, November 2017, and December 2017, respectively. The main reason for this overlap between the forecasted and real values is that the calculations made by using an average of six months smooth the differences between forecasted and real values.

As of October 2017, a six-month horizon forecast forecasts that there is a market shrinkage in all products as given in Figure 68. In the three-month horizon forecast, Product 84.18.40 is predicted to have a sharper decline than in the six-month. Product 84.18.50 has a steady-state in the following three months where six-month BCG Matrix indicates sharp market shrinkage. Therefore, the six-month horizon forecast may cause to overlook these differences. However, the direction of the market is correct in both forecasting horizons.

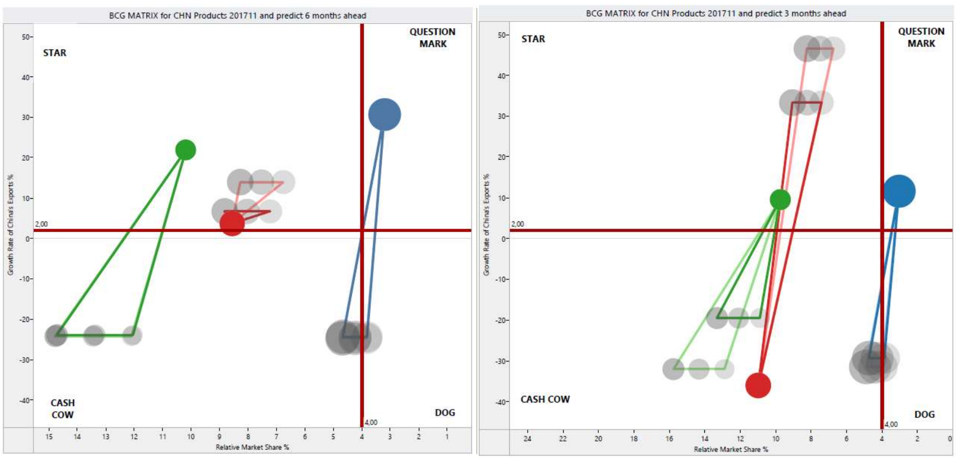
Figure 68. The Comparison of BCG Matrices in October 2017 with Six-Month and Three-Month Forecast Horizons for China Products



a) Six-month forecast horizon b) Three-month horizon forecast

As of November 2017, a six-month horizon forecast forecasts market shrinkages for Products 84.18.10 and 84.18.40 (Figure 69a), and slight market growth for Product 84.18.50 (Figure 69a). In the three-month horizon forecast, Product 84.18.50 is expected to have a sharper market growth where there is slight growth in the six-month version of the BCG Matrix (Figure 69b). Although market behavior in six-month and three-month horizon forecasts are similar for Product 84.18.10 and 84.18.40, six-month horizon forecast may mislead in short term for Product 84.18.50.

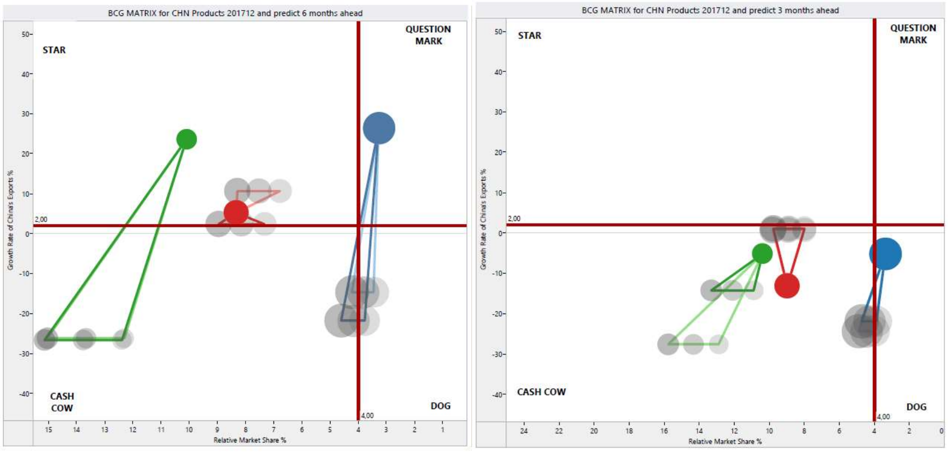
Figure 69. The Comparison of BCG Matrices in November 2017 with Six-Month and Three-Month Forecast Horizons for China Products



a) Six-month forecast horizon b) Three-month horizon forecast

As of December 2017, a six-month horizon forecast forecasts that there are market shrinkages in Products 84.18.10 and 84.18.40 (Figure 70a), whereas a slight market growth in Product 84.18.50 (Figure 70a). In the three-month horizon forecast (Figure 70a), Product 84.18.50 is expected to have a market growth where there is slight growth in six-month BCG Matrix. And, the forecasted market shrinkages for Product 84.18.10 and 84.18.40 are not as large as the forecasted ones in the six-month horizon forecast. Therefore, the six-month horizon forecast may cause to overlook these subtle differences.

Figure 70. The Comparison of BCG Matrices in December 2017 with Six-Month and Three-Month Forecast Horizons for China Products



a) Six-month forecast horizon b) Three-month horizon forecast

Having investigated the results presented in this chapter, the six-month horizon forecast gives more accurate information compared to a three-month horizon forecast. However, the fluctuations in export volumes are very high and looking at a six-month horizon may cause to overlook some sudden ups and downs. These ups and downs can be beneficial in developing market strategies and creating competitive advantage. Despite its slight accuracy disadvantage, three-month horizon forecast gives better insight to decision-makers on the movement and direction of the market in the shorter term.

CONCLUSION

Today's increased global competition forces companies to make better predictions and strategic decisions considering their business environments. Boston Consulting Group Matrix is one of the most well-known management tools that revolutionized the strategic management. However, it has some issues in practice, such as its difficulty to define the market share and growth rate. Also, it shows only the current business environment and does not give any insight into the future. By accurately predicting the future business environment, some future insight information can be added to the BCG Matrix. At this point, companies can identify new trade and business environments from the forecasts of the trade volumes between countries. However, accurate forecasting has become significantly harder due to the increased complexity of the globalization and competition between countries and supply chains. Using Big Data Analytics, accurate forecasts can be achieved and these forecasts can be used in strategic market analysis. The difficulties of applying parametric models in realistic applications and increasing capabilities of machine learning tools have driven companies to use machine learning in Big Data Analytics. To prove this hypothesis, a parametric model is applied to all cases. And, the results showed that ML is a better method for this kind of implementations.

In this dissertation, a holistic methodology for market analysis using Big Data Analytics is proposed. The proposed methodology uses machine learning methods and various open international data to forecast international trade volumes between countries. By using these forecasts as the future market data for exporters (or importers), some future insight information is added to the BCG Matrix for strategic market analysis. This is the first study to use Big Data Analytics and machine learning methods for strategic market analysis.

At the first stage of the methodology, the BCG Matrix is applied with the international trade data to investigate the current market situation. Then, the forecasted future values are added to the BCG Matrix by using the proposed Big Data Analytics method. To demonstrate the proposed methodology, two case studies of hypothetical Turkish and Chinese companies exporting refrigerators and freezers to the United Kingdom are presented. As demonstrated in the case studies, foreseeing market conditions of three-month ahead gives important managerial insights for companies.

With the case studies, it is shown that the potential users (exporters or importers) can easily adapt the proposed methodology of this dissertation to align their strategic market decisions according to the current and future market trends. To test the efficiency of the proposed methodology, three different sub-product groups within the main product group of refrigerators and freezers are selected in the case studies. The main rationale for selecting refrigerators and freezers as a product group is that they are closely related to the end-customer behavior and the second largest export chapter of Turkey. To forecast the trade volumes, 28 different factors are considered and the data (ranged from 2006 April to 2018 March) of each factor are obtained from openly available data sources (OECD, International Trade Center, etc.). Random Forest and Artificial Neural Networks methods are used as the main forecasting algorithms. The results showed that the Random Forest model yields more accurate forecasts than the Artificial Neural Networks model. However, both the RF and ANN models provide successful forecasts for trade volumes. Identifying important features, transforming the dependent variable with logarithm and adding past trade volume information significantly contribute to the accurate forecasts. Instead of using all features obtained from the data sources, using 50 or 75 percent of the features improves the forecasting accuracy for all products in both Random Forest and Artificial Neural Network models. Feature selection results show the total export volume of an exporter for a product is the most common feature in all products and country pairs. Lagged dependent variable, World total trade, Consumer Price Index of the exporter country, Consumer Price Index of the importer country and Composite Leading Indicator of the exporter country are the other most significant features. Also, some very common features used in trade forecasting, such as GDP and exchange rates, have lower significant effects than the above ones. Results show that Economic Politic Uncertainty does not appear in the top ten most important features for all six product and country combinations.

The results also showed that the tuning process helps to find better and more robust results. Therefore, feature selection and tuning procedures improve the forecast accuracy for all products. Results show that each product and country pair should be handled separately because of their independent dynamics. In general, very good R^2 results are achieved in Random Forest models. When the countries are compared, R^2 results for China are better than the ones for Turkey. In both cases, random forest models for Product 84.18.10 and Product 84.18.50 give better R^2 values than Product 84.18.40. The reason is that product dynamics are different, which may result in low

R^2 results on Product 84.18.40

This dissertation is focused on foreseeing a three-month horizon forecast, and a six-month horizon forecast is also used to test the proposed methodology. Although six-month yield better results, it overlooks sudden ups and downs. Therefore tree-month forecast horizon is recommended to capture market dynamics better.

In addition to these practical results, this dissertation shows that various types of decision support systems can be developed by using the proposed holistic framework for Big Data Analytics in Strategic Market Analysis. According to each supply chain function, there are some candidates for decision support systems to develop. For example, procurement function, supplier selection, sourcing cost improvement or sourcing risk management are some key areas in any supply chain that can benefit from the proposed methodology of this dissertation. Other candidate key areas on manufacturing function are product R&D, quality management or maintenance & diagnosis (Nguyen et al., 2018). Depending on the purpose and available data sets, and desired additional value, steps of this framework can be easily applied in the real-life applications. For some newly developed product types such as smart watches, sufficient data may not be available. In such cases, the data set can be enriched by using information about the substitute products that the newly developed product replaces. In this dissertation, BCG matrix is selected, but any kind of positioning map tool can be used according to purpose and available data sets. For example, the classical positioning map, where cost and quality are the main dimensions, can be employed for the future state in which each product can be forecasted with different scenarios. Also, some industry-specific positioning maps can be developed to incorporate into the proposed holistic BDA framework for addressing some industry-specific issues.

The proposed Machine Learning Modeling process used for forecasting can be applied for similar kinds of Big Data Analytics problems or just for only forecasting problems. In the above positioning map example, to find the forecasted positions of products in cost and quality, companies R&D expenditures, innovation indexes, labor cost changes, energy cost changes will be some candidate features for Machine Learning algorithms. After implementing the preliminary analysis and tuning processes, models will be ready to calculate forecasts. Depending on data frequency and results of Machine Learning models, any desired forecast-horizon can be applied to the previously determined positioning tool.

The future work of the dissertation can be done in two perspectives. The first one is the future works on the same domain (Product positioning with international trade data) by using a similar type of dataset, which would be an extension of this dissertation. The second one is the application of this approach on a different domain with data types and different datasets.

For the future works within the same domain, the developed approach can be applied in other country pairs with the same product types. As another extension, the proposed methodology can be applied to different products and country pairs to identify the significant factors affecting the bilateral trade volumes between countries. Lastly, other machine learning methods (e.g., LSTM) can be used to compare with RF and ANN. As a methodological extension, the proposed feature selection ranking algorithm can be extended to develop a new feature ranking in machine learning algorithms.

For the application of the proposed approach of this dissertation to different domains, various types of decision support systems can be developed for manufacturing, service and marketing systems where forecasting is essential. In these studies according to the purpose of Big Data Analytics, appropriate decision tools and compatible data sets should be matched. By applying the steps of the proposed framework, the results of Big Data Analytics can be applied to both practitioners and researchers of the field.

REFERENCES

Addo-Tenkorang, R., & Helo, P. T. (2016). Big data applications in operations/supply-chain management: A literature review. *Computers and Industrial Engineering*, *101*: 528–543.

Akhtar, S. (2003). Is there seasonality in Pakistan's merchandise exports and imports? The univariate modelling approach. *Pakistan Development Review*, *42*(1): 59–75.

Allende, H., Moraga, C., & Salas, R. (2002). Artificial neural networks in time series forecasting: A comparative analysis. *Kybernetika*, *38*(6): 685–707.

Ansoff, H. (1957). Strategies for Diversification. *Harvard Business Review*, *355*(5): 113–124.

Awad, M., & Khanna, R. (2015). *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Apress.

Ayankoya, K., Calitz, A. P., & Greyling, J. H. (2016). Real-time grain commodities price predictions in South Africa: A big data and neural networks approach. *Agrekon*, *55*(4): 483–508.

Barbosa, M. W., Vicente, A. de la C., Ladeira, M. B., & de Oliveira, M. P. V. (2018). Managing supply chain resources with big data analytics: A systematic review. *International Journal of Logistics Research and Applications*, *21*(3): 177–200.

Bohanec, M., Robnik-Sikonja, M., & Kljajic Borstnar, M. (2017). Decision-making framework with double-loop learning through interpretable black-box machine learning models. *Industrial Management & Data Systems*, *117*(7): 1389–1406.

Bovi, M., & Cerqueti, R. (2016). Forecasting macroeconomic fundamentals in economic crises. *Annals of Operations Research*, *247*(2): 451–469.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1): 5–32.

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Points of significance: Statistics Versus Machine Learning. *Nature Methods*, *15*(5): 233–234.

Calandro, J. C., & Lane, S. L. (2007). A new competitive analysis tool: The relative profitability and growth matrix. *Strategy and Leadership*, *35*(2): 30–38.

Chen, C. P., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, *275*: 314–347.

Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the use of big data analytics affects value creation in supply chain management. *Journal of Management Information Systems*, *32*(4): 4–39.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, *19*(2): 171–209.

Choi, T. M., Wallace, S. W., & Wang, Y. (2018). Big Data Analytics in Operations Management. *Production and Operations Management*, *27*(10): 1868–1883.

Chokmani, K., Khalil, B. M., Ouarda, T. B. M. J., & Bourdages, R. (2007). Estimation of River Ice Thickness Using Artificial Neural Networks. *Proceedings - 14th Workshop on the Hydraulics of Ice Covered Rivers, Quebec City, CGU-HS CRIPE*, (July 2014): 1–12.

Chong, A. Y. L., Li, B., Ngai, E. W., Ch'ng, E., & Lee, F. (2016). Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. *International Journal of Operations & Production Management*, *36*(4): 358–383.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *ArXiv Preprint ArXiv:1412.3555*, 1–9. Retrieved from http://arxiv.org/abs/1412.3555 (date accessed: 20/07/2019).

Co, H. C., & Boosarawongse, R. (2007). Forecasting Thailand's rice export: Statistical techniques vs. artificial neural networks. *Computers and Industrial Engineering*, *53*(4): 610–627.

Council, Customs, & Co-operation. (2017). Harmonized commodity description and coding system (HS). Retrieved from https://unstats.un.org/unsd/tradekb/Knowledgebase/50018/Harmonized-Commodity-Description-and-Coding-Systems-HS (date accessed: 20/07/2019).

CPB. (2019). World Trade Monitor. Retrieved from https://www.cpb.nl/en/worldtrademonitor (date accessed: 20/07/2019).

Dale, C., & Bailey, V. B. (1982). A Box-Jenkins model for forecasting U. S. merchandise exports. *Journal of International Business Studies*, *13*(1): 101–108.

Davenport, T. H., & Dyché, J. (2013). Big data in big companies. *International Institute for Analytics*, (3).

Delen, D., & Zolbanin, H. M. (2018). The analytics paradigm in business research. *Journal of Business Research*, *90*: 186–195.

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): 196–204.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*: *14*(2): 179–211.

Emang, D., Shitan, M., Abd. Ghani, A. N., & Noor, K. M. (2010). Forecasting with univariate time series models: A case of export demand for peninsular Malaysia's moulding and chipboard. *Journal of Sustainable Development*, *3*(3): 157–161.

EPU. (2019). Economic Policy Uncertainty. Retrieved from https://www.policyuncertainty.com/ (date accessed: 20/07/2019).

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*(2), 897–904.

Fan, J., Han, F., & Liu, H. (2014). NIH Public Access. *National Science Review*, *1*(2): 293–314.

Fang, K., Jiang, Y., & Song, M. (2016). Customer profitability forecasting using Big Data analytics: A case study of the insurance industry. *Computers and Industrial Engineering*, *101*: 554–564.

Feigenbaum, E. A., & Feldman, J. (1963). Some Studies in Machine Learning using the Game of Checkers. In *Computers and Thought* (pp. 71–105). New York: McGraw-Hill.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2): 137–144.

Ge, Y., & Wu, H. (2019). Prediction of corn price fluctuation based on multiple linear regression analysis model under big data. *Neural Computing and Applications*, *4*(December 2010).

Gens, F. (2012). *IDC Predictions 2012 : Competing for 2020*. *Top 10 Predictions* (Vol. 1). Retrieved from https://www.virtustream.com/sites/default/files/IDCTOP10Predictions2012.pdf (date accessed: 20/07/2019).

Gens, F., Philip, C., & Bill, F. (2018). *Worldwide IT Industry 2018 Predictions*. *IDC FutureScape*. Retrieved from https://www.idc.com/getdoc.jsp?containerId=US43171317 (date accessed: 20/07/2019).

Getscher, A. (2017). Six Actionable Steps to Build a Strategic Marketing Plan at

Your Organization. Retrieved from https://www.bizible.com/blog/strategic-marketing-plan (date accessed: 20/07/2019).

Gite, P., & Roy, C. K. (2014). Export Markets ’ Segmentation , Performance and Marketing of Indian Carpet Industry : A BCG Matrix Approach, *6*(11): 28–33.

Grömping, U. (2009). Variable importance assessment in regression: Linear regression versus random forest. *American Statistician*, *63*(4): 308–319.

Gupta, S., & Kashyap, S. (2015). Forecasting inflation in G-7 countries: An application of artificial neural network. *Foresight*, *17*(1): 63–73.

Hassan, S. S., & Craft, S. (2012). Examining world market segmentation and brand positioning strategies. *Journal of Consumer Marketing*, *29*(5): 344–356.

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*: 72–80.

Hazen, B. T., Skipper, J. B., Boone, C. A., & Hill, R. R. (2018). Back in business: operations research in support of big data analytics for operations and supply chain management. *Annals of Operations Research*, *270*(1–2): 201–211.

Henkle, H. H., & Lubetzky, S. (1946). The Future of the Research Library. *College and Research Libraries*, *7*(3): 301–308.

Hofmann, E. (2017). Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect. *International Journal of Production Research*, *55*(17): 5108–5126.

Hu, H., Wen, Y., Chua, T., & LI, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access*, *2*: 652–687.

Hunt, E., Minsky, M., & Papert, S. (2006). Perceptrons. In *The American Journal of Psychology* (Vol. 84, pp. 11218–11221).

International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Ioana, A., Mirea, V., & Balescu, C. (2009). Analysis of service quality management in the materials industry using the bcg matrix method. *Amfiteatru Economic Journal*, *11*(26): 270–276.

Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, *10*(3): 215–236.

Kabadurmus, O., Erdogan, M. S., & Tasgetiren, M. F. (2017). Design of multi-product multi-period two-echelon supply chain network to minimize bullwhip effect through differential evolution. *2017 IEEE Congress on Evolutionary Computation (CEC)*: 789–796.

Kangas, K., & Baudin, A. (2003). *Modelling and projections of forest products demand, supply and trade in Europe*. *GENEVA TIMBER AND FOREST DISCUSSION PAPERS* (Vol. 30). Retrieved from https://www.unece.org/fileadmin/DAM/timber/docs/efsos/03-sept/dp-30.pdf (date accessed: 20/07/2019).

Kargbo, J. M. (2007). Forecasting agricultural exports and imports in South Africa. *Applied Economics*, *39*(16): 2069–2084.

Keck, A., Raubold, A., & Truppia, A. (2010). Forecasting international trade: A time series approach. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, *2009*(2): 157–176.

Khan, T. (2011). Identifying an appropriate forecasting model for forecasting total import of Bangladesh. *Statistics in Transition New Series*, *12*(1): 179–192.

Kotler, P., & Armstrong, G. (2018). *Principles of marketing, Global edition (17th)*. London: Pearson.

Laney, D., & Kart, L. (2012). *Emerging Role of the Data Scientist and the Art of Data Science*. Retrieved from https://www.gartner.com/doc/1955615?ref=mrktg-srch (date accessed: 20/07/2019).

Laptev, N., Yosinski, J., Li, E. L., & Smyl, S. (2017). Time-series extreme event forecasting with neural networks at Uber. In *International Conference on Machine Learning* (pp. 1–5).

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, *52*(2): 21–32.

Law, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management*, *21*(4): 331–340.

Li, K., Yu, N., Li, P., Song, S., Wu, Y., Li, Y., & Liu, M. (2017). Multi-label spacecraft electrical signal classification method based on DBN and random forest. *PLoS ONE*, *12*(5): 1–19.

Marwah, K. (1976). A World model of international trade: Forecasting market shares and trade flows. *Empirical Economics*, *1*(1): 1–39.

Mehmood, R., & Graham, G. (2015). Big Data Logistics: A health-care Transport Capacity Sharing Model. *Procedia Computer Science*, *64*: 1107–1114.

Meinert, R. (2019). Towards Data Science - Optimizing Hyperparameters in Random Forest Classification. Retrieved from https://towardsdatascience.com/optimizing-hyperparameters-in-random-forest-classification-ec7741f9d3f6 (date accessed: 20/07/2019).

Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L., & Canhoto, A. (2018). Big data, big decisions: The impact of big data on board level decision-making. *Journal of Business Research*, *93*: 67–78.

Miller, A. (1997). *Strategic Management: 3rd (Third) edition*. McGraw-Hill Companies.

Mishra, D., Gunasekaran, A., Papadopoulos, T., & Childe, S. J. (2018). Big data and supply chain management: A review and bibliometric analysis. *Annals of Operations Research*, *270*(1–2): 313–336.

Mishra, S., Mishra, D., & Santra, G. H. (2016). Applications of machine learning techniques in agricultural crop production: A review paper. *Indian Journal of Science and Technology*, *9*(38): 1–14.

Mutandwa, E., Kanuma, N. T., Rusatira, E., Mugenzi, P., Govere, I., & Foti, R. (2009). Analysis of coffee export marketing in Rwanda : Application of the Boston consulting group matrix. *African Journal of Business Management*, *2*(4): 210–219.

Nguyen, T., Li, Z., Spiegler, V., Ieromonachou, P., & Lin, Y. (2018). Big data analytics in supply chain management: A state-of-the-art literature review. *Computers & Operations Research*, *98*: 254–264.

Nummelin, T., & Hänninen, R. (2016). *Model for international trade of sawnwood using machine learning models*. *Natural Resources and Bioeconomy Studies* (Vol. 74). Retrieved from http://jukuri.luke.fi/handle/10024/537749 (date accessed: 20/07/2019).

OECD. (2019a). Developments in individual OECD and selected non-member economies. *OECD Economic Outlook*, *2018*(2): 89–91.

OECD. (2019b). OECD.Stat. Retrieved from https://stats.oecd.org/ (date accessed: 20/07/2019).

Ozemre, M., & Kabadurmus, O. (2019). Using Big Data Analytics to Forecast Trade Volumes in Global Supply Chain Management. In *Managing Operations Throughout Global Supply Chains* (pp. 70–99). IGI Global.

Padhi, S. S., Wagner, S. M., & Aggarwal, V. (2012). Positioning of commodities using the Kraljic Portfolio Matrix. *Journal of Purchasing and Supply Management*, *18*(1): 1–8.

Pakravan, M. R., Kelashemi, M. K., & Alipour, H. R. (2011). Forecasting Iran's rice imports trend during 2009-2013. *International Journal of Agricultural Management and Development (IJAMAD)*, *1*: 39–44.

Palmer, A., Montaño, J. J., & Sesé, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management*, *27*(5): 781–790.

Piezas-Jerbi, N., & Wardyn, S. (2017). *Trends in world trade: Introduction*. *World Trade Statistical Review*. Retrieved from https://www.wto.org/english/res_e/statis_e/wts2018_e/wts2018chapter01_e.pdf (date accessed: 20/07/2019).

Raisinghani, M. S., & Meade, L. L. (2005). Strategic decisions in supply-chain intelligence using knowledge management: an analytic-network-process framework. *Supply Chain Management: An International Journal*, *10*(2): 114–121.

Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Informatiın Storage amd OrganizationIn the Brain. *Psychological Review*, *65*(6): 386–408.

Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter*, *19*(4), 1–34. Retrieved from https://tdwi.org/research/2011/09/~/media/TDWI/TDWI/Research/BPR/2011/TDWI_BPReport_Q411_Big_Data_Analytics_Web/TDWI_BPReport_Q411_Big Data_ExecSummary.ashx (date accessed: 20/07/2019).

Sahu, P. K., & Mishra, P. (2013). Modelling and forecasting production behaviour and import- export of total spices in two most populous countries of the World. *Journal of Agricultural Research*, *51*(1): 81–97.

Sanders, N. R. (2014). *Big data driven supply chain management: A framework for implementing analytics and turning information into intelligence*. New Jersey: Pearson Education.

SAS. (2019). Which machine learning algorithm should I use? Retrieved from https://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/ (date accessed: 20/07/2019).

Schoenherr, T., & Speier-Pero, C. (2015). Data science, predictive analytics, and big data in supply chain management: Current state and future potential. *Journal of Business Logistics*, *36*(1): 120–132.

Scikit_Learn. (2019a). Multi Layer Perceptron. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html (date accessed: 20/07/2019).

Scikit_Learn. (2019b). Random Forest Regressor. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html (date accessed: 20/07/2019).

Senhadji, A. S., & Montenegro, C. E. (1999). Time series analysis of export demand equations : A cross - country analysis. *IMF Staff Papers*, *Vol 46*, *No*(3): 259–273.

Sepp, H., & Jurgen, S. (1997). 1997Hochreiter_LSTM. *Neural Computation*, *9*(8): 1–32.

Shaw, E. H. (2012). Marketing strategy: From the origin of the concept to the development of a conceptual framework. *Journal of Historical Research in Marketing*, *4*(1): 30–55.

Shibasaki, R., & Watanabe, T. (2012). Future forecast of trade amount and international cargo flow in the APEC Region: An application of Trade-Logistics Forecasting Model. *Asian Transport Studies*, *2*(2): 194–208.

Silva, E. S., & Hassani, H. (2015). On the use of singular spectrum analysis for forecasting U.S. trade before, during and after the 2008 recession. *International Economics*, *141*: 34–49.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*: 263–286.

Smith, M. (2002). Derrick's Ice-Cream Company: applying the BCG matrix in customer pro analysis. *Accounting Education*, *11*(4): 365–375.

Sokolov-Mladenović, S., Milovančević, M., Mladenović, I., & Alizamir, M. (2016). Economic growth forecasting by artificial neural network with extreme learning machine based on trade, import and export parameters. *Computers in Human Behavior*, *65*: 43–45.

Speranza, M. G. (2018). Trends in transportation and logistics. *European Journal of*

Operational Research, *264*(3): 830–836.

Stahlbock, R., & Voß, S. (2010). Improving empty container logistics – Can it avoid a collapse in container transportation? In L. Kroon, T. Li, & R. Zuidwijk (Eds.), *Liber Amicorum In Memoriam Jo Van Nunen* (pp. 217–224). Rotterdam: Rotterdam School of Management, Erasmus University.

Stefanovic, N., & Stefanovic, D. (2009). Supply chain business intelligence: technologies, issues and trends. In M. Bramer (Ed.), *Artificial intelligence an international perspective* (pp. 217–245). Berlin: Springer Berlin Heidelberg.

Svatos, M., Smutka, L., & Miffek, O. (2010). Competitiveness of agrarian trade of EU-15 countries in comparison with new EU member states. *Agricultural Economics-Zemedelska Ekonomika*, *56*(12): 569–582.

Tanwar, R. (2013). Porter's generic competitive strategies. *Journal of Business and Management*, *15*(1): 11–17.

The_Asimov_Institute. (2019). - The Neural Network Zoo. Retrieved from http://www.asimovinstitute.org/neural-network-zoo/ (date accessed: 20/07/2019).

Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing Journal*, *38*: 788–804.

Veenstra, A. W., & Haralambides, H. E. (2001). Multivariate autoregressive models for forecasting seaborne trade flows. *Transportation Research Part E: Logistics and Transportation Review*, *37*(4): 311–319.

Villars, R. L., Olofson, C. W., & Eastwood, M. (2011). *Big data: What it is and why you should care*. *IDC Analyze the Future*. Retrieved from http://www.admin-magazine.com/HPC/content/download/5604/49345/file/IDC_Big%2520Data_whitepaper_final.pdf (date accessed: 20/07/2019).

Wagacha, P. W. (2003). *Induction of decision trees*. Nairobi. Retrieved from http://41.204.161.209/bitstream/handle/11295/44263/decisionTrees.pdf?sequence=1&isAllowed=y (date accessed: 20/07/2019).

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. fan, Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*: 356–365.

Wang, C., Li, X., Zhou, X., Wang, A., & Nedjah, N. (2016). Soft computing in big data intelligent transportation systems. *Applied Soft Computing*, *38*, 1099–1108.

Weiss, E. A. (1992). Arthur Lee Samuel. *IEEE Annals of the History of Computing*, *14*(3): 55–69.

Wigan, M. R., & Clarke, R. (2013). Big data's big unintended consequences. *Computer*, *46*(6): 46–53.

Wind, Y. . (1988). *Positioning analysis and strategy*. *The Interface of Marketing and Strategy*. Wharton School, University of Pennsylvania, Marketing Department.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *The Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining* (pp. 29–39).

World Trade Organization. (2018). World Trade Statistical Review 2018 About the WTO. Retrieved from www.wto.org/statistics (date accessed: 20/07/2019).

Wu, X., Zhu, X., Wu, G.-Q., & Wei, D. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1): 97–107.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*(1): 35–62.

APPENDICES

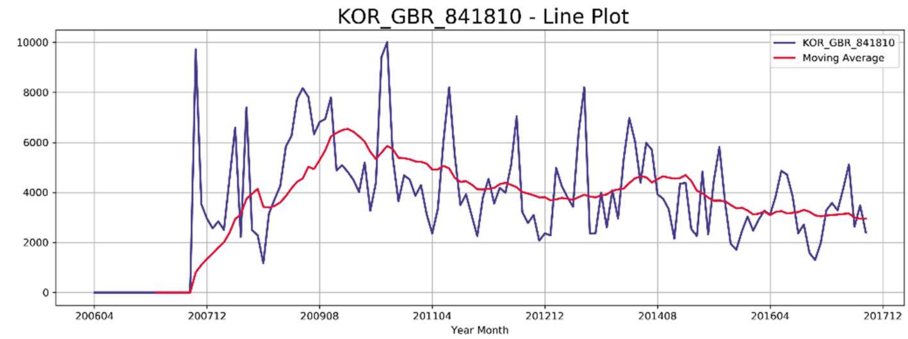
APPENDIX 1 - Line Plots Product Specific Features

In this appendix top five export partners of the United Kingdom, the top five import partners of Turkey and the top five import partners of China for all product groups are given. Units of these features are given in Chapter 7.1.

Product 84.18.10 Line Plots

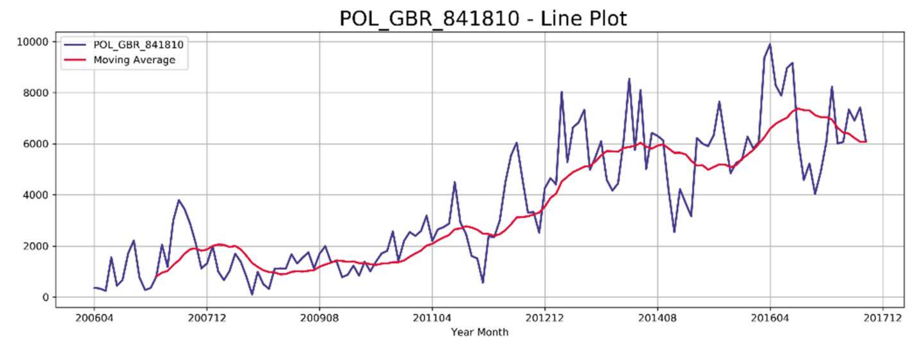
The United Kingdom's top five export partners for Product 84.18.10 are given below. These are common data for Turkish and Chinese company cases.

Figure A1.1 Export Volumes of the Product 84.18.10 from Korea to the United Kingdom



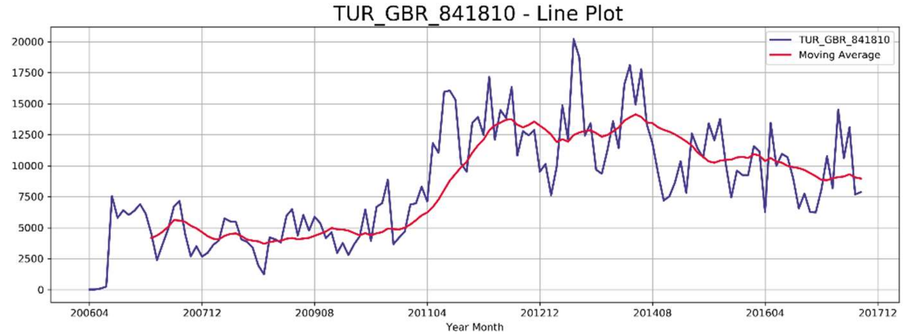
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.2 Export Volumes of the Product 84.18.10 from Poland to the United Kingdom



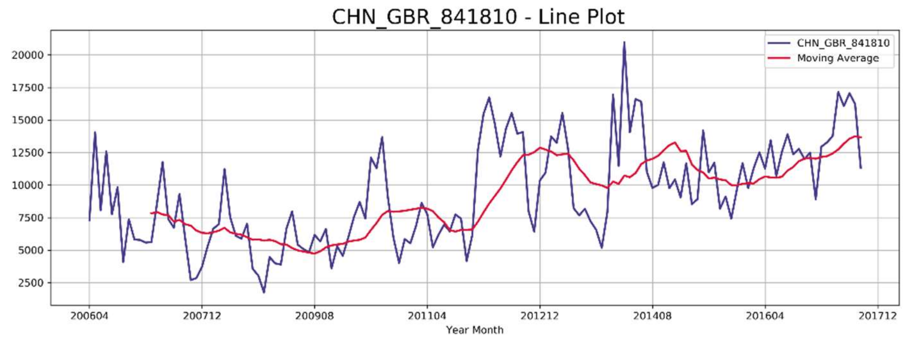
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.3 Export Volumes of the Product 84.18.10 from Turkey to the UK



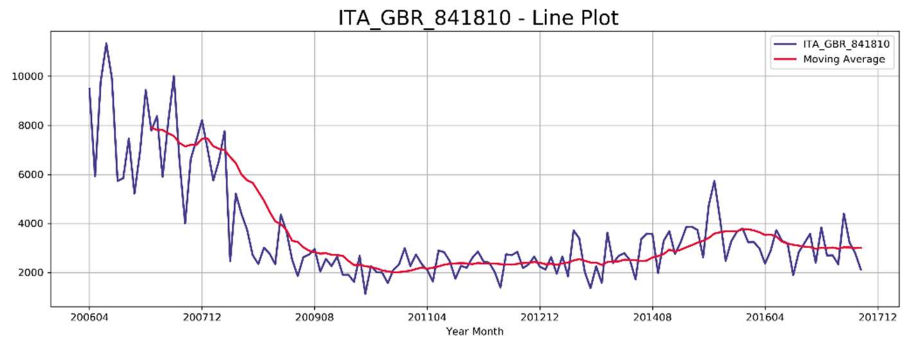
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.4 Export Volumes of the Product 84.18.10 from China to the UK



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.5 Export Volumes of the Product 84.18.10 from Italy to the United Kingdom

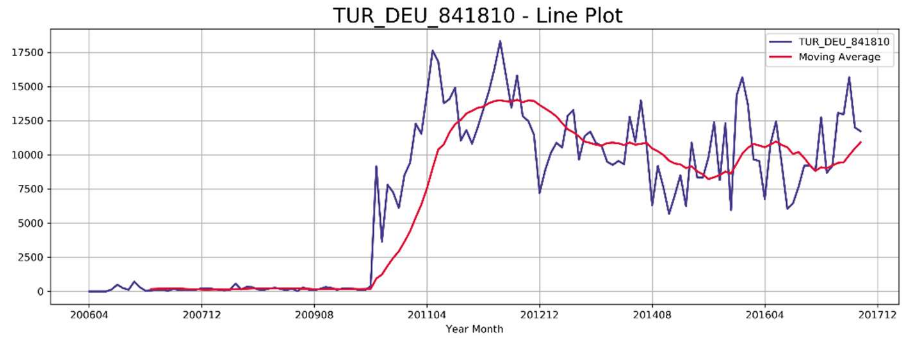


Source: Data retrieved from: International Trade Center (2019). Trade Map.

Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

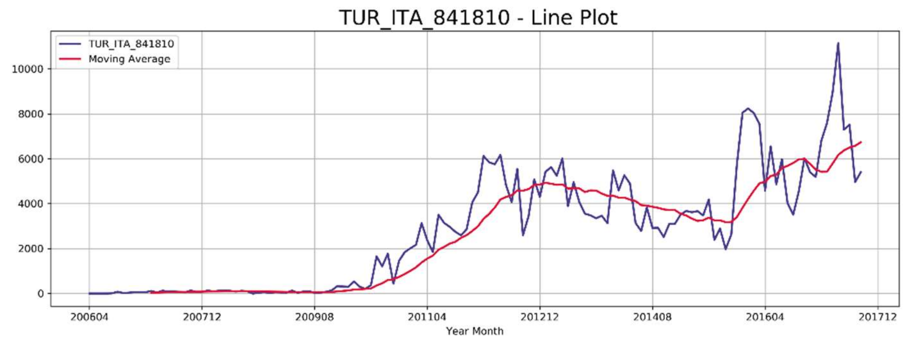
Turkey's top five import partners for Product 84.18.10 plots are given below. These are used for a Turkish company case.

Figure A1.6 Export Volumes of the Product 84.18.10 from Turkey to Germany



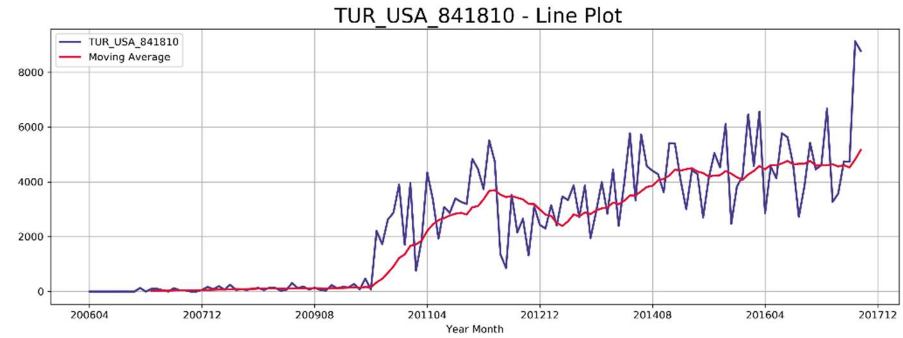
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.7 Export Volumes of the Product 84.18.10 from Turkey to Italy



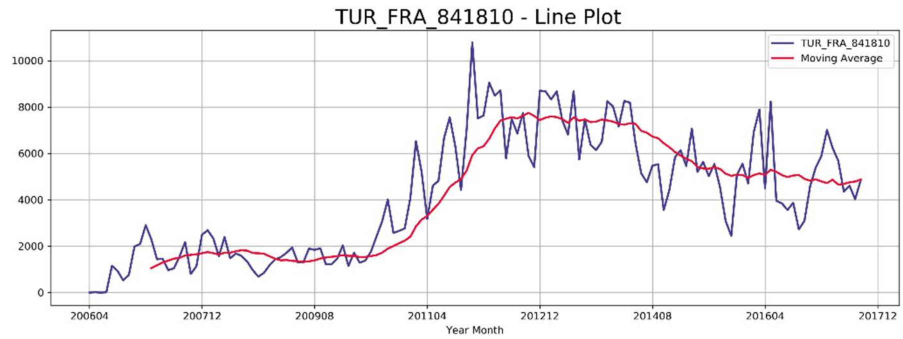
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.8 Export Volumes of the Product 84.18.10 from Turkey to the United States



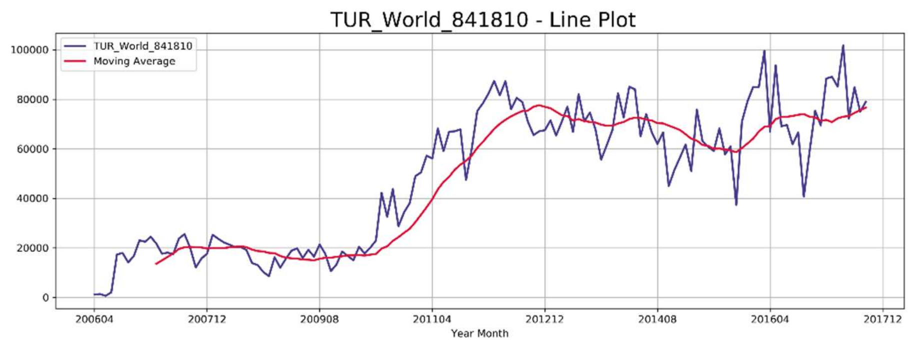
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.9 Export Volumes of the Product 84.18.10 from Turkey to France



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.10 Export Volumes of the Product 84.18.10 from Turkey to World

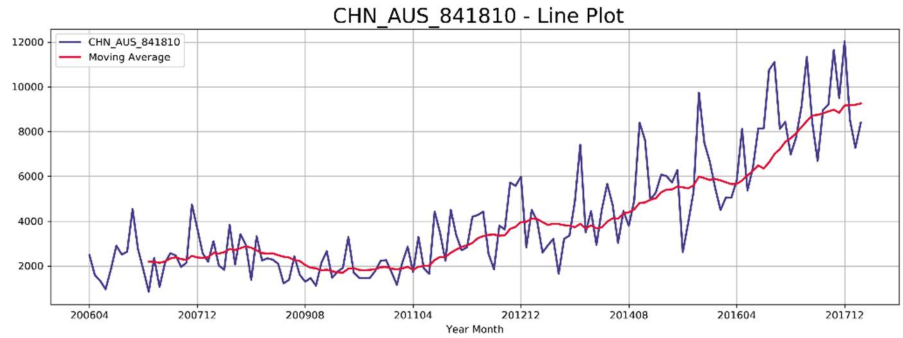


Source: Data retrieved from: International Trade Center (2019). Trade Map.

Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

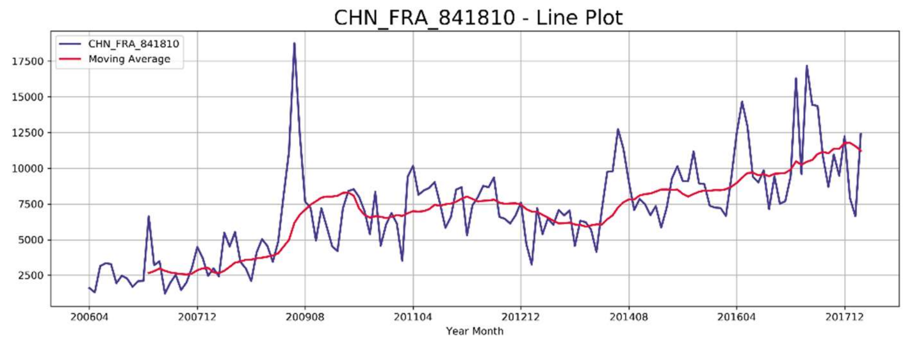
China's top five import partners for Product 84.18.10 plots are given below. These are used for the Chinese company case.

Figure A1.11 Export Volumes of the Product 84.18.10 from China to Austria



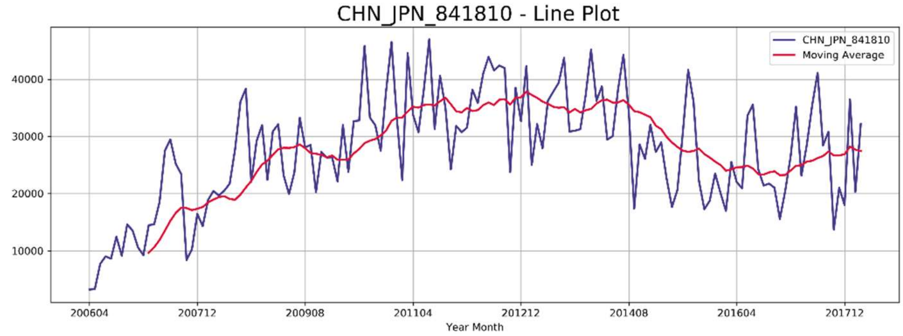
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.12 Export Volumes of the Product 84.18.10 from China to France



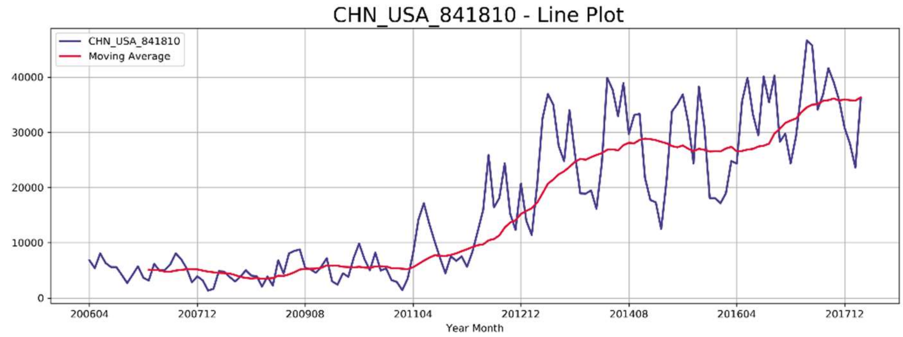
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.13 Export Volumes of the Product 84.18.10 from China to Japan



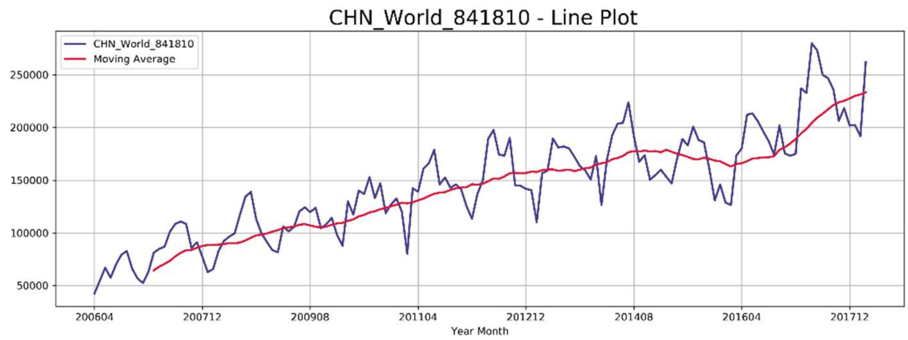
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.14 Export Volumes of the Product 84.18.10 from China to United States



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.15 Export Volumes of the Product 84.18.10 from China to World

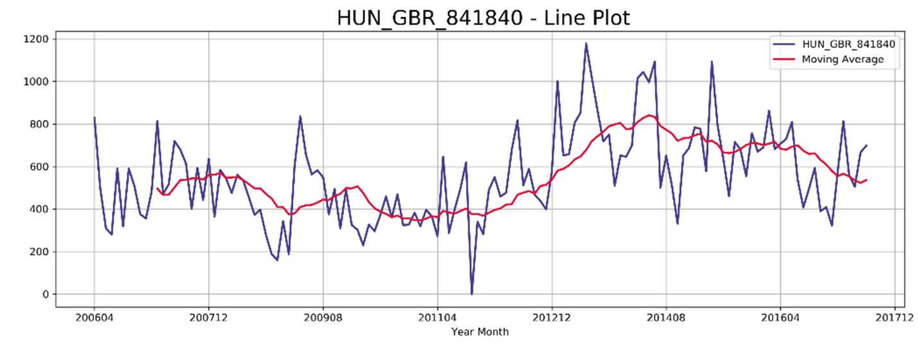


Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Product 84.18.40 Line Plots

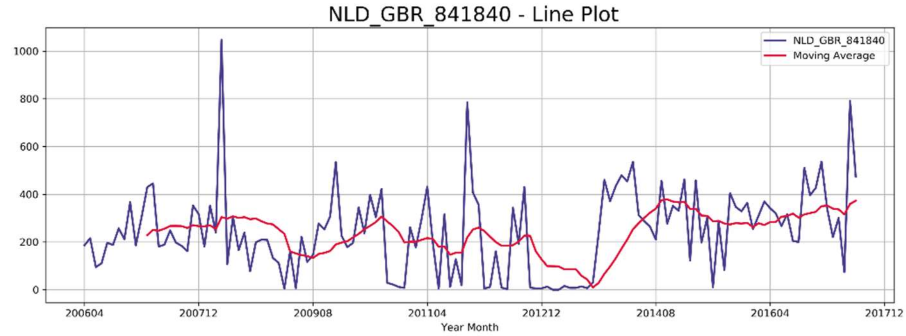
The United Kingdom's top five export partners for Product 84.18.40 are given below. These are common data for Turkish and Chinese company cases.

Figure A1.16 Export Volumes of the Product 84.18.40 from Hungary to the United Kingdom



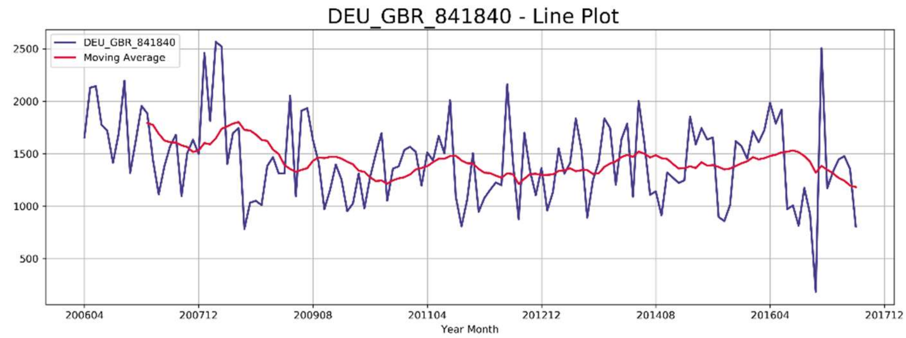
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.17 Export Volumes of the Product 84.18.40 from Netherland to the United Kingdom



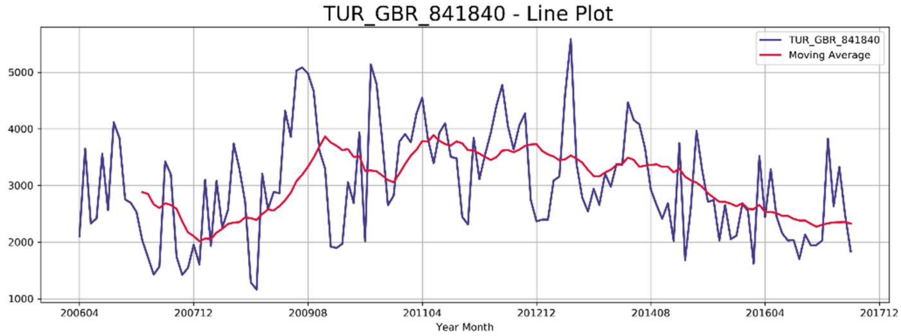
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.18 Export Volumes of the Product 84.18.40 from Germany to the United Kingdom



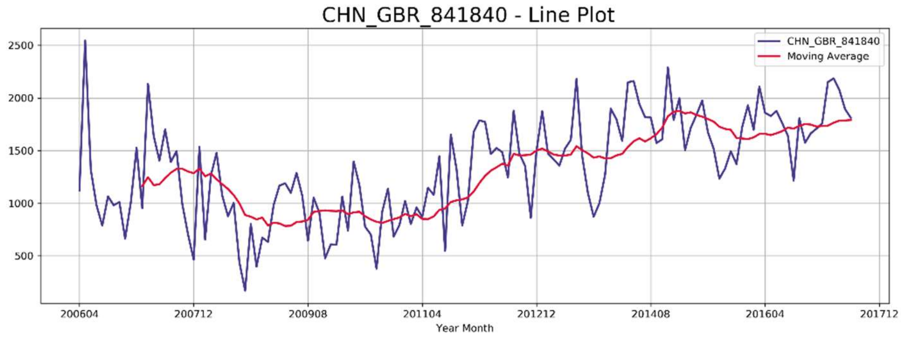
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.19 Export Volumes of the Product 84.18.40 from Turkey to the UK



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

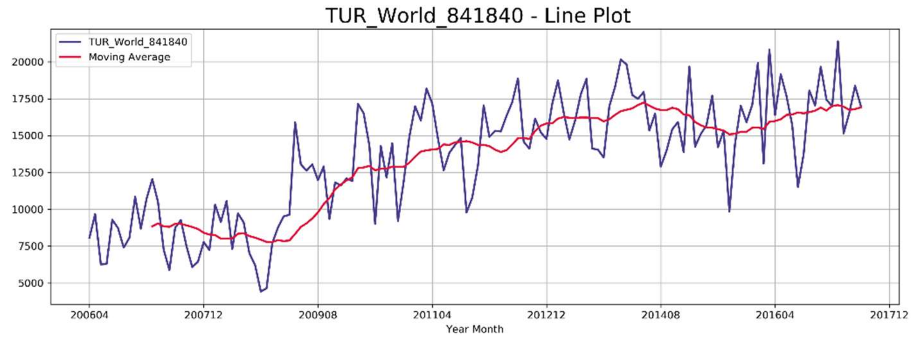
Figure A1.20 Export Volumes of the Product 84.18.40 from China to the UK



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

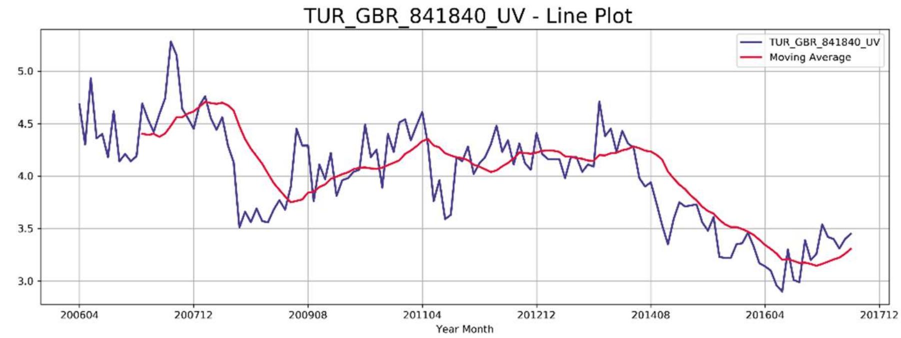
Turkey's top five import partners for Product 84.18.40 plots are given below. These are used for the Turkish company case.

Figure A1.21 Export Volumes of the Product 84.18.40 from Turkey to the UK



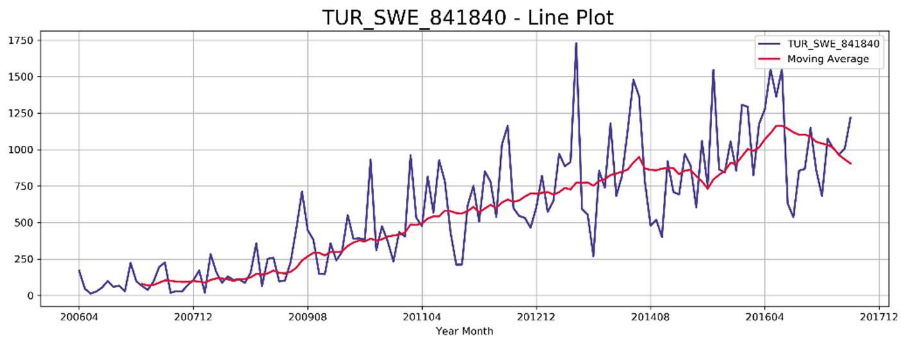
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.22 Export Unit Values of the product 84.18.40 from Turkey to the UK



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

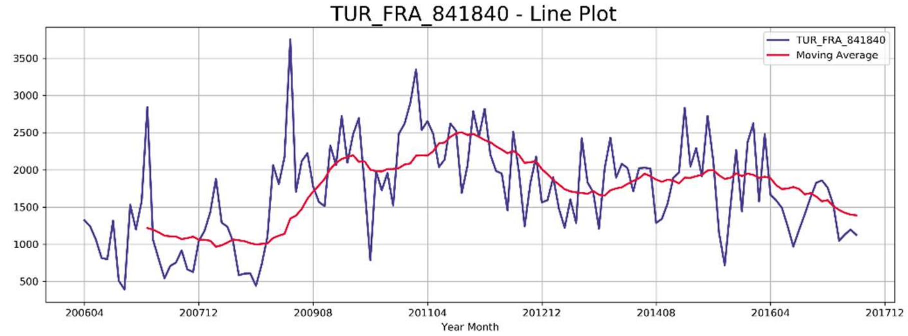
Figure A1.23 Export Volumes of the Product 84.18.40 from Turkey to Sweden



Source: Data retrieved from: International Trade Center (2019). Trade Map.

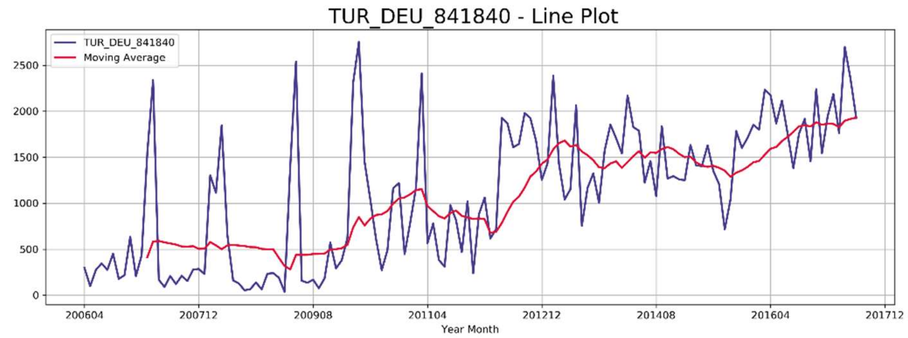
Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.24 Export Volumes of the Product 84.18.40 from Turkey to France



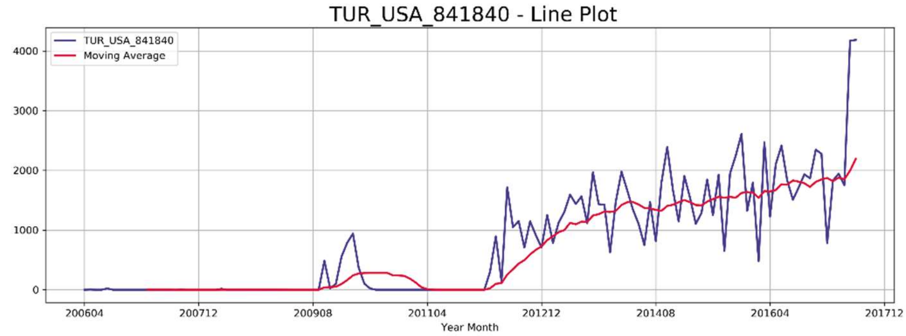
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.25 Export Volumes of the Product 84.18.40 from Turkey to Germany



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

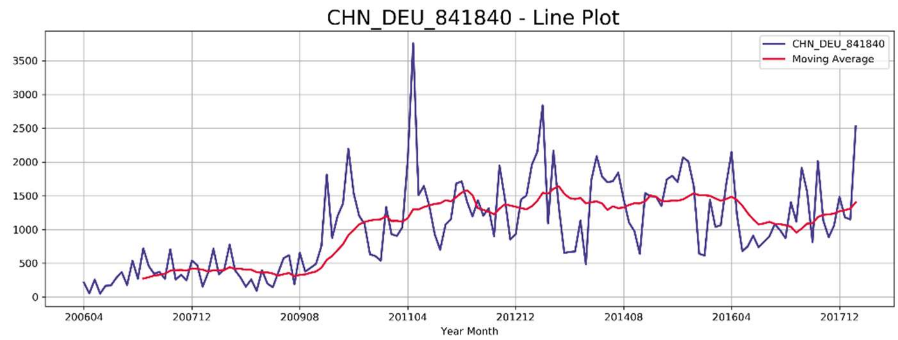
Figure A1.26 Export Volumes of the Product 84.18.40 from Turkey to the United States



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

China's top five import partners for Product 84.18.40 plots are given below. These are used for the Chinese company case.

Figure A1.27 Export Volumes of the Product 84.18.40 from China to Germany

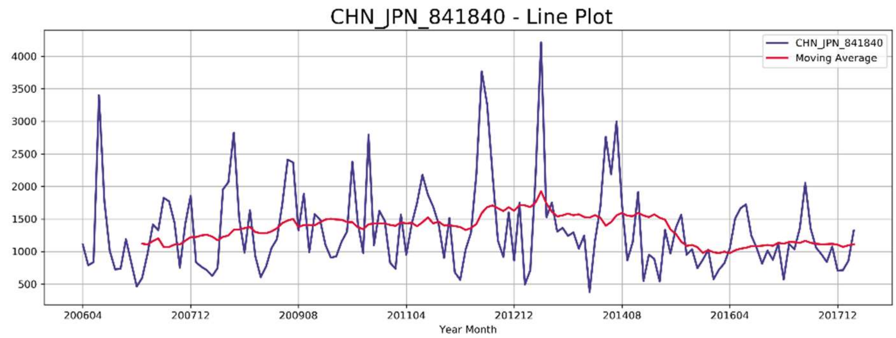


Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.28 Export Volumes of the Product 84.18.40 from China to France

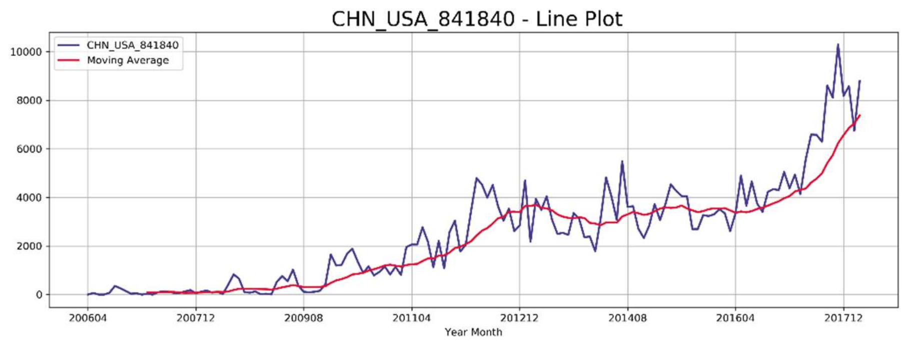
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.29 Export Volumes of the Product 84.18.40 from China to Japan



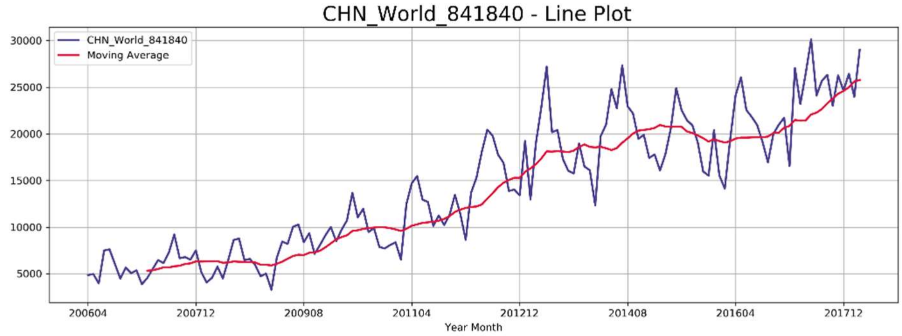
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.30 Export Volumes of the Product 84.18.40 from China to the United States



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.31 Export Volumes of the Product 84.18.40 from China to World

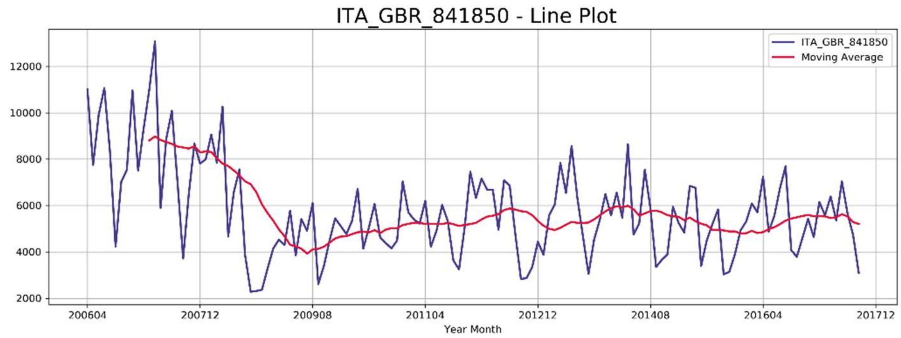


Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Product 84.18.50 Line Plots

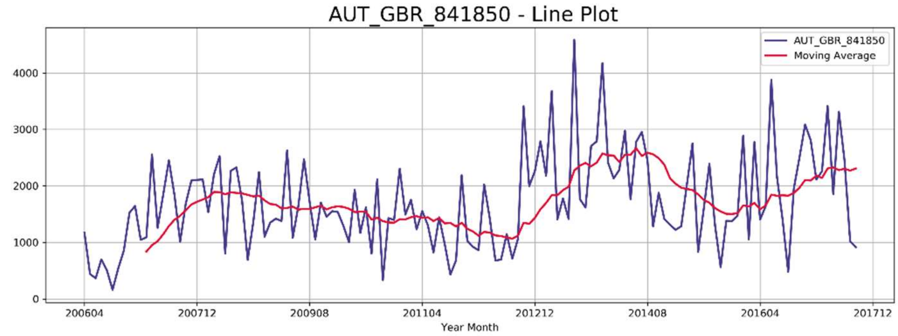
The United Kingdom's top five export partners for Product 84.18.50 are given below. These are common data for Turkish and Chinese company cases.

Figure A1.32. Export Volumes of the Product 84.18.50 from Italy to the United Kingdom



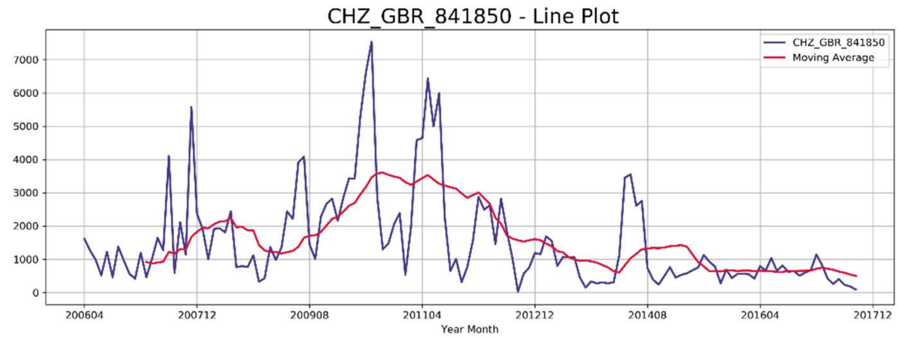
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.33. Export Volumes of the Product 84.18.50 from Austria to the United Kingdom



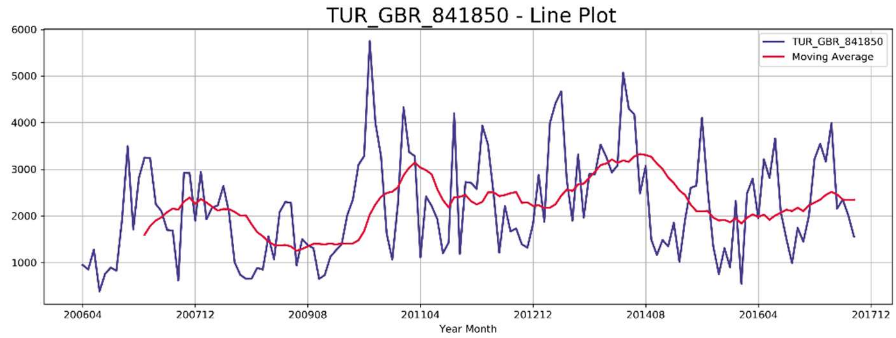
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.34. Export Volumes of the Product 84.18.50 from Chezia to the United Kingdom



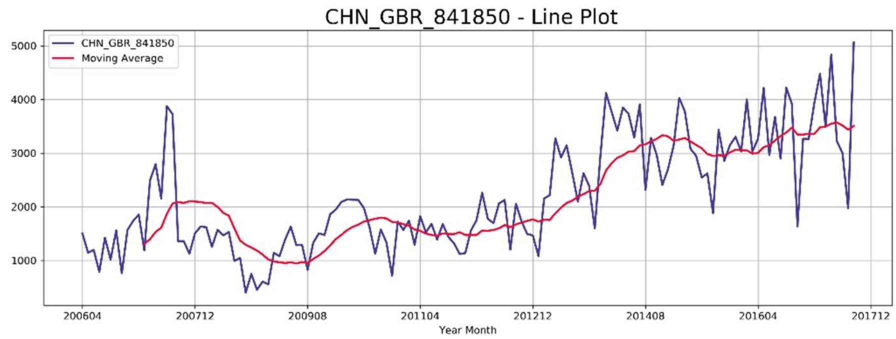
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.35. Export Volumes of the Product 84.18.50 from Turkey to the UK



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

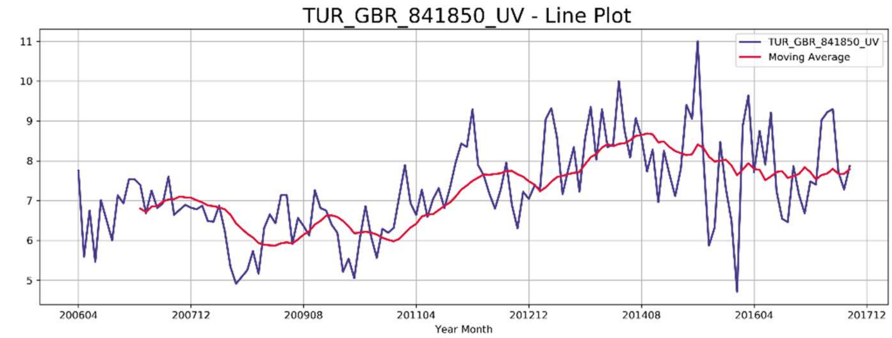
Figure A1.36. Export Volumes of the Product 84.18.50 from China to the UK



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

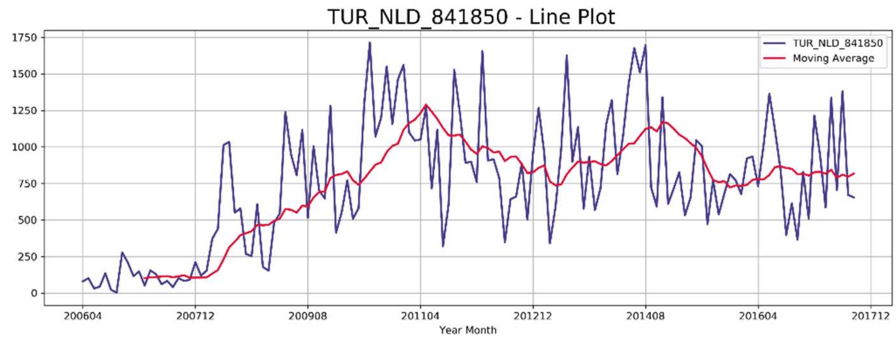
Turkey's top five import partners for Product 84.18.50 plots are given below. These are used for the Turkish company case.

Figure A1.37 Export Unit Values of the Product 84.18.50 from Turkey to the UK



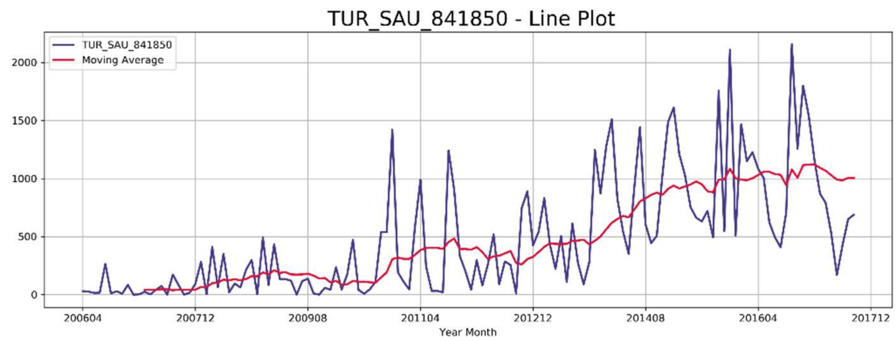
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.38 Export Volumes of the Product 84.18.50 from Turkey to Netherland



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

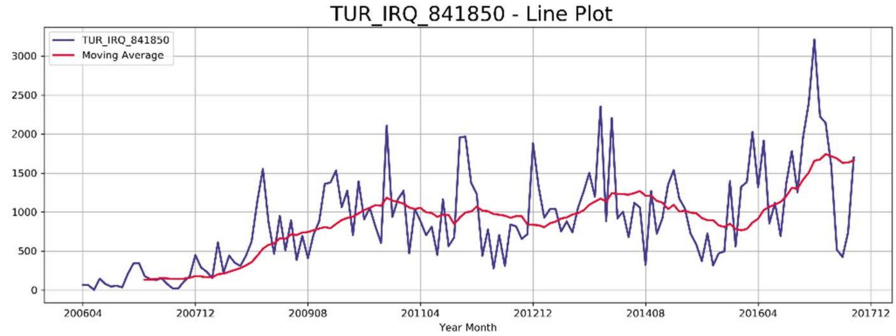
Figure A1.39 Export Volumes of the Product 84.18.50 from Turkey to Saudia Arabia



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date

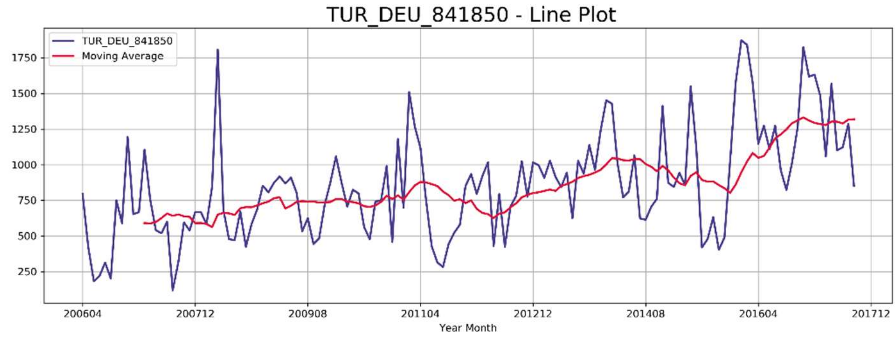
accessed: 20/07/2019).

Figure A1.40 Export Volumes of the Product 84.18.50 from Turkey to Iraq



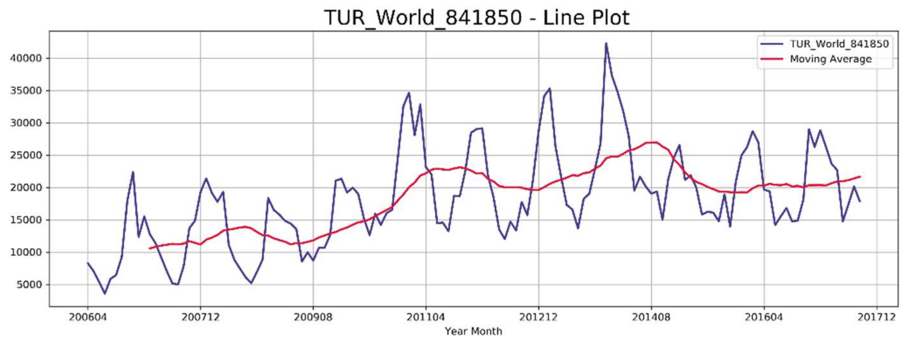
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.41 Export Volumes of the Product 84.18.50 from Turkey to Germany



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.42 Export Volumes of the Product 84.18.50 from Turkey to World

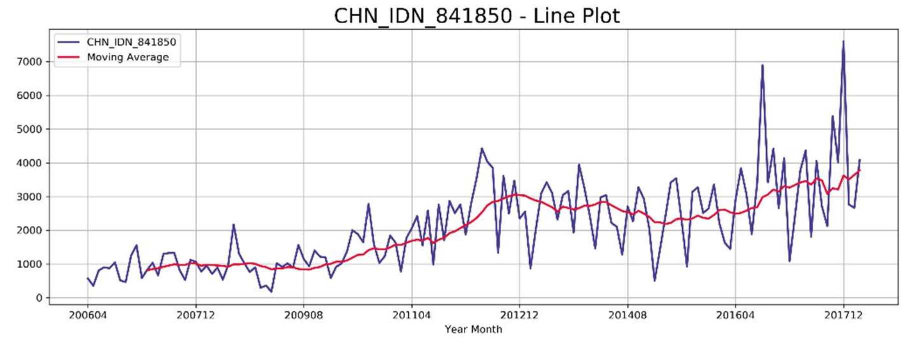


Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date

accessed: 20/07/2019).

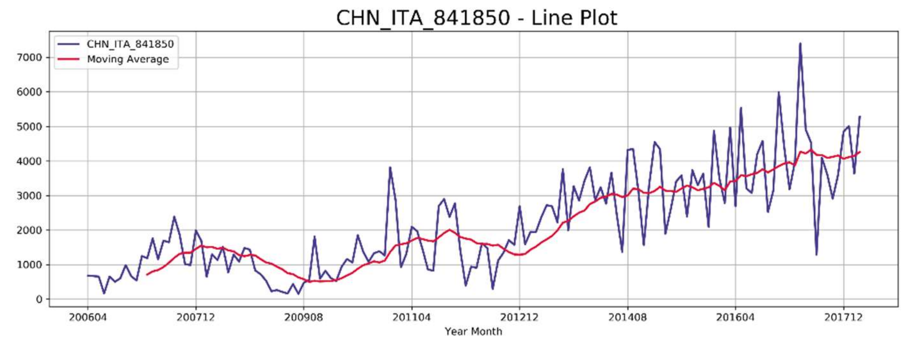
China's top five import partners for Product 84.18.50 plots are given below. These are used for the Chinese company case.

Figure A1.43 Export Volumes of the Product 84.18.50 from China to Indonesia



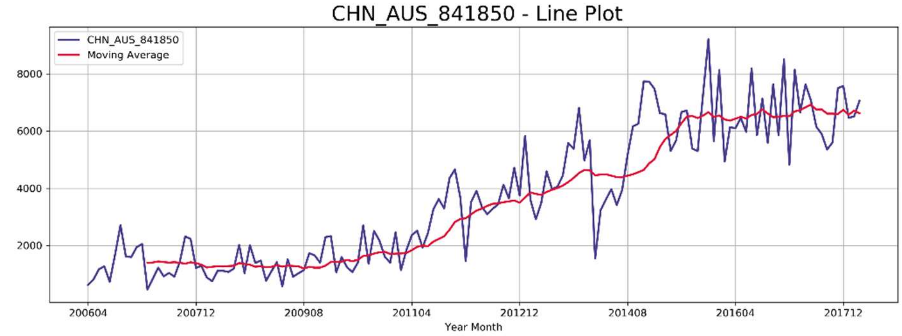
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.44 Export Volumes of the Product 84.18.50 from China to Italy



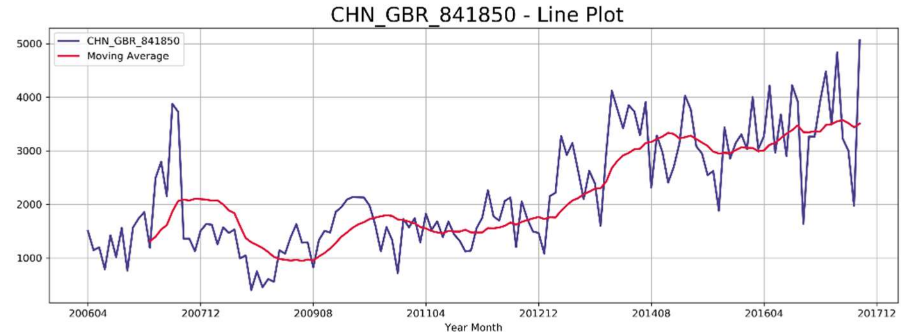
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.45 Export Volumes of the Product 84.18.50 from China to Austria



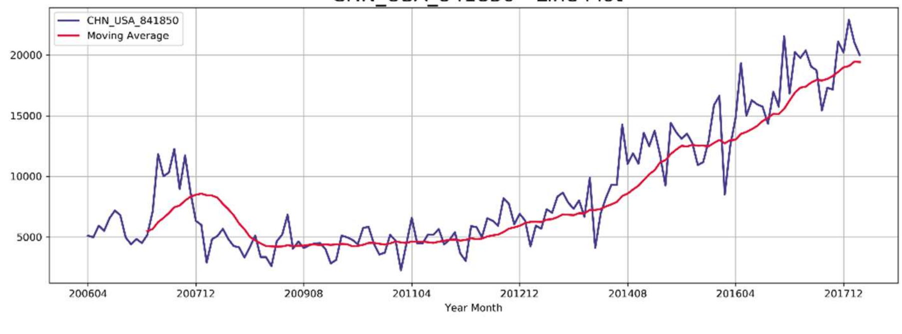
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A1.46 Export Volumes of the Product 84.18.50 from China to the UK



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

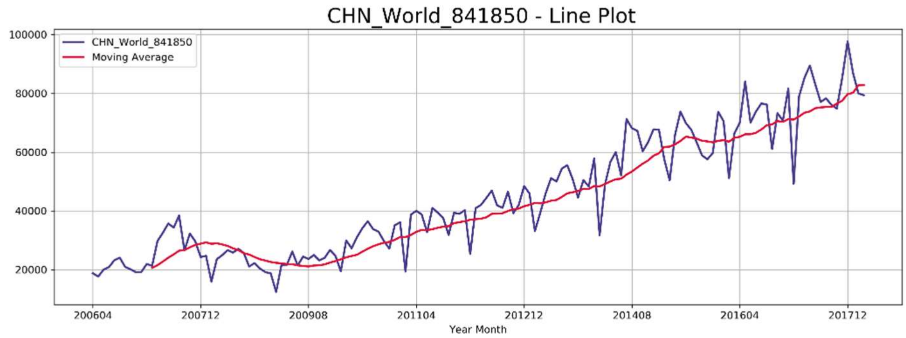
Figure A1.47 Export Volumes of the Product 84.18.50 from China to the United States



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date

accessed: 20/07/2019).

Figure A1.48 Export Volumes of the Product 84.18.50 from China to World



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

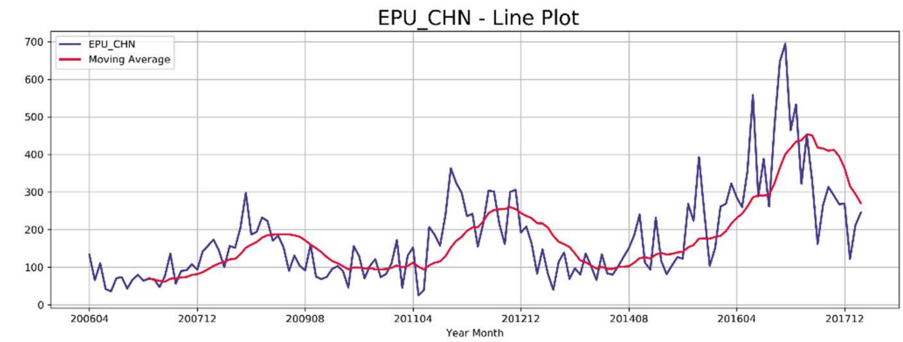
APPENDIX 2 - Line Plots of Global Features

In this appendix, the features that are not related to the products are given. These are Political Environment features, Economic Environment Features, and Business Environment Features. Units of these features are given in Chapter 3.1.2.

POLITICAL ENVIRONMENT: EPU

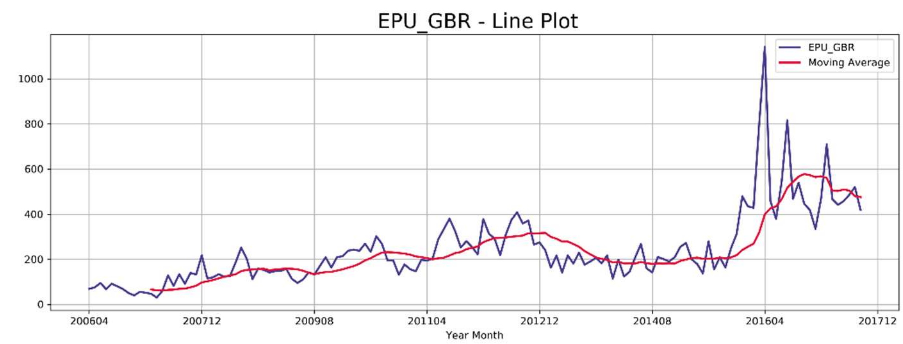
This chapter presents Economic Politic Uncertainty index graphs for China, the United Kingdom and for the World.

Figure A2.1. Economic Politic Uncertainty Index for China



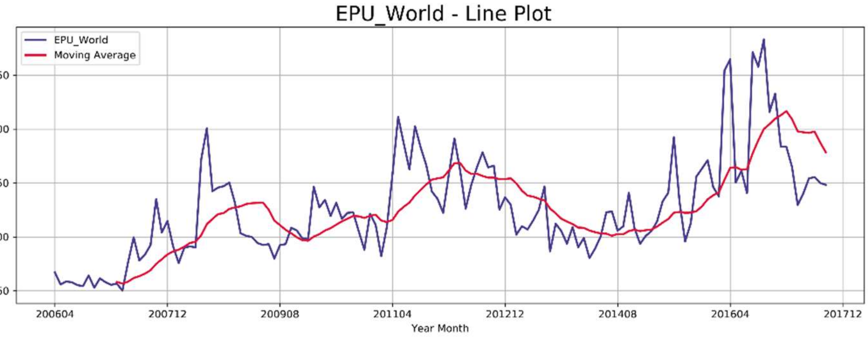
Source: Data retrieved from: EPU. (2019). Economic Policy Uncertainty. Retrieved from https://www.policyuncertainty.com/ (date accessed: 20/07/2019).

Figure A2.2. Economic Politic Uncertainty Index for United Kingdom



Source: Data retrieved from: EPU. (2019). Economic Policy Uncertainty. Retrieved from https://www.policyuncertainty.com/ (date accessed: 20/07/2019).

Figure A2.3. Economic Politic Uncertainty Index for World

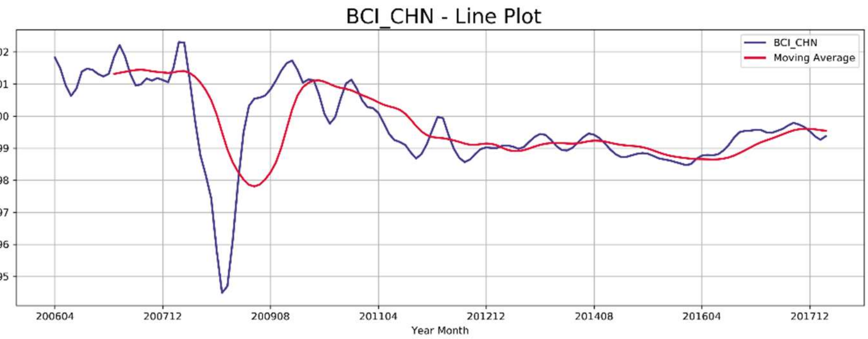


Source: Data retrieved from: EPU. (2019). Economic Policy Uncertainty. Retrieved from https://www.policyuncertainty.com/ (date accessed: 20/07/2019).

BUSINESS ENVIRONMENT: BCI, CCI, CLI

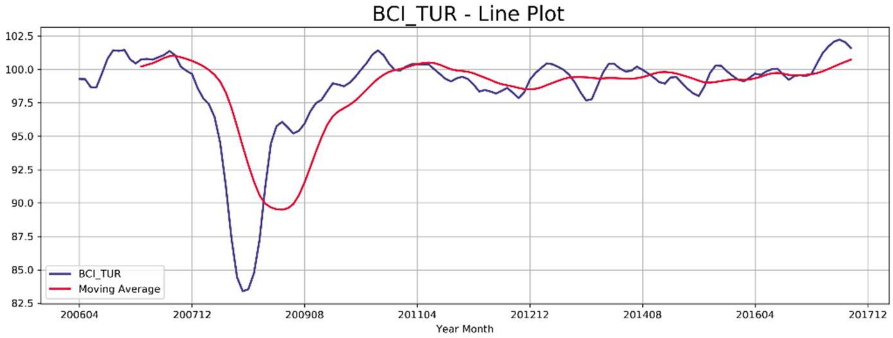
This chapter presents Business Confidence Indicator, Customer Confidence Indicator, Composite Leading Indicator graphs for Turkey, China and the UK

Figure A2.4. Business Confidence Indicator for China



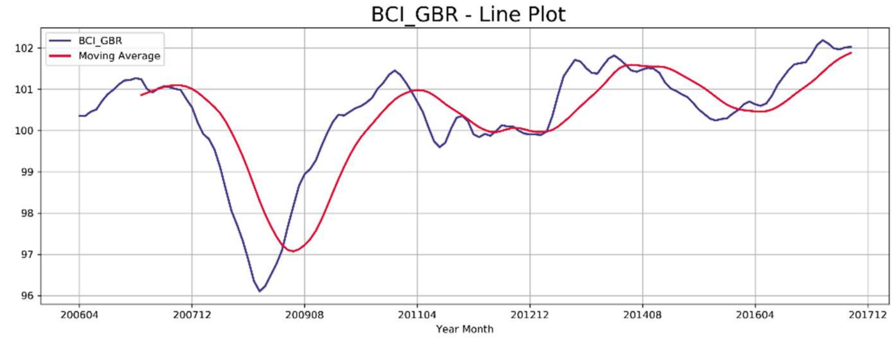
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.5. Business Confidence Indicator for Turkey



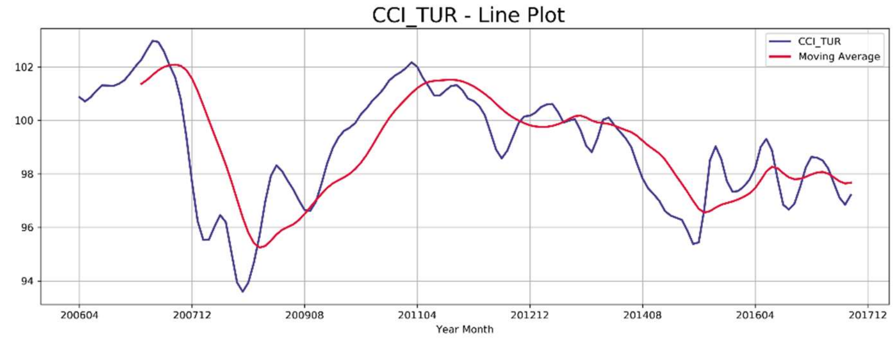
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.6. Business Confidence Indicator for United Kingdom



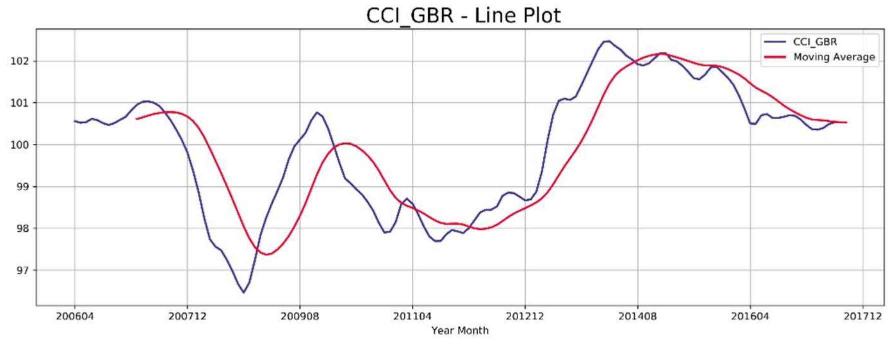
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.7. Customer Confidence Indicator for Turkey



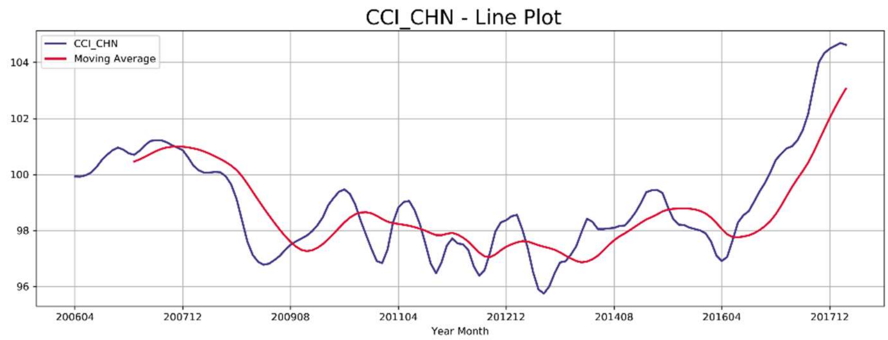
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.8. Customer Confidence Indicator for United Kingdom



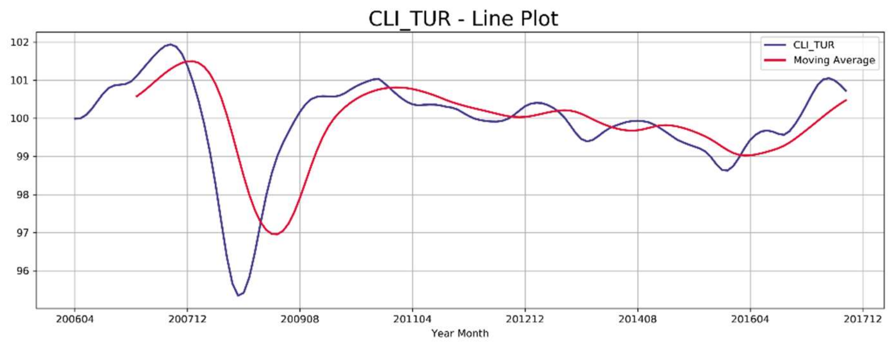
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.9. Customer Confidence Indicator for China



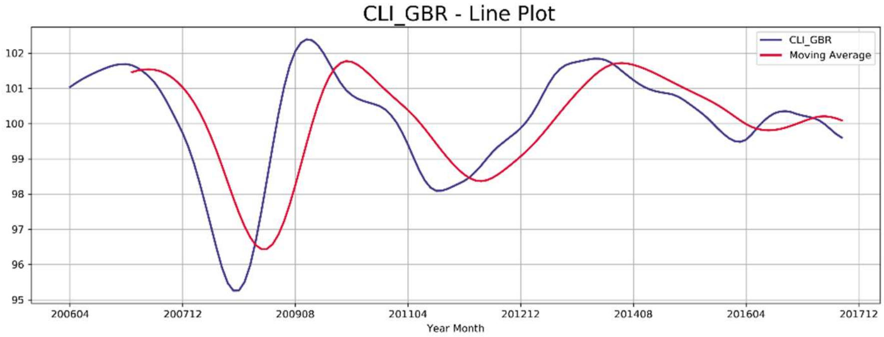
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.10. Composite Leading Indicator for Turkey



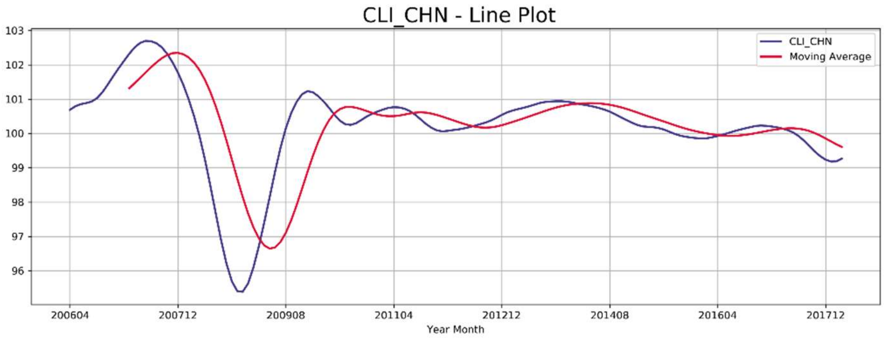
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.11. Composite Leading Indicator for United Kingdom



Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.12. Composite Leading Indicator for China

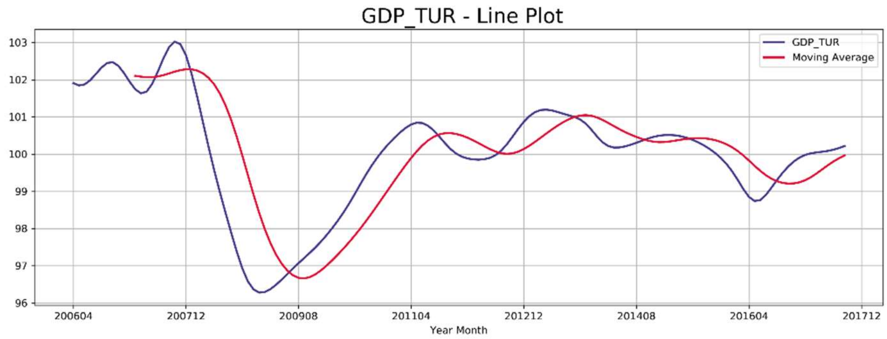


Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

ECONOMIC ENVIRONMENT: GDP, PPI, CPI, EXCHANGE RATES

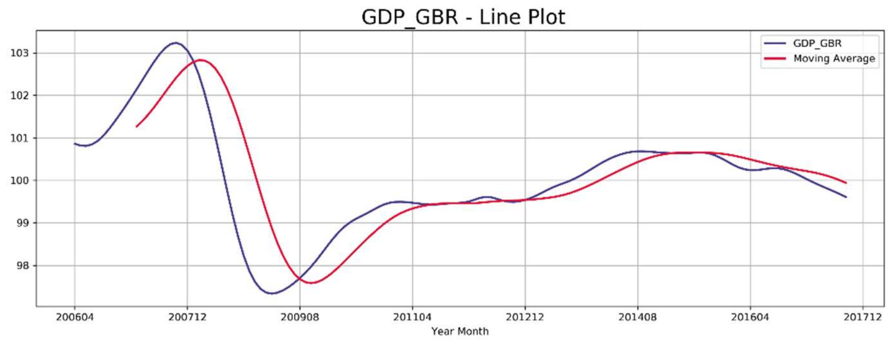
This chapter presents GDP, Producer Price Indices, Consumer Price Indices, Currency exchange rate graphs for Turkey, China and the United Kingdom. World trade volume graph is given in the last part.

Figure A2.13. Gross Domestic Product for Turkey



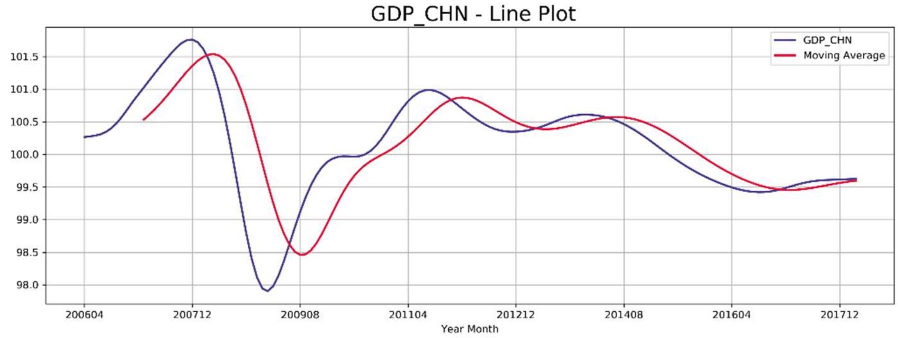
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.14. Gross Domestic Product for United Kingdom



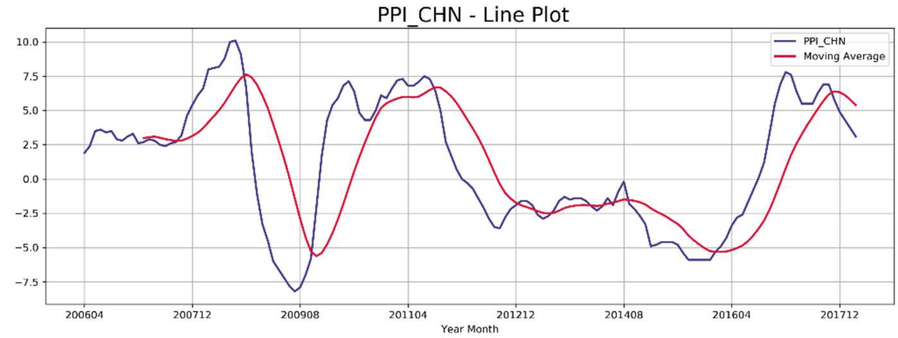
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.15. Gross Domestic Product for China



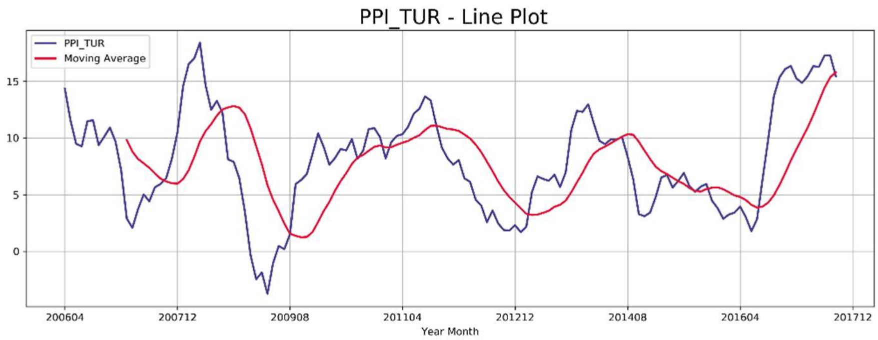
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.16. Producer Price Indices for China



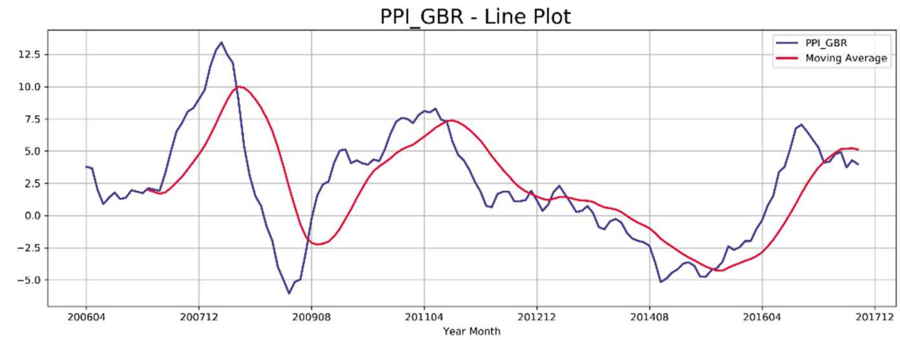
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.17. Producer Price Indices for Turkey



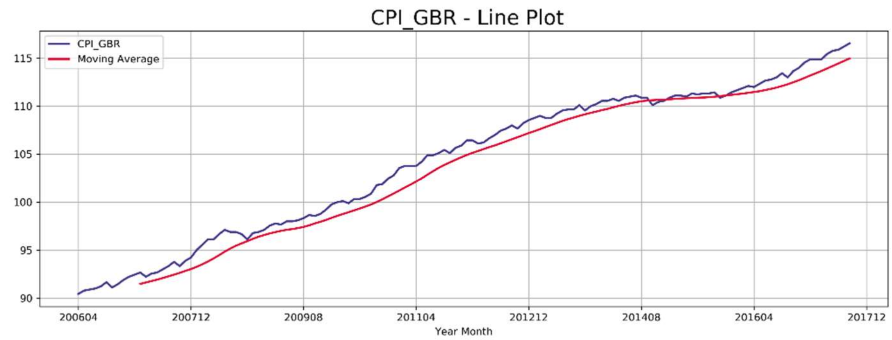
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.18. Producer Price Indices for United Kingdom



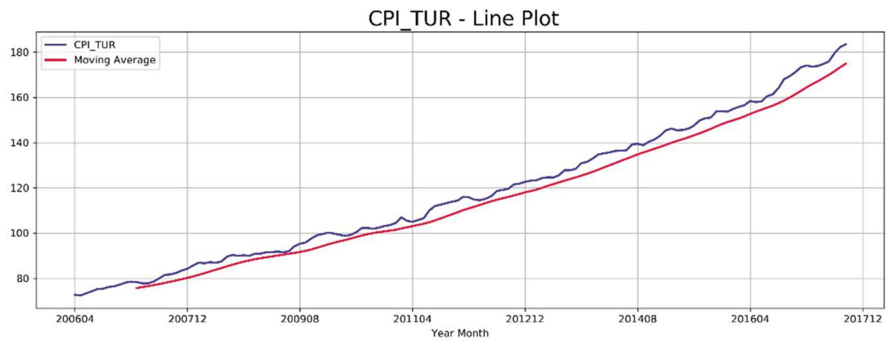
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.19. Consumer Price Indices for United Kingdom



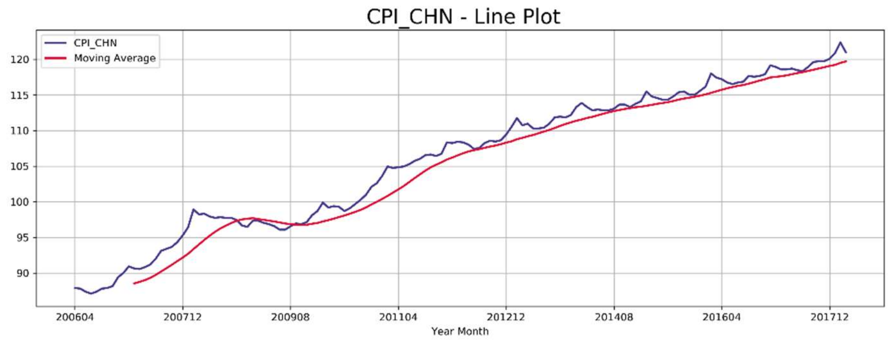
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.20. Consumer Price Indices for Turkey



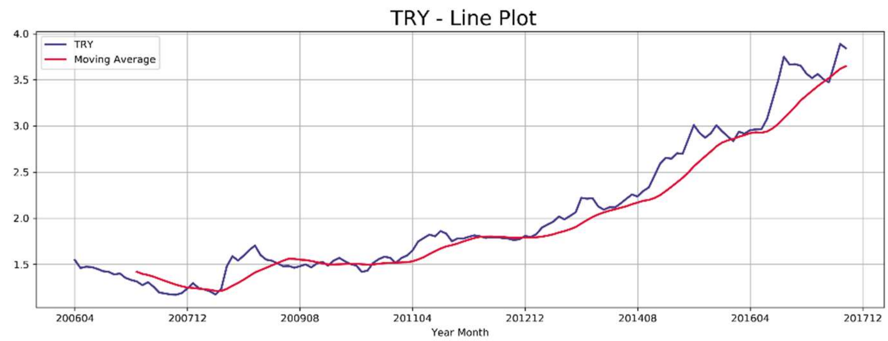
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.21. Consumer Price Indices for China



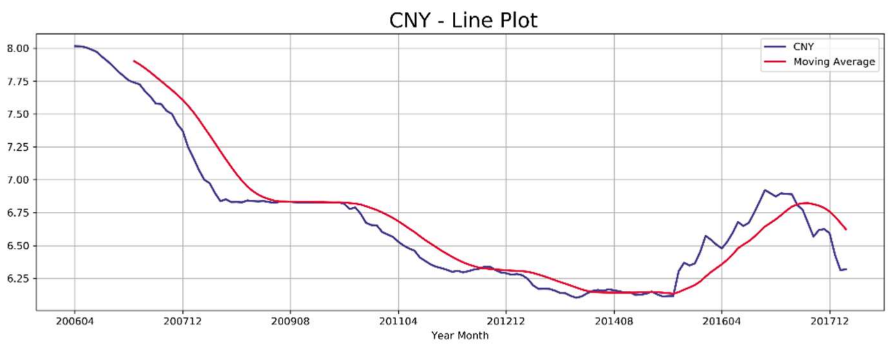
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.22. Turkish Lira vs US Dollars



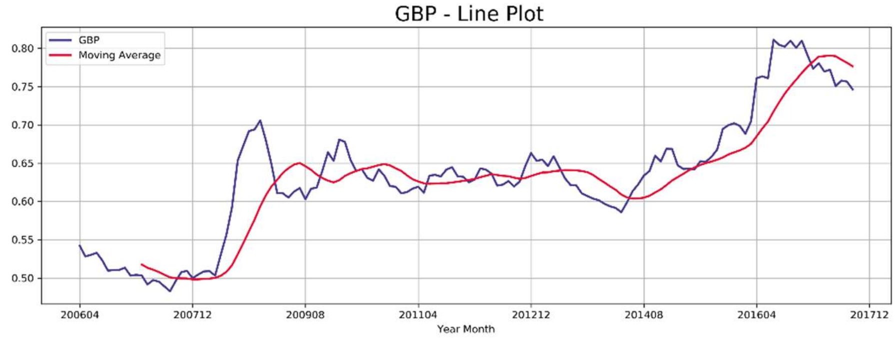
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.23. Chinese Yuan vs US Dollars



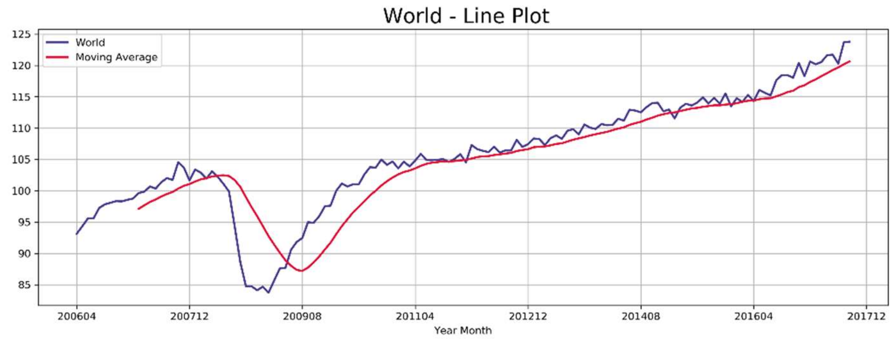
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.24. Great Britain Pound vs US Dollars



Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

Figure A2.25. World Trade



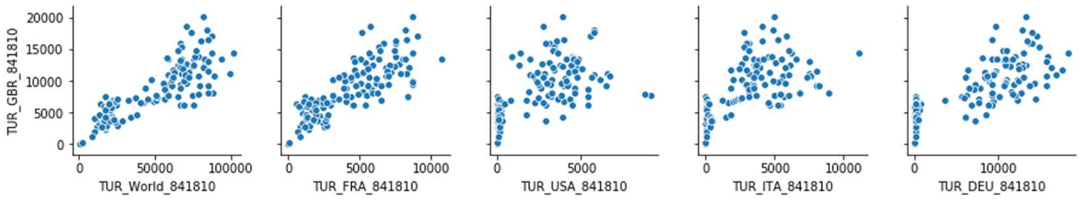
Source: Data retrieved from: OECD. (2019b). OECD.Stat. Retrieved July 20, 2019, from https://stats.oecd.org/ (date accessed: 20/07/2019).

APPENDIX 3 - Pair Plots of Dependent Variables versus Trade Couple

In this appendix pair plots of dependent variables versus product-related features are given.

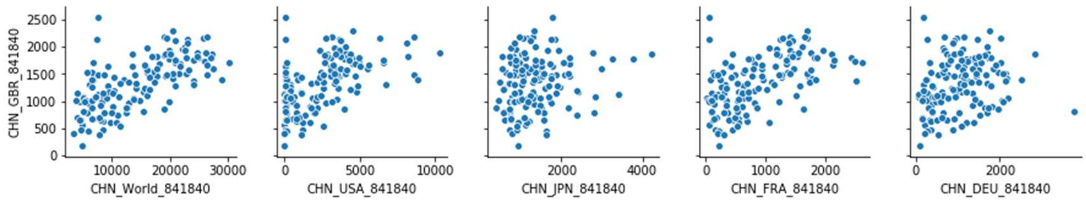
PRODUCT 84.18.10 PAIR PLOTS

Figure A3.1. Top Five Importer from Turkey



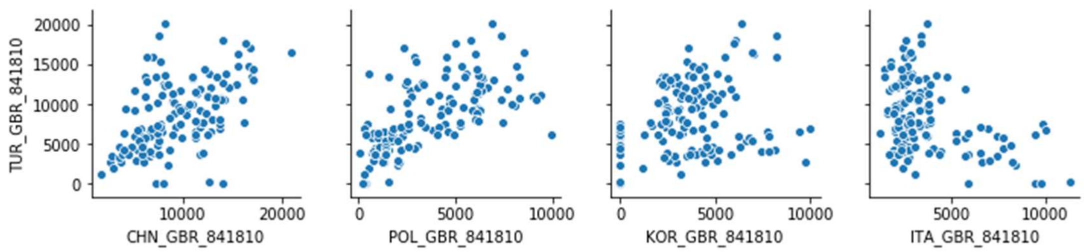
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A3.2. Top Five Importer from China



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

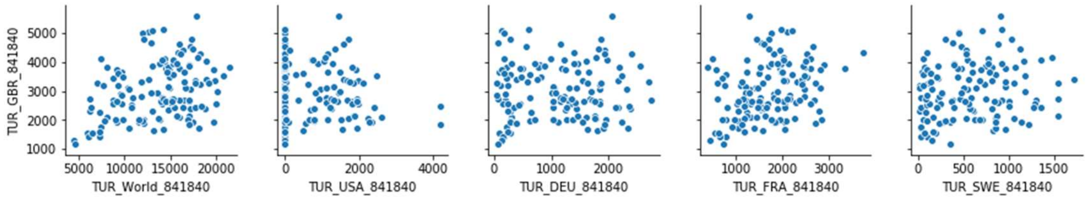
Figure A3.3. Top Five Exporter to the United Kingdom



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

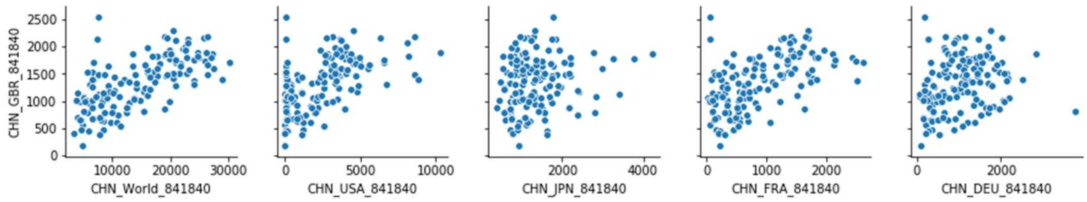
PRODUCT 84.18.40 PAIR PLOTS

Figure A3.4. Top Five Importer from Turkey



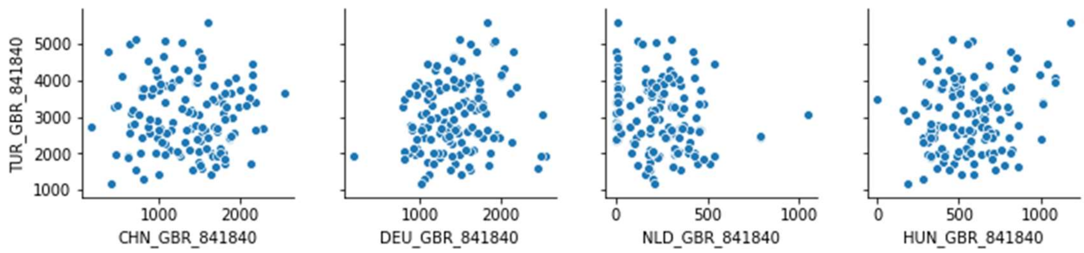
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A3.5. Top Five Importer from China



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

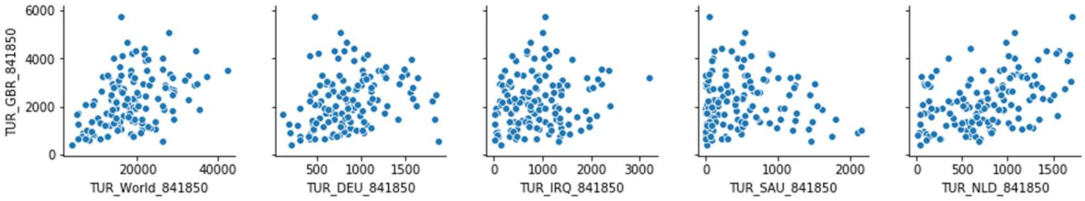
Figure A3.6. Top Five Exporter to the United Kingdom



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

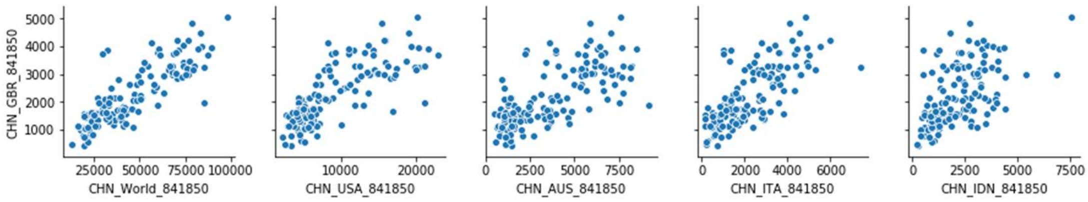
PRODUCT 84.18.50 PAIR PLOTS

Figure A3.7. Top Five Importer from Turkey



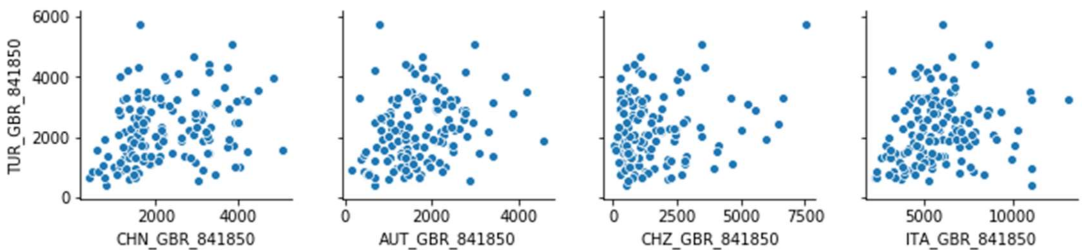
Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A3.8. Top Five Importer from China



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Figure A3.9. Top Five Exporter to the United Kingdom



Source: Data retrieved from: International Trade Center (2019). Trade Map. Retrieved from http://www.intracen.org/itc/market-info-tools/trade-statistics/ (date accessed: 20/07/2019).

Murat ÖZEMRE

Izmir | Turkey

PROFESSIONAL EXPERIENCE

03.2015 – Present	**Assistant General Manager Software Development** **BİMAR IT Services Inc., (ARKAS Holding), Izmir**

Establishing company's technical vision, strategic plan and leading the roadmap for technology and product development, ensuring and maintaining the effectiveness and efficiency of ARKAS Group, reporting to CIO.

11.2005 – 03.2015	**Software Development Manager** **BİMAR IT Services Inc. (ARKAS Holding), Izmir**

Responsible for strategic planning and execution of software technologies for ARKAS Group with an annual budget of 2,5 million USD.

11.2003 – 11.2005	**Technical Project Manager** **STM Defense Technologies and Engineering Inc., Ankara.**

Managing a 2 million USD budgeted, Military Command and Control Software Development Project on Java

08.2001 – 10.2003	**Senior Systems Engineer** **STM Defense Technologies and Engineering Inc., Ankara.**
03.2000 – 08.2001	**Project Management Officer** **Military Service & Turkish Land Forces, Ankara.**
10.1997 – 03.2000	**Systems Engineer & Software Developer** **Roketsan Missile Industries Inc., Ankara**
04.1997 – 10.1997	**Software Developer** **Roketsan & US Army, MD, USA.**
08.1996 – 04.1997	**Systems Engineer** **Roketsan Missile Industries, Engineering Development Division, Ankara.**

TECHNICAL PROJECTS

Projects in BİMAR, where main role is Program and Portfolio management

Transformation from Mainframe Systems to Service Oriented Software products;

- YNA Next Generation Agency. Developed new software for four ARKAS Agency Companies 2005-2010
- EDS Integrated Depot System. Developed new software for Arkas Logistics using in 21 Depots 2009-2011
- ARLES ARKAS Port Management System. Developed System for LİMAR to use in 2 Government 7 Private ports 2010-2012

Transformation from outsourced systems to in house development software products;

- YNL Next Generation Logistics. Developed new software for Arkas Logistics Replaced Soft platform. 2014-2015
- YNU Next Generation Transportation. Developed new software for Arkas Logistics. Replaced Catlogic platform. 2017-2018
- ARFLEET ARKAS Fleet Management System. New software for Arkas Ship Fleet. Cloud born project. (2018 – cont.). Will replace Shippernetix platform.
- ARMA New software for Arkasline. Cloud born project. (2018 - cont.) Will replace SEALINER platform.

Dissemination of software products to abroad offices;

- YNA Next Generation Agency. Dissemination program to 10 Countries (Italy, Algeria, Morocco, Egypt, Libya, Georgia, Spain, Portugal, Bulgaria, Tunisia). Each country held by a separate project. 2010 - still continues
- YNL Next Generation Logistics. Dissemination program to four Countries (Georgia, Ukraine, Russia, Azerbaijan). Each country held by a separate project. 2015-2016

Other projects;

- ARKAS Intranet Portal for Turkey. ARKAS Companies in Turkey. 2013-2014
- BOX. Feedering Management System. Developed new software for EMES Feedering Italy 2013- 2015
- FOT Electronic Invoice Distribution System. Developed new software for Arkas Holding Companies. 2017
- NAR News Analysis Reporting Portal for ARKAS Holding Companies 2017-2018
- Azure Migration Project (Turkey's biggest one in 2018), Program Manager, BİMAR, 2018
- ARKAS Intranet Portal for International offices. 2018-2019

Granted Projects from Governmental Organizations;

- YNA, Next Generation Agency, TUBITAK TEYDEB 2005-2008
- ENTAC Integrated Agency System Project TUBITAK TEYDEB 2008-2010
- EIMIS, European Inter-Modal Information System. EUREKA Project. 2010-2012

- LESBAM, Enterprise Service Bus and Business Activity Monitoring Framework for Container Logistics Business Processes Project. SANTEZ Project. 2011 - 2012
- INNOLOGI Innovative Logistics - Freight Village Management System. EUREKA Project. 2013-2015
- FORBIS Container Freight Transported by the Company Is Organizer (Forwarders) Private Enterprise Information Management System Izmir Kalkınma Ajansı 2015-2016
- BAKI, Linked Data Based Container Tracking System Project. SANTEZ Project. 2013 - 2015
- YNU, Next Generation Transportation Project TUBITAK TEYDEB 2017-2018
- NAR News Analysis Reporting TUBITAK TEYDEB 2017- 2018

Internal Development Platform Projects;

- Microsoft BIZTALK TAP (Technical Adaption Program) 2010 and 2012, the only participant company from Turkey so far.
- ROTA, Framework for Web and Windows Based development
- BIZFRAME, Framework for Enterprise Integration Development
- BIZWATCH, BizTalk Server and Integration Management Console

Projects in STM and ROKETSAN

- TAIKS Fire Support Command Control Communication System. STM Project Manager. Software developed for Turkish Army. 2003-2005
- C4ISR Naval Interoperability. Project Engineer. NATO Project. 2001-2003
- K System Artillery Rocket System System Engineer. Turkish Army Project. 1998-2000
- NABK Nato Artillery Ballistic Kernel. Software Developer. Software developed for NATO. 1997
- SHORAD Short Range Air Defense System. System Engineer. NATO Project. 1996

EDUCATION

YASAR UNIVERSITY, İzmir, Turkey

2013 – 2019

PhD: Business Administration
Thesis subject: A Big Data Analytics Based Methodology for Strategic Market Analysis

MIDDLE EAST TECHNICAL UNIVERSITY, Ankara, Turkey

1998 – 1999

Department of Management
Advanced Management - MBA Program

MIDDLE EAST TECHNICAL UNIVERSITY, Ankara, Turkey

1997 – 2000

Master’s Degree: Department of Aeronautical Engineering
Thesis subject: Neural Network Initialization of Strapdown Inertial Navigation Systems

MIDDLE EAST TECHNICAL UNIVERSITY, Ankara, Turkey

1991 – 1996

Bachelor’s Degree: Department of Aeronautical Engineering
Honor student, 2nd rank in class.

CERTIFICATIONS

- ISTQB Certified Tester, International Software Testing and Qualifications Board, 2007
- ISEB Foundation Certificate in IT Service Management ITIL, Information Systems Examinations Board, 2007

TECHNICAL ARTICLES

- Using Big Data Analytics to Forecast Trade Volumes in Global Supply Chain Management, Book Chapter in Managing Operations Throughout Global Supply Chains, 2019.
- Cohesive software measurement planning framework using ISO standards: A case study from logistics service sector, Journal of Software Maintenance and Evolution, April 2012
- An Approach to Find Integration and Monitoring Points for Container Logistics Business Processes, Service Computation, The Fourth International Conference on Advanced Service Computing, 2012, Nice, France
- Join into Software Process World with Three Steps, Software Quality and Software Development Tools Symposium,2008, Istanbul
- Artificial Neural Networks for Transfer Alignment and Calibration of Inertial Navigation System AIAA Missile Guidance, Navigation, and Control Conference, 2000, Montreal, Quebec, Canada