International Journal of Scientific and Technological Research www.iiste.org
ISSN 2422-8702 (Online), **DOI:** 10.7176/JSTR/5-3-19
Vol.5, No.3, 2019

A Comparison of SVM and Traditional Methods for Demand Forecasting in a Seaport: A case study

Hande Guzey
Faculty of Engineering, Department of Industrial Engineering
Uludag University, Bursa, Turkey
E-mail: hndgzy@gmail.com

Mehmet Akansel
Faculty of Engineering, Department of Industrial Engineering
Uludag University, Bursa, Turkey
E-mail: akansel@uludag.edu.tr

Abstract

Demand forecasting is important for planning future of a seaport facility. In this paper, different methods are compared for demand forecasting problem of a seaport in Turkey. Three types of data (general cargo, container, vehicle) were collected from the period of 2012-2017. Using machine learning for demand forecasting was found to be an important missing link in earlier studies and it was observed that the studies about demand forecasting on container terminals is a lot more than the studies on maritime terminals. Statistical forecasting methods and machine learning methods are applied for all types of data to determine the best estimation method and forecast the handling volumes for the next two years. The comparison of the forecasting performances of statistical forecasting methods and machine learning methods have been comparatively analysed. According to chosen accuracy measures, Multiplicative Holt Winter's was recognized as the best forecasting method for container and vehicle handling volumes, whereas machine learning method ensured the best forecasting values for the general cargo.

Keywords: Demand forecasting, seaport, machine learning, statistical modeling

DOI: 10.7176/JSTR/5-3-19

1. Introduction

As in other sectors, in ports, to plan the future, it is very important to determine whether the current capacity will be sufficient or not for the future demand. By forecasting demand which is is the first step of analyzing the capacity companies can plan their future [1].

Most of the earlier studies about capacity analysis were done at container terminals [2]. Studies about demand forecasting and capacity analysis are very limited especially for general cargo and vehicle handling sectors.

It was observed that traditional regression methods were widely used in the studies about forecasting on container volumes [3,4]. Meanwhile, demand forecasting is a requirement of many different sectors. For example, Tratar and Strmcnik [5] compared the methods of Holt&Winter's and Multiple Linear Regression for forecasting daily, weekly and monthly electricity consumptions.

Support Vector Machine (SVM) can be applied to classification and regression. Different SVM kernels can be used for time series prediction [6]. According to earlier studies applying kernel methods for time series prediction get similar results with other techniques [8, 9].

In this paper we compared different methods for demand forecasting problem of a seaport in Turkey. The seaport is handling three different types of cargo. For each cargo type, yearly berth capacity is calculated by using these handling volumes. Therefore forecasting the volumes correctly has a very important role for the capacity calculations. For this purpose statistical forecasting methods, regression analysis and machine learning methods applied for general cargo, container and vehicle handling volumes. For multiple linear regression and SVM methods, gross domestic product (GDP), population, inflation and foreign trade of Turkey were selected as the model input (predictor) variables [1,7]. The monthly values of predictor variables are created by applying cubic spline interpolation to yearly data [12]. The results were compared by using mean absolute percentage error (MAPE) and mean absolute error (MAE). Forecasting results with the best model for all operation types were used to analyse seaport handling capacity for the next 2 years.

We expect that this study will contribute to demand forecasting applications in maritime terminals. We also hope that data mining can be an integral part of a combined approach for demand forecasting in all other sectors of the economy.

This paper is organized as follows: Section 2 presents the details of the methodology. Section 3 presents the data and the comparison of the results from all methods. Finally, the findings of the paper are discussed and summarized in Section 4.

2. Methodology

In this section the forecasting methods used in the study are explained.

In the statistical forecasting part, the complete set of data which contains of 72 months was used for forecasting. But in the machine learning part, data were divided into two sets, such as training data set (first 5 years) and testing data set (last year). For each method training set was used for forecasting and then forecasted values were compared with the real data in the test set.

For performance comparison of the forecasting methods RMSE (Root Mean Squared Error), MAPE (Mean Absolute Percentage Error) and MAE (Mean Absolute Error) accuracy measures are applied.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} {e_t}^2} \quad (1)$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} |e_t| \quad (2)$$

$$\text{MAPE} = \left(\frac{1}{n}\sum_{t=1}^{n} \left|\frac{e_t}{y_t}\right|\right) * 100 \quad (3)$$

The lower values of the accuracy measures RMSE, MAPE and MAE represent a better forecasting performance. All the methods described below were applied to the data by using R Studio.

2.1 Exponential Smoothing Methods

2.1.1 SES (Simple Exponential Smoothing) Method

The SES method is used for forecasting a time series when there is no trend or seasonal pattern in data. In the equation [15], t is the time period, α is the smoothing constant for level in range (0,1) and Y_t is the observed data at time t. The initial value of the forecasted value is equal to last observed data. The forecasted value at time *t+1* is based on the value at time *t.*

2.1.2 Holt Method

The Holt method is useful for the time series when there is trend in data. The Holt method a contains one more smoothing parameter for trend in addition to the SES method as shown in the equations in [15].

2.1.3 Holt-Winters (HW) Methods

In this study two different HW methods were applied to data: 1) Additive Holt-Winters(AHW), 2) Multiplicative Holt-Winters (MHW).

Holt-Winters method is an exponential smoothing approach for handling seasonal data. AHW is used for time series with constant seasonal variations. MHW is used for time series with increasing seasonal variations. HW methods includes three smoothing parameters for level (α), trend (β) and seasonality (γ) as shown in the equations in [15].

2.2 ARIMA

Autoregressive Integrated Moving Average (ARIMA) is one of the common methods that can be fitted to time series data in order to predict future points in the series. ARIMA stands for auto-regressive integrated moving average and is specified by these three order parameters: *(p, d, q)*. *p* is the number of autoregressive terms, *d* is the number of differences and *q* is the number of moving averages [15].

2.3 Simple Linear Regression
It is possible to develop a linear regression model for time series analysis. The relationship between variables is described by a linear function. In the equation [14] general cargo/container/vehicle handling volumes are used as dependent variable and time is used as independent variable.

After calculate intercept and the slope coefficient, we can estimate the future value of the dependent variable in any given period.

2.4 Multiple Linear Regression
Multiple Linear Regression is the most widely used method for forecasting when there are two or more independent variables associated with a dependent variable [14].

In this study the total volume of import and export, the GDP and the population of Turkey are used as independent variables [1] and general cargo, container and vehicle volumes of the port are used as three dependent variables, which are obtained separately.

2.5 SVM (Support Vector Machine) Methods
For the machine learning we used the caret package [13] which supports the machine learning algorithms.
In machine learning we should split the data into a training set and a test set to correctly assess the performance of the model. Cross-validation is a technique for evaluating machine learning (ML) models by training several models on subsets of the available input data. There are different techniques for cross validation like N-Fold Cross Validation and Leave One Out Cross Validation.
N-Fold Cross Validation splits the data set into n number of subsets and perform training on all the subsets (n-1) but leave one subset for the evaluation of the trained model. LOOCV (Leave One Out Cross Validation) performs training on the whole data-set but leaves only one data-point of the available data set and then iterates for each data point. But for time dependent problems this methods has some disadvantages like testing model with a set of data older than the data used to train the model and time differences between the set of data used. In order to reduce these effects we used moving window approach as proposed in [11].
When SVM algorithms are used for regression problems, it is called Support Vector Regression (SVR). One of the **attractive** properties of SVR is the use of kernel functions. There are several types of kernel functions used in SVM. In this study we used linear, polynomial and radial basis kernels. For the hyper-parameters C and σ^2 he optimal pairs were used as proposed in [10]: $C \in \{2^{-3}, 2^{-2}, \dots, 2^{0}, \dots, 2^{9}\}$ and $\sigma^2 \in \{2^{-5}, \dots, 2^{0}, \dots, 2^{9}\}$.
In this study, we used the first 60 observations for training and the last 12 observations for testing. As a cross validation technique we used timeslice method that supported by caret package. We applied linear, polynomial and radial basis kernels to training data individually. For each method, the test set was used for forecasting and later the forecasted values were compared to real data in the test set. The model which has the lower errors was selected to forecast future values.

3. Data and Results
In this section, we describe the data used for the study. Then results from all the models and accuracy comparison of the models are reported.

3.1 Data
The monthly data of general cargo, container and vehicle handling volumes were collected from a seaport in Turkey for the period from January 2012 to December 2017 (72 observations). The data on the container, general cargo and vehicle handling volumes of seaport, are presented in Table 1. Because of the privacy policy of the port, we applied a scaling to the data in the tables.
For multiple linear regression and SVM methods, model input (predictor) variables were selected as gross domestic product (GDP), population, inflation foreign trade of Turkey similar to Esmer [1].
For GDP, actual rates for the years of 2012-2016 are obtained from Turkish Statistical Institute, predicted rates are obtained from World Bank [17] for the years 2017-2019. The actual and predicted values of population are obtained from Turkish Statistical Institute [16]. The actual values of foreign trade are obtained from Turkish Statistical Institute and predicted rates for the years 2018-2019 are obtained from OECD [18].

ISSN 2422-8702 (Online), DOI: 10.7176/JSTR/5-3-19
Vol.5, No.3, 2019

Table 1. Output variables used for forecasting models

Years	Months	General Cargo	Container	Vehicle	Year	Month	General Cargo	Container	Vehicle
2012	1	173921,00	12182,3	6187,35	2015	1	127138,41	11594,7	10648,95
2012	2	120551,94	8471,45	10182,3	2015	2	152927,59	9419,8	16443,7
2012	3	181563,62	9811,1	11063	2015	3	205145,35	10881,65	20375,55
2012	4	162171,24	8961,55	10396,1	2015	4	182817,98	12934,35	15479,1
2012	5	166516,03	10608	8380,45	2015	5	179776,42	14313,65	11419,2
2012	6	154416,70	10830,95	9525,1	2015	6	188224,31	12930,45	18243,55
2012	7	140598,25	10204,35	5796,7	2015	7	195835,06	10723,7	15103,4
2012	8	184395,29	10292,1	4317,3	2015	8	148931,58	12591,8	9896,9
2012	9	166383,81	10330,45	8671	2015	9	139461,13	13783,9	16075,8
2012	10	158389,44	10673	9456,85	2015	10	191816,03	12055,55	19778,2
2012	11	199489,29	10961,6	8599,5	2015	11	199383,86	12496,25	21039,2
2012	12	135982,81	9586,2	8103,55	2015	12	167828,30	12990,9	18733
2013	1	146049,01	11023,35	6073,6	2016	1	171603,94	11991,2	7527,65
2013	2	154535,92	9066,85	10326,6	2016	2	156358,18	10278,45	14647,75
2013	3	211676,28	11180,65	15547,4	2016	3	204132,41	14101,75	15959,45
2013	4	164533,60	10304,45	13740,4	2016	4	292970,91	14218,1	14105
2013	5	143055,43	12987,65	13822,3	2016	5	172556,96	12931,1	18708,3
2013	6	180032,49	12467	12879,8	2016	6	174783,56	16075,15	23000,25
2013	7	175108,78	13230,75	10691,2	2016	7	207243,50	11854,05	18476,9
2013	8	147455,58	12158,9	5912,4	2016	8	157337,46	15318,55	13510,9
2013	9	192864,37	12442,95	9159,15	2016	9	184285,92	12698,4	10887,5
2013	10	145347,40	11966,5	8703,5	2016	10	161445,20	14943,5	19162,65
2013	11	187517,67	12761,45	11687	2016	11	225984,19	13783,9	18515,9
2013	12	213221,32	12332,45	12062,1	2016	12	190181,55	13958,75	20623,85
2014	1	106917,28	13995,15	5666,7	2017	1	241316,04	12769,25	12708,8
2014	2	160088,05	10547,55	9340,5	2017	2	174615,35	9161,1	16521,05
2014	3	192645,31	12005,5	14888,9	2017	3	189018,22	15374,45	18566,6
2014	4	147980,55	12080,9	15505,8	2017	4	189873,45	13700,05	19734
2014	5	170704,01	13997,75	15330,9	2017	5	213999,18	11748,75	18811
2014	6	161507,29	13892,45	17084	2017	6	211054,56	13847,6	19383
2014	7	161070,77	12355,85	15271,8	2017	7	227774,56	12701,65	14216,15
2014	8	182299,76	11820,25	5172,7	2017	8	187170,53	13809,25	11969,75
2014	9	156111,68	10561,85	13446,6	2017	9	211018,05	12231,05	12110,8
2014	10	194104,82	11655,15	14606,2	2017	10	227835,30	13057,85	17642,3
2014	11	139594,43	12723,1	18120,7	2017	11	189684,02	12669,8	19934,85
2014	12	203769,89	11954,8	20352,8	2017	12	196135,52	16210,35	15466,1

We needed to convert the yearly data to a monthly data for GDP and population by using cubic spline interpolation [12]. Table 2 and Table 3 show the data before and after transformation. For the predicted values of foreign trade we assumed that the values for all months equal and we used the yearly data average for all months.

Table 2. Yearly Input Variables Used for Interpolation

Year	GDP	Population	Foreign Trade
2012	117,6	75,60	-
2013	122,4	76,70	-
2014	126,1	77,70	-
2015	133,79	78,70	-
2016	138,07	79,80	-
2017	148,28	80,80	-
2018	154,95	81,80	433,28
2019	161,14	82,80	461,87

Table 3. Monthly Input Variables Used for Forecasting

Year	Month	General Cargo	Container	Vehicle
2018	1	199028,90	13958,98	9432,28
2018	2	199413,76	10786,35	14528,86
2018	3	199803,70	13569,06	18372,64
2018	4	200198,70	13425,17	16778,68
2018	5	200598,65	14290,16	16314,55
2018	6	201003,66	14921,74	18374,49
2018	7	201413,68	13402,06	14222,31
2018	8	201828,77	13897,01	9216,76
2018	9	202224,88	13272,69	14241,78
2018	10	202673,90	13654,44	16603,82
2018	11	203104,01	13984,23	18077,24
2018	12	203539,18	13829,34	17605,62
2019	1	207387,31	14138,98	9721,19
2019	2	207832,50	10921,72	14962,34
2019	3	208282,69	13734,78	18906,62
2019	4	208737,95	13584,75	17253,76
2019	5	209198,21	14455,48	16764,61
2019	6	209663,48	15089,76	18868,37
2019	7	210133,76	13548,94	14594,79
2019	8	210609,10	14045,26	9451,99
2019	9	211089,45	13410,52	14596,00
2019	10	211574,81	13792,46	17006,30
2019	11	212065,23	14121,83	18504,32
2019	12	212560,66	13961,81	18011,04

3.2 Results for the Models

In this section, we present the results and the comparison of performance measures from all the models. Table 4-6 show the accuracy results of all models for general cargo, container and vehicle handling volumes, respectively. The additive Holt-Winter's method is clearly the best forecasting model for the general cargo volumes since it has the lowest values of all the three performance measures as shown in Table 4. Table 5 and 6 show that the multiplicative Holt-Winter's method obtains the best results among all methods applied to container and vehicle handling volumes.

According to these performance results the forecasting results of each method with its best parameters are shown in Table 7 for all operation types.

Table 4. Performances of the Models for General Cargo Forecasting with the Best Model Parameters

	MAPE (%)	MAE ($x10^3$)	RMSE ($x10^3$)	Parameters
Exponential Smoothing	12,09	32,95	43,35	α = 0.0986
Holt's Linear Method	12,08	32,35	42,07	α = 0.0282, β = 0.0108
Additive Holt-Winters	10,94	28,89	37,90	α = 0.0318, β = 2e-04, γ = 1e-04
Multiplicative Holt-Winters	11,10	29,11	37,72	α = 0.0222, β = 2e-04, γ = 1e-04
ARIMA	11,25	29,80	39,13	p = 2, d = 1, q = 2
Simple Linear Regression	12,06	31,98	40,76	β_0 = 235146, β_1 = 1063
Multiple Linear Regression	11,81	31,41	40,25	β_0 = -547958.5, β_1 = 2202.2, β_2 = 849.5, β_3 = 8282.1,
Support Vector Machine	7,99	27,03	37,03	C = 1

Table 5. Performances of the Models for General Cargo Forecasting with the Best Model Parameters

	MAPE (%)	MAE ($x10^3$)	RMSE ($x10^3$)	Parameters
Exponential Smoothing	9,23	1,72	2,22	α = 0.1755
Holt's Linear Method	8,74	1,60	2,07	α = 1e-04, β = 1e-04
Additive Holt-Winters	7,04	1,32	1,64	α = 1e-04, β = 1e-04, γ = 1e-04
Multiplicative Holt-Winters	6,87	1,30	1,61	α = 1e-04, β = 2e-04, γ = 1e-04
ARIMA	9,13	1,69	2,22	p = 0, d = 1, q = 1
Simple Linear Regression	8,98	1,64	2,10	β_0 = 16082.8, β_1 = 73.7
Multiple Linear Regression	8,63	1,58	2,04	β_0 = -57422.05, β_1 = 160.08, β_2 = -1.61, β_3 = 917.84,
Support Vector Machine	9,64	1,84	2,78	C = 0.25, σ = 0,3922293

Table 6. Performances of the Models for General Cargo Forecasting with the Best Model Parameters

	MAPE (%)	MAE($x10^3$)	RMSE ($x10^3$)	Parameters
Exponential Smoothing	26,72	4,74	5,73	α = 0.164
Holt's Linear Method	27,65	4,12	5,47	α = 1e-04, β = 1e-04
Additive Holt-Winters	16,77	2,92	3,51	α = 0.2808, β = 2e-04, γ = 0.0013
Multiplicative Holt-Winters	15,20	2,83	3,57	α = 0.5301, β = 1e-04, γ = 1e-04
ARIMA	23,82	4,12	5,07	p = 2, d = 1, q = 2
Simple Linear Regression	26,41	4,45	5,42	β_0 = 12961.2, β_1 = 219.5
Multiple Linear Regression	23,29	4,01	5,00	β_0 = -635373.8, β_1 = 650.3, β_2 = -1408.4, β_3 = 10495.3,
Support Vector Machine	16,90	4,01	4,74	C = 0.25, σ = 0,3954284

Table 7. Forecasting Results with Best Model

Year	Month	General Cargo	Container	Vehicle
2018	1	199028,90	13958,98	9432,28
2018	2	199413,76	10786,35	14528,86
2018	3	199803,70	13569,06	18372,64
2018	4	200198,70	13425,17	16778,68
2018	5	200598,65	14290,16	16314,55
2018	6	201003,66	14921,74	18374,49
2018	7	201413,68	13402,06	14222,31
2018	8	201828,77	13897,01	9216,76
2018	9	202224,88	13272,69	14241,78
2018	10	202673,90	13654,44	16603,82
2018	11	203104,01	13984,23	18077,24
2018	12	203539,18	13829,34	17605,62
2019	1	207387,31	14138,98	9721,19
2019	2	207832,50	10921,72	14962,34
2019	3	208282,69	13734,78	18906,62
2019	4	208737,95	13584,75	17253,76
2019	5	209198,21	14455,48	16764,61
2019	6	209663,48	15089,76	18868,37
2019	7	210133,76	13548,94	14594,79
2019	8	210609,10	14045,26	9451,99
2019	9	211089,45	13410,52	14596,00
2019	10	211574,81	13792,46	17006,30
2019	11	212065,23	14121,83	18504,32
2019	12	212560,66	13961,81	18011,04

Figure 1-3 show the actual values in the dataset and the forecasted values produced by using the best forecasting methods for all operation types, respectively.

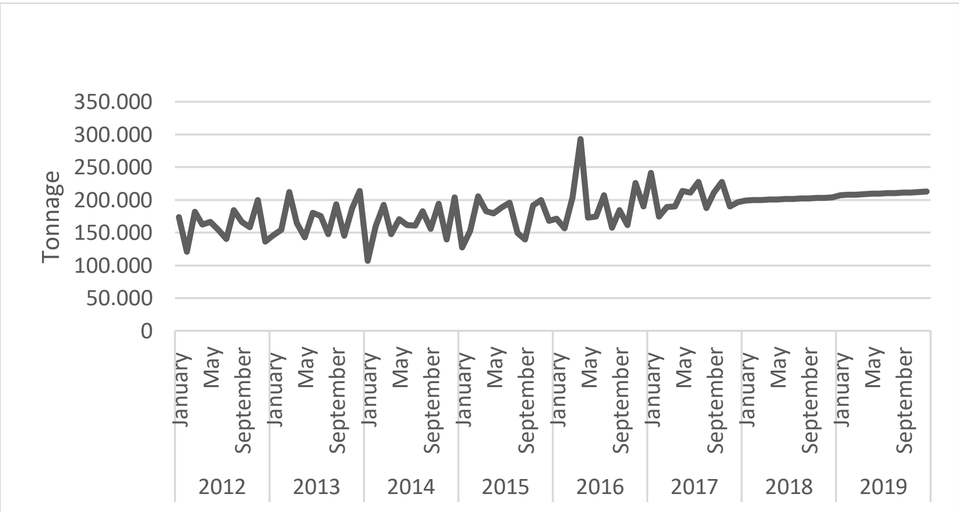

Figure 1. Real data and forecasts for the general cargo handling volumes

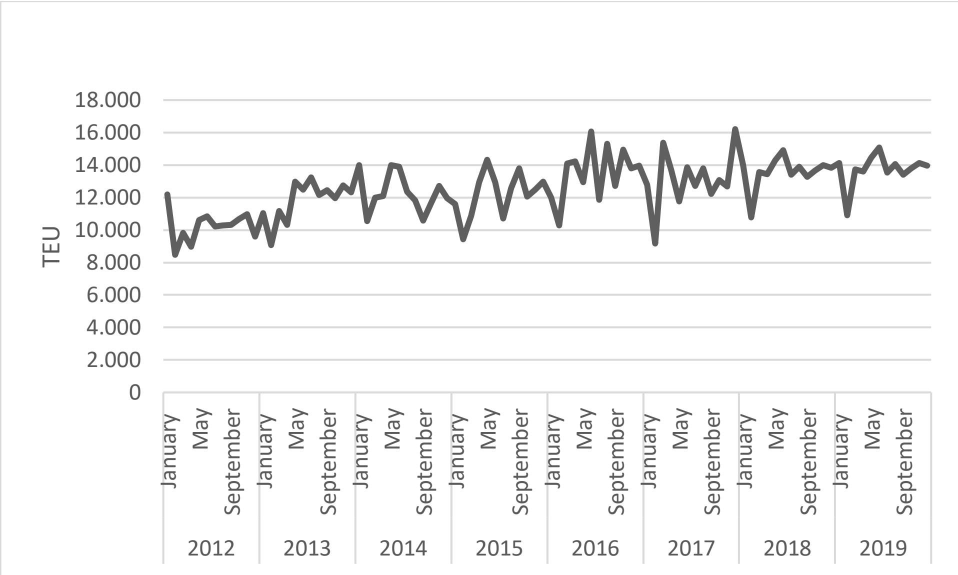

Figure 2. Real data and forecasts for the container handling volumes

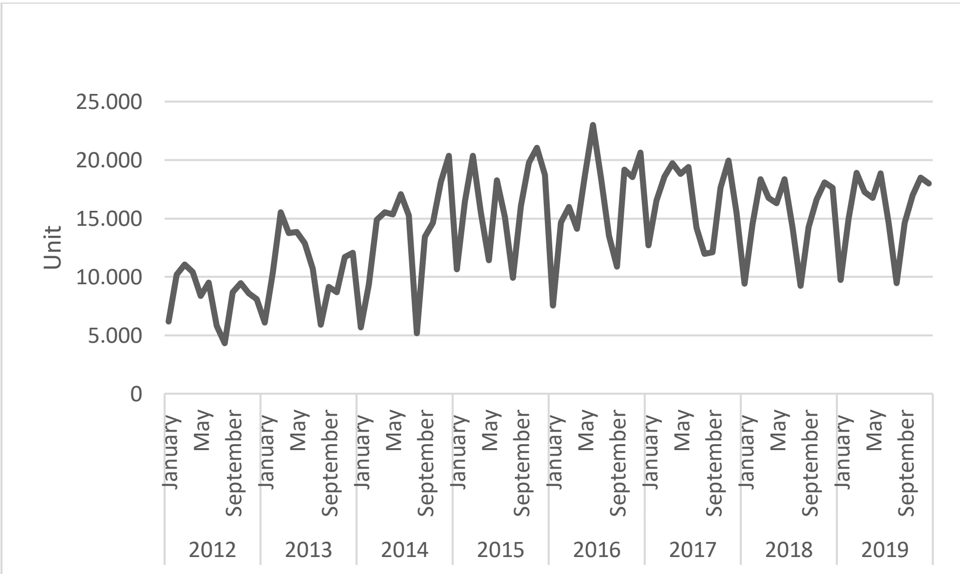

Figure.3. Real data and forecasts for the vehicle handling volumes

4. Conclusion

The aim of this paper is to determine the best estimation method for general cargo, container and vehicle handling volumes and forecast the values for the next 2 years. We applied exponential smoothing, regression and support vector regression methods to three types of operation data, respectively. According to the chosen accuracy measures, Multiplicative Holt-Winter's was recognized as the best forecasting method for container and vehicle handling volumes, whereas machine learning method with linear kernel ensured the best forecasting values for general cargo. We can say that applying SVM kernel methods for prediction get similar results with other techniques. The results show that the SVM methods can be applied to demand forecasting but Holt-Winter's is more suitable than SVM for the data which has a seasonality.

ISSN 2422-8702 (Online), DOI: 10.7176/JSTR/5-3-19
Vol.5, No.3, 2019

References

[1] Orkut Akar, Soner Esmer. (2015). Cargo Demand Analysis of Container Terminals in Turkey. *Journal of ETA Maritime Science*, 2105, 3(2):117-122.

[2] Emin Deniz ÖZKAN, Selçuk NAS, Nil GÜLER. (2016). Capacity Analysis of Ro-Ro Terminals by Using Simulation Modeling Method. *The Asian Journal of Shipping and Logistics, (2016), 32(3):139-147.*

[3] Chien Chang Chou, Ching-Wu Chu, Gin-Shuh Liang. (2008). A modified regression model for forecasting the volumes of Taiwan's import containers. *Mathematical and Computer Modelling 47 (2008), 797-807*

[4] Wen-Yi Peng, Ching-Wu Chu. (2009). A comparison of univariate methods for forecasting container throughput volumes. *Mathematical and Computer Modelling* 50 (2009) 1045-1057.

[5] Liljana Ferbar Tratar, Ervin Strmcnik. (2016). The Comparison of Holt-Winters method and Multiple regression method: A case study. *Energy* 109 (2016) 266-276.

[6] Gang Xie, Shouyang Wang, Yingxue Zhao, Kin Keung Lai. (2013). Hybrid approaches based on LSSVR model for container throughput forecasting: A comparative study. *Applied Soft Computing* 13 (2013) 2232-2241.

[7] Ümit Gökkuş, Mehmet Sinan Yıldırım, Metin Mutlu Aydın. (2017). Estimation of Container Traffic at Seaports by Using Several Soft Computing Methods: A Case of Turkish Seaports. *Discrete Dynamics in Nature and Society Volume 2017,* Article ID 2984853, 15 pages.

[8] Gin´es Rubio, H´ector Pomares, Luis J. Herrera, and Ignacio Rojas. (2007). Kernel Methods Applied to Time Series. *International Work-Conference on Artificial Neural Networks IWANN 2007: Computational and Ambient Intelligence* 782-789.

[9] Stefan Rüping. (2001). SVM Kernels for Time Series Analysis. *LLWA 01-Tagungsbandder GI-Workshop-Woche Lernen-Wissen-Adaptivity (2001),* 43–50.

[10] K.L. Mak, D.H. Yang. (2007). Forecasting Hong Kong's Container Throughput with Approximate Least Squares Support Vector Machines. *Proceedings of the World Congress on Engineering 2007 Vol I.*

[11] Luis F abregues de los Santos. (2017). *Forecasting Financial Time Series Using Multiple Kernel Learning, Master Thesis, Dept. of Computer Science, Universitat Polit_ecnica de Catalunya*.

[12] Ajao, I.O., Ibraheem, A.G., Ayoola, F. J. (2012). Cubic Spline Interpolation: A Robust Method of Disaggregating Annual Data to Quarterly Series. *Journal of Physical Sciences and Environmental Safety*, Volume 2, Number 1.

[13] Max Kuhn. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software, November 2008*, Volume 28, Issue 5.

[14] Ronald E. Walpole. Probability & Statistics for Engineers & Scientists Eight Edition.

[15] Makridakis S, Wheelwright SC, Hyndman RJ. Forecasting Methods and Applications. 3rd ed. New York: Wiley; 1998.

[16] TSI. Turkish Statistical Institute. http://www.turkstat.gov.tr/UstMenu.do?metod=temelist

[17] World Bank. https://datacatalog.worldbank.org/dataset/global-economic-prospects

[18] OECD. https://data.oecd.org/trade/trade-in-goods-and-services-forecast.htm