



Full Research Paper

Application of Sales Forecasting Model Based on Machine Learning Algorithms

Abdullahi, Maimuna I.
Aimufua, Gilbert I.O.
Muhammad, Umar A.

Dept of Computer Science
Nasarawa State University
Keffi, Nasarawa State
Nigeria.



E-mail

aimufuagio@yahoo.com
Maimunaibrahim1105@gmail.com
umarfaruq54@gmail.com

Phone Nos

+2348066359657
+2348189890617
+2348032977816

ABSTRACT

Machine learning has been a subject undergoing intense study across many different industries and fortunately, companies are becoming gradually more aware of the various machine learning approaches to solve their problems. However, to fully harvest the potential of different machine learning models and to achieve efficient results, one needs to have a good understanding of the application of the models and the nature of data. This paper aims to investigate different approaches to obtain good results of the machine learning algorithms applied for a given forecasting task. To this end, the paper critically analyzes and investigate the applicability of machine learning algorithm in sales forecasting under dynamic conditions, develop a forecasting model based on the regression model, and evaluate the performance of four machine learning regression algorithms (Random Forest, Extreme Gradient Boosting, Support Vector Machine for Regression and Ensemble Model) using data set from Nigeria retail shops for sales forecasting based on performance matrices such as R-squared, Root Mean Square Error, Mean Absolute Error and Mean Absolute Percentage Error.

Keywords: Sales Forecasting, Model Based, Algorithms
Machine Learning

Proceedings Reference Format

Abdullahi, M.I., Aimufua, G.I.O. & Muhammad, U.A. (2021): Application of Sales Forecasting Model Based on Machine Learning Algorithms. Proceedings of the 28th iSTEAMS Intertertiary Multidisciplinary Conference. American International University West Africa, The Gambia. October, 2021. Pp 205-216
www.isteams.net/gambia2021.
DOI - [https://doi.org/ 10.22624/AIMS/iSTEAMS-2021/V28P17](https://doi.org/10.22624/AIMS/iSTEAMS-2021/V28P17)



1. INTRODUCTION

Earlier supermarkets used to order goods without considering the number of sales and demand. Modern supermarket requires past and present sales data before ordering for sales to profit (Mohit, 2017). In today's highly competitive environment and dynamic consumer landscape, accurate and timely forecasting of future revenue or sales can offer valuable insight to supermarkets' orders, planning, stock management, growth, and decision-making (Wacker & Lummus, 2018). Sales forecasting is particularly important because of the limited shelf-life of many household goods which leads to a loss of income in both shortage and surplus situations (Martne, 2019).

Most supermarket managers make sales predictions randomly by the rule of the thumb especially in developed countries like Computer-assisted predictions can be more accurate than expert predictions. To achieve a computer system that plays a better managerial role is to try and model the professional managers' skills in a computer program (Fallah & Ahrens, 2016). Using Machine Learning techniques with the available sales data and information, one can automatically develop accurate sales, forecasting models. This approach is simpler, more flexible (adapt to data change), and not prejudiced by a single sales manager's particularities (Qu *et. al.*, 2017).

Previously, forecasting models such as statistical models, time series, linear regression, feature engineering, and Random Forest Regressors were used to forecast sales but the complexity of predicting sales calls for a more intelligent model that can learn from previous data and automatic predictions (Wang *et. al.*, 2017). The challenge this research tries to address is how information about different items sold in different supermarkets in Nasarawa State can be used to forecast sales in new stores through the application of features engineered to propose an intelligent Machine Learning model that is more accurate than the traditional method.

2. LITERATURE SURVEY

Although researchers have tried modelling sales and demand forecasting systems much is still left not researched. Works of literature on a few sales forecasting systems are reviewed below. XGBoost is used for binary classification in a supervised machine learning system capable of matching internet devices to web cookies. XGBoost is used to improve the marketer's ability to identify individual users as they switch between devices and show relevant content/recommendations to users wherever they go. Support Vector Machine (SVM) had been applied to demand forecasting. Garcia *et al.* (2012), in their study, proposed an intelligent model that relies on supporting vector machines to deal with issues relating to the allocation and revelation of new models. Kandananond (2012) showed that SVM surpassed Artificial Neural Networks in estimating demand for consumer goods. Holmberg (2017) observed that regular restaurant sales be influenced by the weather. They considered two Machine Learning algorithms as XGBoost and neural network, and the results showed that the XGBoost algorithm is more accurate than the other algorithm, and they also found that they had improved their model performance by 2-4 percentage points by considering weather factors. To improve accuracy, they had considered numerous variables such as date characteristics, sales history, and weather factors.



Grigorios (2016) had used Machine Learning techniques to perform a survey on the forecasting of food sales. They had addressed data analyst design decisions such as temporal granularity, output variable, and input variables in this survey. In this paper, the authors experimented by taking the point of sale (POS) as internal data and even external data by considering different environments to enhance the efficiency of demand forecasting. They considered different Machine Learning algorithms such as Boosted Decision Tree Regression, Bayesian Linear Regression, and Decision Forest Regression for evaluation. In particular, Support Vector Machine (SVM) had been applied to demand forecasting. Garcia et. al., (2012), in their study, proposed an intelligent model that relies on supporting vector machines to deal with issues relating to the allocation and revelation of new models. Kandananond (2012) showed that SVM surpassed Artificial Neural Networks in estimating demand for consumer goods.

XGboost is being used in research areas of growing importance such as miRNA-disease association prediction, a problem that has been addressed using either Network Analysis or Machine Learning techniques (Xing et. al., 2018). Also applied in old but continuing problems such as bankruptcy prediction (Maciej et. al., 2018). In addition to other prediction problems such as crude oil price (Mesut and Mustafa, 2017), grid-connected photovoltaic plant production (Ferlito et. al., 2017), and as a part of a forecasting system for bird's migration (Benjamin & Kyle, 2018). Previously, most of the studies focused on considering the metrics as mean absolute error, mean squared error, median absolute error, and k-fold cross-validation is used for training and testing data. Metrics like max error, accuracy, and mean absolute error are considered in this research. In this study stratified, the K-fold cross-validation technique is used for training and testing to increase the efficiency of the results. In this study, a suitable algorithm is chosen for sales forecasting.

Ali and Vehmas (2016) incorporate weather data into retail sales forecasting. They predict daily sales of products based on time series forecast and a regression model accounting for the weather effects. They find that forecasting accuracy for weather-sensitive product categories can be improved by using this method. Fildes et. al., (2019) listed several machine learning and nonlinear regression models used for sales prediction. They opined that most published research has found improvements from using nonlinear methods rather than linear regressions but remark that this may well arise from publication bias. The models used for product sales prediction include back-propagation neural networks, fuzzy neural networks, regression trees, grey relation analysis and multilayer functional link networks, two-level switching models that select between a simple moving average and k-nearest neighbour or decision trees, support vector machines (Pillo et. al., 2016), wavelets (Michis, 2015), fuzzy set approach (Veiga et. al., 2016) and Bayesian P-splines (Lang et. al., 2015). Also, extreme learning machines are used in several studies.

Due to computational limitations, studies with nonlinear regression models often use only small datasets (i.e., datasets with tens of products). The exception to this is Mukherjee et. al., (2018), where authors develop a demand density forecasting model with a multi-layered hybrid deep neural network for inventory management in Flipkart.com. They model demand density using a mixture of Gaussian distributions and report significant improvement in prediction accuracy.

3. METHODOLOGY

The Agile model system development life cycle was adopted in this research where there is a continuous interaction of model development, implementation procedures, and testing with collected sales data which represent customers' expectations. The entire modelling process was divided into small incremental stages with time frames that represent sprint in the Agile model development life cycle. Each sprint in this research consists of the planning, requirement analysis, designing, building, and testing phase which lasted for four (4) weeks. After every development stage, the behaviour of data generated from input data is tested and compared with historical data and other standard methods to test data accuracy before evaluating the performance of machine learning algorithms. The research development life cycle is as shown in figure 1.

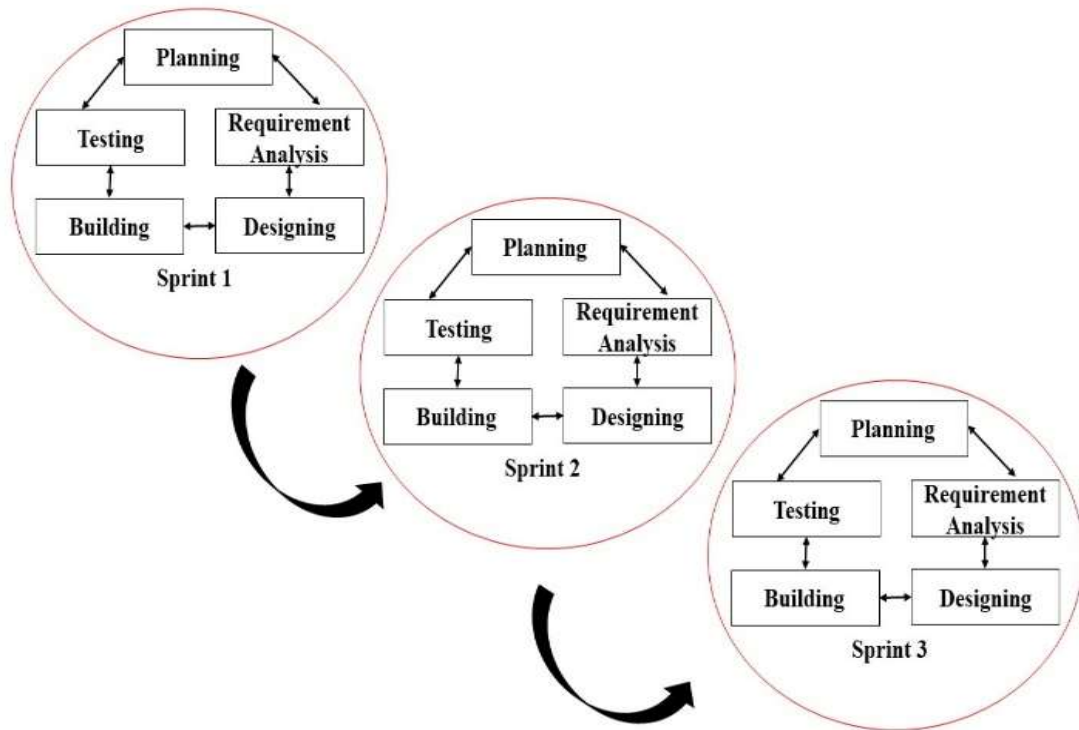


Fig 1: The research development life cycle

3.1 Population Sample of the Study

In this study, the sales data collected from the point-of-sale system of ten outlets of Bakan Gizo supermarket within Nasarawa State, Nigeria was used for the study. The supermarket outlets sell a very wide range of brands (more than 20,000 brands) of products ranging from beverages and frozen foods to household commodities and other Nigeria traditional items. This dataset includes 2018-2020 sales data and includes 2000 products distributed across the ten stores in Nasarawa states in Nigeria.



The data set consists of 8523 observations (instances or rows) and 12 attributes (variables or columns). Each of these observations represents a unique product/item that possesses a unique identification number. Also, each item has a value for all the 12 variables. These variables of the data set can be classified into two groups; those related to the item itself (8 variables), and the others are related to the store in which those items are exhibited (4 variables). The complete set of variables is as shown in table 1.

Table 1: List of Variables

S/No	Variables	Description	Feature Type
1.	Store identifier.	Unique identifier for an outlet	Categorical
2.	Store location Type.	Unique identifier for outlet location	Categorical
3.	Store Type.	Unique identifier for outlet type	Categorical
4.	Store Size.	Unique identifier for outlet size	Categorical
5.	Store establishment year.	Unique identifier for outlet year of establishment	Categorical
6.	Item identifier.	Unique identifier for the item	Categorical
7.	Item weight.	Unique identifier for item weight	Numeric
8.	Item fat content.	Unique identifier for item fat content	Categorical
9.	Item visibility.	Unique identifier for item visibility	Numeric
10.	Item type.	Unique identifier for the item type	Categorical
11.	Item MRP.	Unique identifier for item MRP	Numeric
12,	Item store sales	Unique identifier for item store sale	Numeric

4. RESULTS AND DISCUSSIONS

4.1 Data Analysis and Results

The resulting measurements of each model are presented in four separate tables where each table represents conditions under which the models were developed. Each table below shows the values associated with each performance evaluation metric (RMSE, R-square, MAPE, and MAE) for each one of the three models (Random Forest Regressor, Extreme Gradient Boosting model, Support Vector Machine Regression model). Each metric has two lines showing the results of validation and testing sets.

Table 2: Models performance – full set of features – tuning not applied

Parameter	Algorithm		
	RF	Xgboost	SVR
RMSE	1.9358	1.9632	1.9831
	1.94012	1.9446	1.9782
R-Square	0.70621	0.6981	0.6916
	0.70344	0.6910	0.7026
MAPE	13.79424	13.9871	14.0261
	13.46512	13.7326	13.7301
MAE	1.51725	1.5376	1.5182
	1.50772	1.5173	1.5171

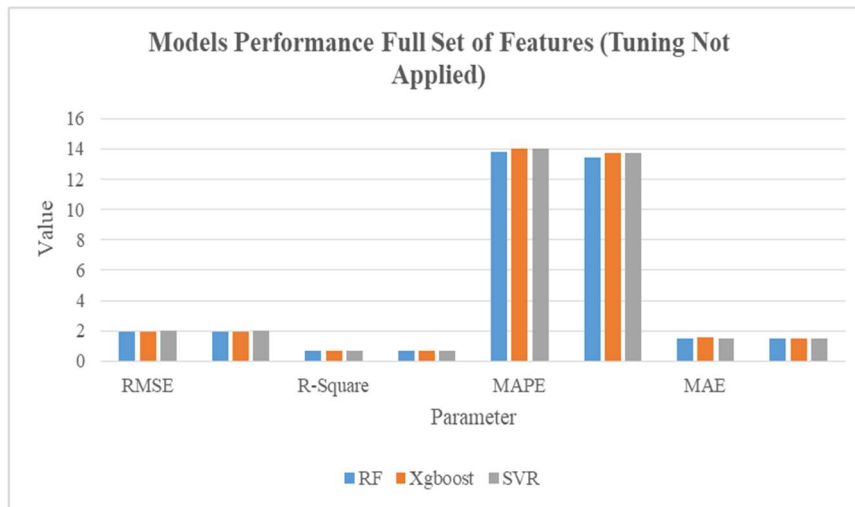


Figure 2: Models Performance Full Set of Features (Tuning Not Applied)

Table 4.1 and figure 4.1 show the result and performance of the models with their default hyperparameter and the basic configuration. From the RMSE perspective, the RF algorithm is associated with the lowest score of 1.935, while slightly higher values with the two models Xgboost and SVR. In respect to R2 scores, again RF best performed with a value of 0.706 which is slightly better than SVR's value of 0.702, but both of the algorithms show very close values. boost performance is slightly weaker with a value of 0.698. MAPE and MAE values are consistent with the ratio of error for all of the models; ranging between 13.456 per cent for RF to 14.026 per cent for the SVR model.

Table 3: Models performance – full set of features – tuning applied

Parameter	Algorithm		
	RF	xgboost	SVR
RMSE	1.9262	1.9331	1.9401
	1.9285	1.9451	1.9543
R-Square	0.7074	0.7054	0.7004
	0.70761	0.7034	0.70345
MAPE	13.8188	13.4431	13.8900
	13.5788	13.8566	13.6640
MAE	1.0505	1.4971	1.5029
	1.4911	1.4944	1.5046

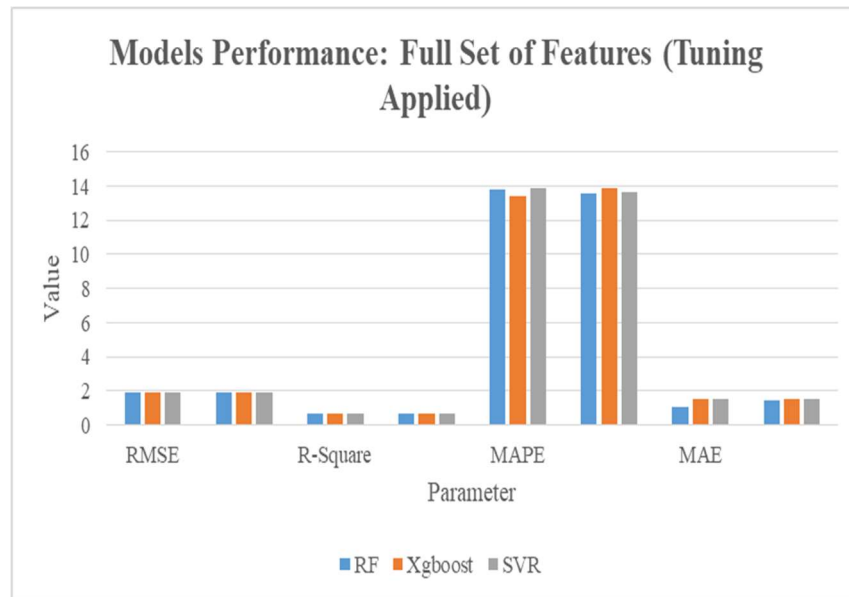


Figure 3: Models Performance: Full Set of Features (Tuning Applied)

From table 3 and figure 3, hyper-parameter optimization has also improved the performance of the three algorithms. With performing tuning optimization, the algorithms performed at the same level with an R² square of around 0.70.

Table 4: Models performance – the subset of features (16, the union)

Parameter	Algorithm		
	RF	Xgboost	SVR
RMSE	1.9188	1.9172	1.9079
	1.8687	1.8917	1.8587
R-Square	0.71503	0.7123	0.7123
	0.7307	0.7210	0.7210
MAPE	13.1181	13.1611	13.0712
	13.5822	13.6121	13.5454
MAE	1.4792	1.4816	1.4515
	1.4553	1.4881	1.4298

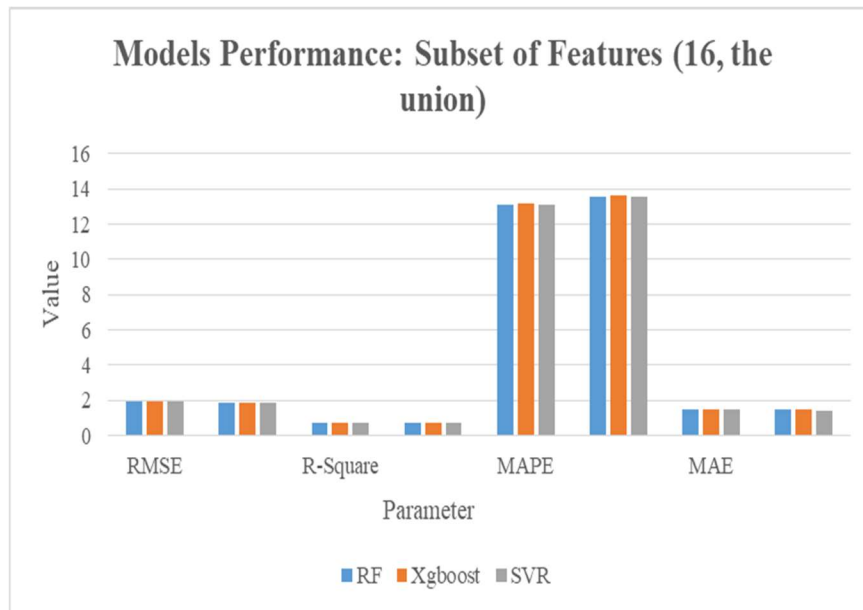


Figure 4: Models Performance: Subset of Features (16, the union)

Table 4.3 and figure 4.3 present the results associated with models' developments using the subset of features selected (union approach; using 16 features). Looking at the R^2 scores, it can be observed that SVR and Xgboost have almost identical performances with recorded values around 0.72, while the RF model performed better with an R^2 value of 0.73. Consequently, the RMSE score has been reduced best by the RF model to 1.858, while Xgboost and SVR have reduced the score to 1.868 and 1.891 respectively.

Table 5: Models performance – the subset of features (5, the intersection)

Parameter	Algorithm		
	RF	Xgboost	SVR
RMSE	1.8619	1.8784	1.8881
	1.9098	1.9239	1.9239
R-Square	0.7103	0.7102	0.7102
	0.7103	0.7102	0.7102
MAPE	13.5906	13.7652	13.5906
	13.2112	13.2685	13.2218
MAE	1.3561	1.4804	1.4851
	1.3578	1.4682	1.4697

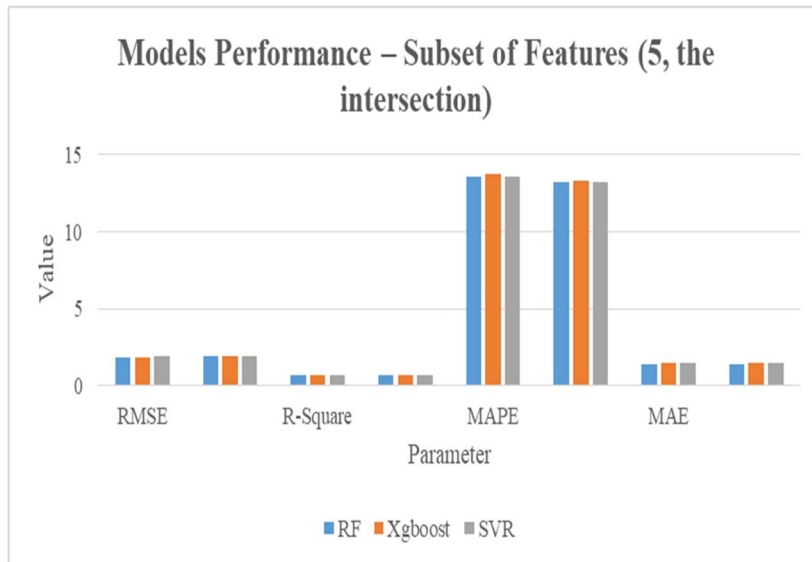


Figure 4.4: Models Performance – Subset of Features (5, the intersection)



Finally, table 5 and figure 5 present the results associated with the model's developments using the subset of features selected (intersection approach; using only five features). Comparing R2 values of the three models to the R2 values in Table 4.3, we can see that they are very close to each other; around 0.7102 for Xgboost and SVR and almost 0.7103 for RF

4.2 Discussion of Findings

In general, Random Forest, Xgboost, and Support Vector Machine have performed well and resulted in the best fitting on the data. To conclude on the performance of the algorithms, it is important to combine the evaluation of the performance with an understanding of the data itself. For example, with much larger samples and data sets, the algorithms can learn more from bigger training data.

Another important aspect regarding the data itself is the type of information recorded and used as predictors highly affects the performance of algorithms; when sales are the target for a prediction task, it would be of great benefit to record some other important features that would have a more direct relationship with the target variable. Those features are, for example, the number of customers or visitors on a given day. Another feature is whether or not a particular item is on sales or not, or more generally whether or not a discount policy is applied or offers are provided. Sales would be affected by these features and thus it worth for the supermarket to record such piece of more informative features.

Based on the results presented in the four tables, the fourth one includes the highest R2 values and the most reduced errors, and this has to do with what has been mentioned on the subject of whether or not the set of predictors are informative or not. Models could have achieved better performance using only a subset of features (16 features) out of the original complete set of 28 features. Feature selection impact was better than performing hyperparameter optimization, which is computationally very expensive, and the effect of ensemble learning, and also the effect of two of them combined (the last column of Table 4.2).

Not only limited to the union approach for feature selection but also the intersection approach has resulted in very close results to the union approach. The results of feature selection achieved an improvement of (3%) in the R2 value compared to the first configuration of the models, and by 2(%) compared to the tuned models. In other words, we could predict sales with a better level of accuracy using only five features.

In simple terms, Random Forest Regression Algorithm performed well after doing all the studies when compared with Extreme Gradient Boosting and Support Vector Machine algorithm. Hence, the Random Forest Regression Algorithm is considered the best suitable algorithm for forecasting product sales.



5. CONCLUSIONS

In this paper, we used three machine learning algorithms (Random Forest, Extreme Gradient Boosting, and Support Vector Machine algorithm) for sales forecasting, RF performed better, as it had a lower mean absolute error than the other two models. We also observed that getting more data would generally increase the predictive power of our models. Also, guarantee a world-free of ADE.

REFERENCES

1. Benjamin M.V., & Doren, K. G. (2018). A continental system for forecasting bird migration. *Journal of Science*, 361.
2. Fallah, T. A., & Ahrens, D. (2016). Enhanced predictive models for purchasing in the fashion field by using kernel machine regression equipped with ordinal logistic regression. *Journal of Retailing and Consumer Services*, 32, 131-138, <http://dio.org/10.1016/j.jretconser.2016.05.008>.
3. Ferlito, G. A. (2017). Comparative analysis of data-driven methods online and offline trained to the forecasting of grid-connected photovoltaic plant production. *Applied Energy*, 205.
4. Fildes, R. & Soopramanien, D. (2014). the value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research*, vol. 237(2): 738 -748, DIO: 10.1016/j.ejor.2014.02.022.
5. Karm J. P. & Maldonado, S. (2019). Hierarchical time series forecasting via Support rector Regression in the European travel retail industry. *Expert Systems with Applications*, 137, 5973, DIO: 10.1016
6. Lang, S., Steiner, W. J., Weber, A., & Wechselberger, P. (2015). Accommodating heterogeneity and nonlinearity in price effects for predicting brand sales and profits. *European Journal of Operational Research*, 246(1):232-241.
7. Martne, A., Schmck, S., Pereere, C., Pirker, M., & Haltmeier (2018). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*, <http://dio.org/10.1016/j.ejor.2018.04.034>.
8. Michis, A. A. (2015). A wavelet smoothing method to improve conditional sales forecasting. *Journal of the Operational Research Society*, 66(5):832-844.
9. Mohit G. (2017). Forecasting of sales by using the fusion of machine learning techniques. *International Conference on Data Management, Analytics and Innovation (ICDMAI)*, IEEE., 93-101.
10. Mukherjee, S., Shankar, D., Ghosh, A., Tathawadekar, N., Kompalli, P., Sarawagi, S., & Chaudhury, K. (2018). Associative and recurrent mixture density networks for retail demand forecasting.
11. Qu, T., Zhang, J. H., Chan, F. T. S., Srivastava, R. S., Tiwari, M. K. & Park, W.Y. (2017). Demand prediction and price optimization for the semi-luxury supermarket segment. *Computers and Industrial Engineering*, 113, 91-102. <http://dio.org/10.1016/j.cie.2017.09.004>
12. Rising, O. (2020) Applied machine learning for supermarket sales prediction. Retrieved from <https://www.researchgate.net/publication/338681895>



Proceedings of the 28th SMART-iSTEAMS
Interteritary Multidisciplinary Conference
American International University West Africa,
The Gambia

13. Valenzuela, O. (2016). Hybridization of intelligent techniques and ARIMA models for time series prediction, fuzzy sets and systems, 159(7): 821-845, <http://dio.org/10.1016/j.fss.2007.11.003>
14. Wacker, J. G. & Lummus, R. R. (2018). Sales forecasting for strategic resource planning. *International Journal of Operations & Production Management*, 22, 1014–1031
15. Xing C. L. H. (2018). EGBMMDA: Extreme gradient boosting machine for MiRNA-disease association prediction. *Journal of Cell Death and Disease*, 9.