Building Predictive CRM Model for Egyptian Electronic Corporations

Ahmed Abdelazim Mohammed Mohammed Eldaw
Sadat Academy for Management Science, Faculty of Management sciences
Department of computers and Information systems It
ahmed2013@yahoo.com

Abstract

It is known that to create an Intelligent CRM applying Machine learning algorithms for the problem that exists will be needed. Machine learning is becoming a more popular feature in CRMs and unlocks more insight into the data as Machine learning provides companies with a 360° view of its customers and can detect patterns for companies before they become obstacles or problems. This paper is a research building a predictive CRM model for Egyptian Electronic Corporations by building the predictive part of CRM that helps to forecast the future, Results: the research proves the assumption of the relationship between series existence and creates a proposed model called N-ARIMA which is the second model in lower error rate but has more stable performance than first model Neural network model as well as comparing many common model results with research proposed model on the data-set.

Keywords: *Customer Relationship Management, Time series, Forecasting, Machine learning*

1. Introduction

CRM is Customer Relationship Management. Others prefer to call it Client Relationship Management. Essentially, CRM Software enables businesses to manage business relationships, the data, and the knowledge obtained through its information connected with them. Strong CRM software solutions are built around the people and relationships as in any business, a need to build strong relationships with the customers is an important factor.

CRM is a strategy and technology that is applied to build stronger relationships between corporations and their clients. corporations will save information that is related to their clients, and employees will analyze it to utilize it in forecasting and creating reports. Some of the strategies of CRM software are designed for targeted marketing campaigns towards particular clients. The strategies applied will be relying on the information that is coming from the CRM system. CRM software solution is usually utilized by all kinds of corporations that focus on keeping a strong relationship with their customers.[1]

Machine learning within a CRM assists in serving the customers in several ways like looking at the past choices and data to see which behaviors led to more reliable and better solutions, supporting the company for making the most knowledgeable decisions with new customers by recommending the next most suitable actions as well as continuously learning.

This paper is aiming at building the predictive part of CRM that helps to forecast the future for the data-set of Egyptian Electronic Corporations that can be later integrated with CRM easily through API. The importance of research is that the predictive part is very critical for any decision-making situation in many Egyptian Corporations for example to see client devices' interests in the future to decide which device the Corporation will buy more from it.

1.1 Literature review

An Efficient CRM-Data Mining Framework for the Prediction of Customer Behavior is a paper that finds close customer relationships and manages the relationship between corporations and customers using machine learning.[2] Data mining has obtained a reputation in various CRM applications nowadays and by using machine learning with the CRM model to create the classification model that is important in data mining and valuable in the field. The model is applied to predict the behavior of customers to improve the decision-making processes for maintaining valued customers. the research proposed two classification models, Naive Bayes and Neural Networks which show the importance of machine learning with CRM.[2]

Statistical And Machine Learning Forecasting Methods is a research that shows a comparison between machine learning algorithms common methods like (naive, LSTM, KNN, MLP, SVM, etc.) and classical time series algorithms like (ETS, ARIMA, Comb, theta, etc.) on time series problem show that statistical or classical time series methods achieve the best results and machine learning is outperformed by simple statistical methods.[3]

ARIMA is a famous univariate time series forecasting models that has been implemented in many fields, including CRM and port container throughput [4], ARIMA applies the old data of the variables to make short-term forecasting effectively [5, 6]. Abdel-Aal and Al-Garni employed a univariate method (ARIMA) to forecast energy consumption in eastern Saudi Arabia.[7] The 5 years of energy consumption data were utilized to train the models and then forecast the sixth year's energy consumption. The performance of the suggested approach is compared with regression and abductive network machine-learning methods.[7]

The outcomes revealed that the ARIMA models need less data, hold fewer coefficients, and more accurate. Song et al. applied six alternative econometric methods to forecast tourism demand in Denmark.[8]

All the models were created from an Autoregressive Distributed Lag Specification (ADLS). The annual tourism data from 1969 to 1993 were obtained to build the models and the data from 1994 to 1997 were used to compare the performances of six models. The results revealed that the Time-Varying Parameter approach had the best performance of one-year-ahead forecast.[8]

the previous paper showed the possibilities of using an econometric model to do the time series forecasting. Rashed used different derivative models of ARIMA to forecast container throughput of the port of Antwerp.[9]

1.2 Data-set

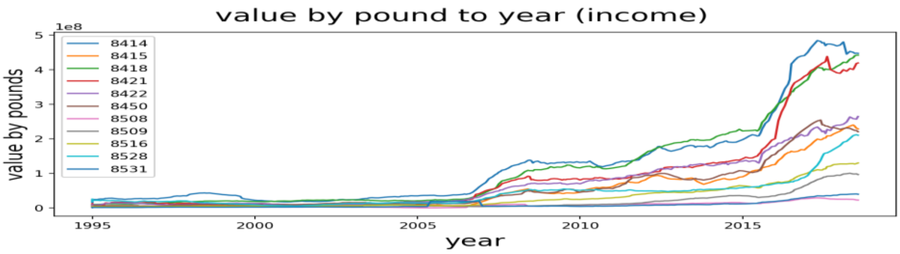

Figure 1: income series data-set Visualization

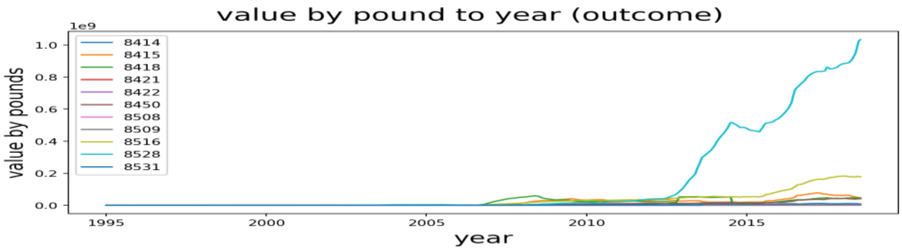

Figure 2: outcome series data-set Visualization

The data-set has five attributes (product code, year, month, income outcome, price-in-pound) what the models forecast is price-in-pound value and the data-set is two parts, first part is income-data-set, the second part is outcome data-set and the data-set has the same 11 products codes in income and outcome data-sets, it also has 3284 records in income and 2482 records in the outcome that start from the year 1995 to the year 2019 as shown in figure 1 and figure 2.

1.3 Contributions

This paper presents a proposed model for Predictive CRM Model for Egyptian Electronic Corporations data-set that has proved the best, stable, and non-consuming computational power its name is N-ARIMA that shows the good result with advantages but yet is not the lowest but the proposed model was chosen based on the flexibility over the time and computational consumption from the lowest algorithm.

This paper also proved the assumption that has been made on the data-set as well it could prove one of the related work phenomena of outperforming of common machine algorithms by statistical methods for time-series problem in the data-set but not completely next sections will show why.

The contributions of this paper are:

- The best methodology for the data-set in long term usage that outperforms almost all other models in error rate and all the models in advantages, this methodology or proposed model comes from a comparison that has been made between common machine learning models and statistical time series models.
- A phenomenon is demonstrated which is outperform of common machine learning models by classical or statistical time series models but yet not completely.
- An assumption of a series relationship between each other is made on data-set and proves that it is right due to the data-set limitations.

1.4 Article structure

this paper is structured as Abstract then Introduction comes with 3 sub-sections (Literature review, Dataset, Contributions) and then section Machine Learning Models come with 3 sub-sections (pre-possessing, Regression models, Statistical time-series models) finely end with section Results which has 3 sub-sections (Proposed Model N-ARIMA vs Common Models, The Assumption Proof, The phenomenon of outperforming)

2. Machine Learning Models

In this paper, nine machine learning models have been applied as six common regression models and three statistical time-series models with different prepossessing for both regression and statistical time-series models so this section will show methods used in prepossessing as well as any other methods used in prediction if there is and show the proposed model.

2.1 pre-possessing

2.1.1 Regression pre-possessing

First split the data-set into 75% train and 25% test then applies certain preprocessing in regression called standardized train set and test set by train std and not using normalization because it needs the means and the means is changing or moving in the regression problem and because Standardization is the method that puts different variables on the same scale. In regression analysis or models, there are some situations where it is essential to standardize the independent variables to avoid risk obtaining misleading results.[10]

2.1.2 time-series pre-possessing

First melt the data into a new form in which the rows represent the months and columns represent the time series which are income and outcome for each product so rows are 299 months and 22 columns which is the number of products multiply by 2(income and outcome) and then fill the NAN values of the products that do not exist in this month by zero, then split the data-set into 75% train and 25% test then apply certain pre-processing called differentiating on the train set until all series are stationary using Adfuller test[11] to make sure that all series are stationary because time-series models can only work on series that are stationary

2.2 Regression models

This section will show the regression models as shown in figure 3, the preprocessing that the previous section explained is done then the models trained then evaluated

2.2.1 Bayesian Regression

The purpose of the Bayesian Linear Regression is not to determine the single “best” value of the model parameters, but somewhat to find the posterior distribution for the model parameters as well as response. The posterior probability is conditional for the parameters of the model in the training inputs and outputs.

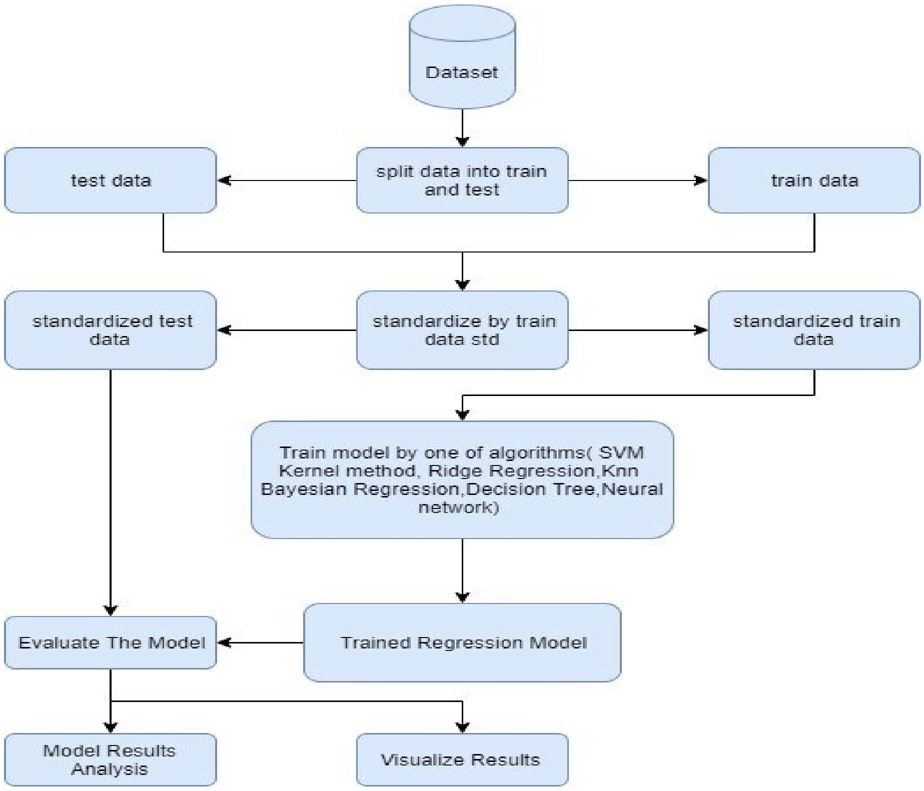

Figure 3: Regression Model

2.2.2 Ridge Regression

Ridge regression is an expansion of linear regression where the loss function is adjusted to minimize the complexity of the model. This adjustment is done by adding a penalty parameter that is equal to the square of the magnitude of the coefficients.

Loss function = OLS + alpha * summation (squared coefficient values)

2.2.3 SVM Regression

SVMs are based on the concept of finding a hyper plane that best separates a dataset. in regression, the main concept of SVM is always the same to minimize error, individualizing the hyper plane which makes the margin maximized, keeping in mind that part of the error is allowed to exist.

2.2.4 KNN Regression

KNN regression is to measure the mean of the numerical target of the K nearest neighbors and the nearest is decided by distance methods. Another method uses an inverse distance weighted mean of the K nearest neighbors with a parameter of k = 3.

2.2.5 Decision Tree

The decision tree algorithm produces regression or classification models in the structure of a tree. It splits a dataset into tinier and tinier subsets while at the same time the decision tree is incrementally improved. The terminal result is a tree with decision nodes and leaf nodes decided by the depth and the parameter of the depth=9.

2.2.6 Deep Neural Network

Neural networks are reducible to regression models, a neural network can “represent” any type of regression model. this very simple neural network in the next figure, with only one information input neuron, one hidden neuron, and one output neuron, is equal to logistic regression. It takes several variables that have relation = input parameters, multiplies them by their coefficients = weights, and works them through a sigmoid activation function and a unit step function, which equals the logistic regression, and used a heavy neural network that has numbers of each neural in each layer in this model are (25, 100, 200, 100, 150, 250, 250, 250, 150, 100, 50, 10, 1])

2.3 Statistical time-series models

This section will show the Statistical time-series models as shown in figure 4, the pre-processing that the previous section explained is done then the models trained then evaluated

2.3.1 VAR

VAR is a Vector Auto regression model [12] that is one of the time-series models that is modeled as a linear combination of its lags. that use the past values of the series in the model to forecast the future.

2.3.2 VARMA

This model uses the Auto regression Moving Average model [13] but for multiply series just like VAR but instead of using vector on the Auto regression model it will use a vector of series on the Auto regression Moving Average model which present a parsimonious description of a (weakly) stationary stochastic method in terms of two polynomials or two models, one model for the auto regression (AR) and the second for the moving average (MA).

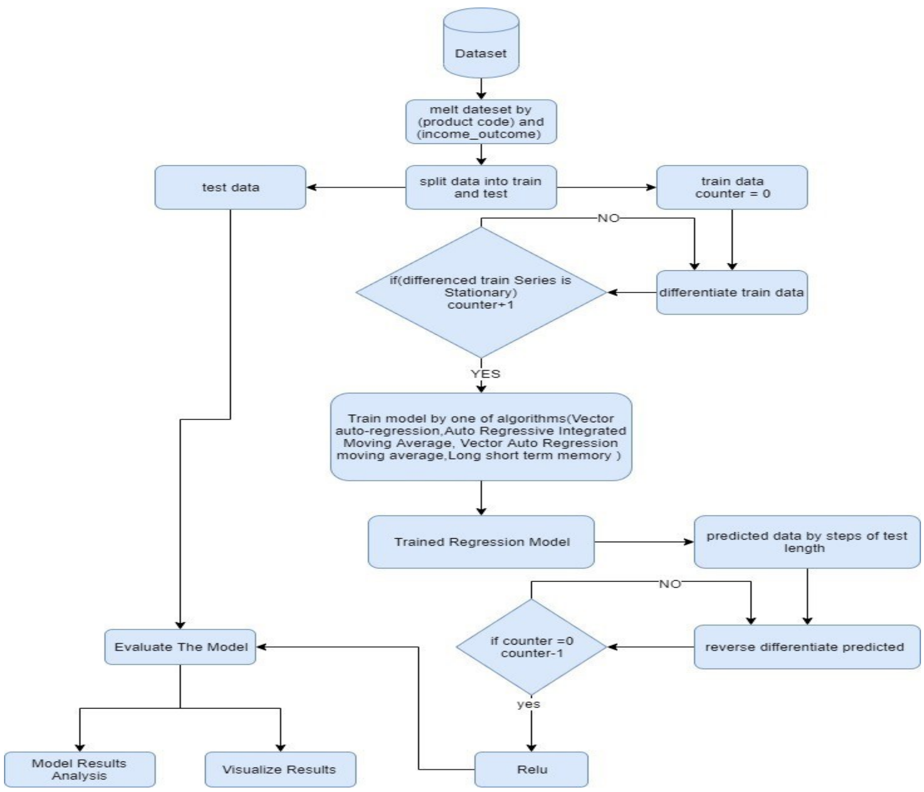

Figure 4: time-series Model

2.3.3 The Assumption

Before showing the last model proposed model N-ARIMA model, let's discuss why the vector algorithms are used, this because that the data-set has multiple series right, but did not the previous sections explain that the data-set is small which means it may not be big enough to have dynamic relationships between multiple series.

The assumption is there are no dynamic relationships between the multiple series so each series is on its own which means AR, ARMA, or ARIMA can be used with the data-set, and from research Statistical and Machine Learning Forecasting Method, it shows high results for the ARIMA model.

2.3.4 Proposed Model N-ARIMA

The proposed model is a combination of ARIMA models as shown in figure 5, The ARMIA model is a method of regression analysis that measures the depth of one dependent variable to other changing variables. The ARIMA model's target is to forecast future stock or market changes by observing the differences between values in the time series instead of real values.

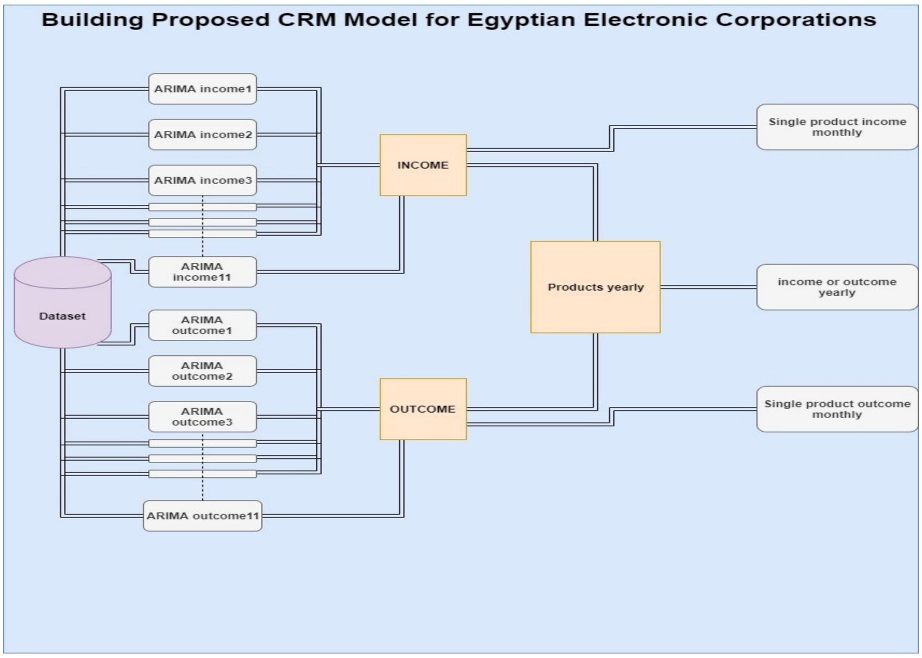

Figure 5: Proposed Model N-ARIMA

The ARIMA model can be explained by describing each of its components:

- Auto regression (AR) applies to a model that presents a changing variable that regresses the model on its own lagged, or prior values.
- Integrated (I) describes the difference of new observations to make the time series stationary, data values are substituted by the difference between the data values and the previous data values.
- The moving average (MA) combines the dependency between the data observation and a residual error from a moving average model used to lagged data observations.

This combination as shown in figure 5 is used by the income and outcome module that gives an output which is income for a single product or outcome monthly for a single product to the user or their outputs go to the module of products yearly that give all products income or outcome yearly or monthly.

3. Results

3.1 Proposed Model N-ARIMA vs Common Models

After training each model from both common regression models and statistical time-series models and evaluating them by 25% from data using matrices of Root Mean Square Error (RMSE) the results of each model have been put into the table then visualize it by order of smaller error and due to high price numbers that we deal with the error is too big as the max price on the data-set which is 10 digits number 15783650000 and this is only one price from 5766 records so RMSE might not be familiar to you which is a value between 0 and 1 but RMSE is not fixed to value as it is related value and train RMSE is similar to test RMSE.

Table 1: Proposed Model N-ARIMA vs Common Models

information / models	test error RMSE	rank
SVM	207948180.910	8
Ridge	210899504.777	9
KNN	192829803.850	6
Bayesian	227666043.348	10
Decision Tree	188368306.277	4
MIN Neural Network	175647859.808	1
MAX Neural Network	190587877.925	5
VAR	200917332.141	7
VARMA	186331767.329	3
N-ARIMA	175859969.678	2

Before evaluating the time-series model, the results of time-series models were passed on the reverse function of differentiating pre-processing function to obtain real values then pass real values to the RELU function that make negative values zeros because some of the series was near to zero which makes negative values which do not exist on the logic of data-set or the corporation.

As shown in figure 6, first 3 models are Neural Network, N-ARIMA, and VARMA which mean first one is Neural Network not N-ARIMA so why this paper recommend N-ARIMA instead of Neural Network that because Neural Network as shown in figure 6 and as seen in table 1 has them min result and max result which mean result is not stable when it was trained multiple times and by taking factor of train time which is nearly 1 hour by GPU support due to Neural Network size which is 12 hidden layers with max neural value 250 used 3 times in layers which mean non stable with computational power and time consuming and need to change architecture every time data increases while on the other side N-ARIMA took just minutes to finish 22 models which are 11 products for income series and another 11 for outcome series with stable results which put N-ARIMA as best choice for the 5k records data-set that may increases in products and records for old products in the near future and timeseries is preferred because it is made to forecast future so this paper are not really sure about Deep Neural Network because it has many disadvantages with risks so the proposed model which is very near to Deep Neural Network RMSE through figure 6 it looks like they equal so this paper highly recommend the proposed model for this data-set and to test it on the other data-sets for being the best stable non-changing non-computational consuming model.

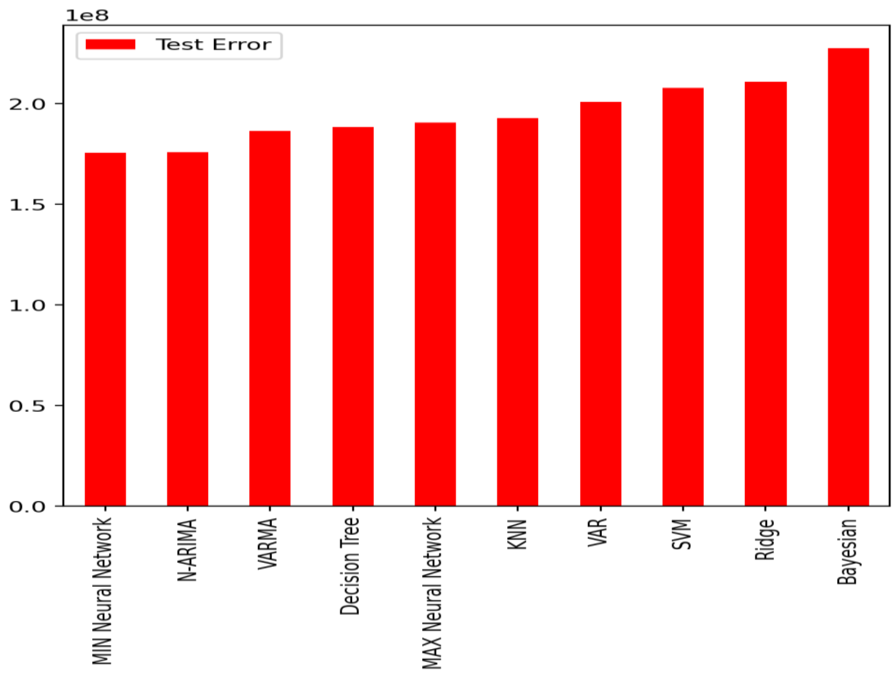

Figure 6: Proposed Model N-ARIMA vs Common Models Visualization

But other regression models like Decision Tree that is second-best in the regression model with stable and non-computational consuming all algorithms but Neural Network and SVM are non-computational consuming models and get stable results.

3.1 The Assumption Proof

Through figure 6 N-ARIMA is the proof that the assumption of the nonexistence of dynamic relationships between the multiple series so each series is on its own, is the right call, and the results of N-ARIMA is way better than VARMA.

3.2 The phenomenon of outperforming

Through figure 6 two time-series models were better than the five regression models which mean 66% of time series models were better than 83% common regression models which proved Makridakis, S., Spiliotis, E. and Assimakopoulos, V., 2020.” Statistical and Machine Learning Forecasting Methods” research that shows statistical methods like ARIMA outperform machine learning models [3] which makes us want to think about how could we improve those statistical methods in the near future.

4. Conclusion

The results and discussion that have been shown in This paper demonstrate that the proposed model N-ARIMA is the best choice for the data-set in many factor like factor of time, computational power, and stability in forecasting the future data, while it show the proposed model it discuss and compare common machine learning models with statistical time-series models showing that two statistical time-series models is better than almost the rest but Deep Neural Network is the best but it costs too much resources, which indeed worth researching with but due to its limitation, this paper do not recommend it for long term usage that may lead to continuous changing in the architecture with the need to change parameters of Neural Network too many times to find best parameters and test each one multiple times which is tiring and time consuming and by adding GPU factor which is the computational power used that does not exist with everyone nowadays and yet this paper did not talk about possibility of data increasing in way that increase work for Neural Network. ARIMA and VARMA, those two statistical models proved the phenomenon of outperforming of common machine learning models by classical or statistical time series models, also this paper has proved the assumption of the non-existence of series dynamic relationship between each other in the data-set which leads to the N-ARIMA model creation as well as the combination of this model with module class make it easier to simply add it in any CRM easily using API to allow multiple usages if needed because proposed model N-ARIMA which is the second model in lower error rate but has more stable performance than first model Neural network model that is ranged between first place as max low error and fifth place as the highest low error for Neural network.

Research future work is to test more models from classical time-series model and deep learning models after a couple of years to have more data to apply deep learning models and test proposed model on more data, as we explain this research is highly recommending the proposed model in case of non-stable environments which is the real-world environment and for stable non-changing environment the research recommends Neural Network. if resources available if it does not exist, this research is highly recommending proposed model in this case too.

Acknowledgments

We would like to express our appreciation to our professor Dr.Christine Albert for giving us valuable advice, invaluable contribution, Share of her time with us, her constructive criticism, her suggestions, her unwavering support, her guidance and share her experience as well as her patience that cannot be underestimated and her profound belief in our abilities during our Research.

References

[1] G. Khoury, “An Introduction to Customer Relationship Management” 2005.ResearchGate,pp.101-109.

[2] T. Bahari and M. S. Elayidom, “An efficient crm-data mining framework for the prediction of customer behaviour,” Procedia Computer Science, vol. 46, pp. 725–731, 2015.

[3] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “Statistical and machine learning forecasting methods: Concerns and ways forward,” PLOS ONE, vol. 13, no. 3, pp. 1–26, 03 2018.

[4] H. K. Chan, S. Xu, and X. Qi, “A comparison of time series methods for forecasting container throughput,” International Journal of Logistics Research and Applications, vol. 22, no. 3, pp. 294–303, 2019.

[5] P. J. Sheldon and T. Var, “Tourism forecasting: a review of empirical research,” Journal of Forecasting, vol. 4, no. 2, pp. 183–195, 1985.

[6] S. F. Witt and C. A. Witt, “Modeling and forecasting demand in tourism,” International Journal of Forecasting, vol. 8, no. 4, pp. 643–644, 1992.

[7] R. Abdel-Aal and A. Al-Garni, “Forecasting monthly electric energy consumption in eastern saudi arabia using univariate time-series analysis,” Energy, vol. 22, no. 11, pp. 1059–1069, 1997.

[8] H. Song, S. F. Witt, and T. C. Jensen, “Tourism forecasting: accuracy of alternative econometric models,” International Journal of Forecasting, vol. 19, no. 1, pp. 123 – 141, 2003.

[9] Y. Rashed, H. Meersman, E. Voorde, and T. Vanelslander, “Short-term forecast of container throughout: An arima-intervention model for the port of antwerp,” Maritime Economics Logistics, vol. 19, 03 2016.

[10] K. Wüllenweber, D.Beimborn, T.Weitzel and W. König, “The Impact of Process Standardization on Business Process Outsourcing Success”. Information Systems Frontiers, 2008, pp.211-224.

[11] R. Mushtaq, “Augmented dickey fuller test,” SSRN Electronic Journal, 08 2011,pp.1-19.

[12] E. Zivot and J. Wang, Vector Autoregressive Models for Multivariate Time Series, 01 2003, pp. 369–413.

[13] W. Scherrer and M. Deistler, Vector autoregressive moving average models, 01 2019,pp. 145-191.