Faculty of Natural and Applied Sciences Journal of Scientific Innovations
Print ISSN: 2814-0877 e-ISSN: 2814-0923
www.fnasjournals.com
Volume 5; Issue 2; December 2023; Page No. 30-37.

PREDICTIVE MODELING OF BUSINESS SUCCESS USING RANDOM FOREST, JRIP AND NAÏVE BAYES ALGORITHMS

[1]Taiwo, E. O., [2]Ogunsanwo, G.O. & [2]Alaba, O. B.
[1]Computer Science Department, Alvan Ikoku University of Education, Owerri.
[2]Computer Science Department, Tai Solarin University of Education, Ijagun

***Corresponding author email:** emmanuel.taiwo@alvanikoku.edu.ng

Abstract
The paper focused on the development of a predictive model for business success using machine learning algorithms. The model classifies and predicts business as either gain or loss. The historical dataset was collected for a period of ten (10) years (2012-2021) from Ogun State Chambers of Commerce, Industry Mines, and Agriculture Ltd/Gte. The dataset was further divided into 80% for training and 20% for testing. The metrics used for evaluation include: classification accuracy, execution time, error rate, ROC Area, mean absolute error (MAE), root mean squared error (RMSE) and confusion matrix. The dataset was used to formulate predictive models for business success using Random Forest, JRip and Naïve Bayes algorithms. Waikato Environment for Knowledge Analysis (WEKA) statistical tool was used to carry out the formulation and simulation of the predictive model. Results show Classification Accuracy (%) of 60.9, 63.9, 68.4. Execution Time (Seconds) of 0.68, 0.03, 0.1. Error Rate (%) of 39.0, 30.6, 31.6. ROC Area of 0.504, 0.509. 0.488. Mean Absolute Error (MAE) of 0.1715, 0.1683, 0.1717. Root Mean Squared Error (RMSE) of 0.3298, 0.2938, 0.2953 for Random Forest, JRip and Naïve Bayes algorithms respectively. The three models were compared and the best model in terms of accuracy and ROC Area was selected and validated. The study revealed that Naïve Bayes model has higher accuracy followed by JRip and Random Forest algorithms. The model is recommended for Business evaluation and any other machine learning algorithms can be used for business success predictive model.

Keywords: Predictive Modeling, Business Success, Machine Learning, Algorithms

Introduction

The growth of any country depends solely on the business sector of the economy. Most of the small and medium businesses in Nigeria find it difficult to succeed due to some cogent factors. Business involves the creation and promotion of goods and services. Singh et al. (2018) defined business as an activity that generates money based on the production, buying and selling of goods and services. Business plays a major role in employment creation, poverty reduction, and national development. The business sector also contributes a larger percentage to the gross domestic product (GDP). The majority of businesses fail as a result of poor business planning, lack of managerial skill, inadequate capital, location of business, poor management knowledge, etc.

Machine Learning (ML) is a sub-field of Artificial Intelligence that permits computers to ponder and analyze on their own (Alzubi et al., 2018). Ravi et al. (2021) described ML as the utilization of artificial intelligence wherein a computer learns from the input data and makes predictions. According to Alqudah and Yaseen (2020), ML means that without being programmed computer is capable of bringing out a solution i.e. machines can learn consistently and address large datasets with the use of classifiers and algorithms. The fundamental support of ML is classifiers that categorize observations even as algorithms construct models of behaviours and based on new input data make use of them for predictions (Wang et al., 2019). On the part of the machine, ML could be used to resolve diverse problems that require learning. Therefore, ML solutions are data-driven and based on the data fed to the model which uses algorithms to forecast expected results (Andrew & Parvathi, 2020). Predictive modelling is a system utilized in predictive analytics to create a statistical model of future behaviour. A predictive model is made from variables which are predictors that can influence future results or behaviour. Data are collected in predictive modelling for the relevant predictors. Then, a statistical model is formulated, predictions are made and the model is validated with historical data. There are three features of learning problems. These include tasks that must be learnt,

Cite this article as:
Taiwo, E. O., Ogunsanwo, G.O. & Alaba, O. B. (2023). Predictive modelling of business success using random forest, jrip and naïve Bayes algorithms. *FNAS Journal of Scientific Innovations, 5*(2), 30-37.

the process of gaining experience and performance measures to be enhanced. Existing literature shows that a wide range of methods have been used for business success prediction. However, most of the studies focused on the model with higher percentage accuracy evaluation parameters. In this study, we predict business success and various evaluation parameters were used.

Kaneko et al. (2017) developed a model to identify the relationship between sales and the movement of in-store customers. They utilized the Bayesian algorithm to build the model. The result showed that the model performed better in terms of accuracy. Lu (2014) proposed a model to predict computer product sales details. He used the support vector regression (SVR) algorithm to construct the model. The experimental result indicated that the model gives greater accuracy. Clark and Ravazzolo (2015) developed a model to forecast macroeconomics. They used Bayesian autoregressive and vector autoregressive algorithms to construct models and their performance was evaluated with the time-varying volatility. Fan et al. (2017) proposed a model from online reviews to predict product sales. They employed the Bass model and sentiment analysis to predict the result. Results showed that the model achieved better performance. Schneider and Gupta (2016) worked on sales of existing and new products. They used consumer records to predict sales of any product. The method yielded much-needed results.

Yu et al. (2013) proposed a model to predict sales of newspapers/magazines. They utilized a support vector regression algorithm to construct the model. Results indicated that in terms of accuracy, the model performed better than the conventional method. Singh et al. (2017) developed a model to predict sales data. They used regression algorithms to build models. The result showed that the model achieved higher accuracy. Choi et al. (2014) proposed a system to predict the sales data. They employed intelligent algorithms to build models. The experimental result indicated that the model performed better in terms of accuracy. Islam and Habib (2015) worked on data mining techniques to forecast business sectors' prospective. To validate the findings in search of a consistent pattern, the system uses data mining and customer transactional-related data methods. Tomy and Pardede (2018) proposed a model to forecast success in technological entrepreneurship. They utilized support vector machine (SVM), k-nearest neighbours (k-NN), and naïve Bayes algorithms to construct the model. Results showed that naïve Bayes outperformed the other two algorithms.

Aim and Objectives of the study

The aim of this study is to develop a model for predicting business success in Nigeria. While the specific objectives are to elicit variables causing business success in Nigeria, formulate model for predicting business success based on the variables identified, simulate the model and validate the model.

Methodology

Seven hundred and fifty (750) datasets of small and medium businesses were collected from Ogun State Chambers of Commerce, Industry Mines, and Agriculture Ltd/Gte for a period of ten (10) years (2012-2021). The data collected included information such as Name, Nationality, Address, Gender, Age, City, State, Marital Status, Business Category, Established Year, and Business Classification. From paper-based storage, the data were converted into electronic format and stored as Microsoft Excel files. Attribute selection processes were performed on the processed data to identify five major crucial input variables: Age, Gender, Business Category, and City. Class is the "business" with two options of "gain" or "loss".

Random Forest, JRip and Naïve Bayes supervised learning algorithms were used. The research focused on classification wherein the output of prediction is already known to either be gain or loss. In developing the predictive models, Waikato Environment for Knowledge Analysis (WEKA) statistical tool was used. The dataset that contains 750 business owners was used to develop a predictive model. The percentage splits of 80% training and 20% testing were used for the prediction. Each model was compared and based on evaluation criteria and results, the most efficient model was chosen.

Cite this article as:
Taiwo, E. O., Ogunsanwo, G.O. & Alaba, O. B. (2023). Predictive modelling of business success using random forest, jrip and naïve Bayes algorithms. *FNAS Journal of Scientific Innovations, 5*(2), 30-37.

Table 1: Dataset

S/N	Variable	Description	Measurement
1	Age	Business owner Age	Numeric
2	Gender	Male or female	Nominal
3	Business Category	Small or medium	Nominal
4	City	Business location city	Nominal
5	Class	Class of business gain/loss	Nominal

Results

The results of the business success model developed for three machine learning algorithms in this paper are shown in Figure 1-6. The model evaluations are also shown in Table 2-5.

The accuracy of the Random Forest Algorithm is 60.9%. It has an execution time and error rate (%) of 0.68 secs and 39.0% respectively. It also has ROC Area, MAE and RMSE of 0.504, 0.1715, 0.3298 as shown in Figure 1

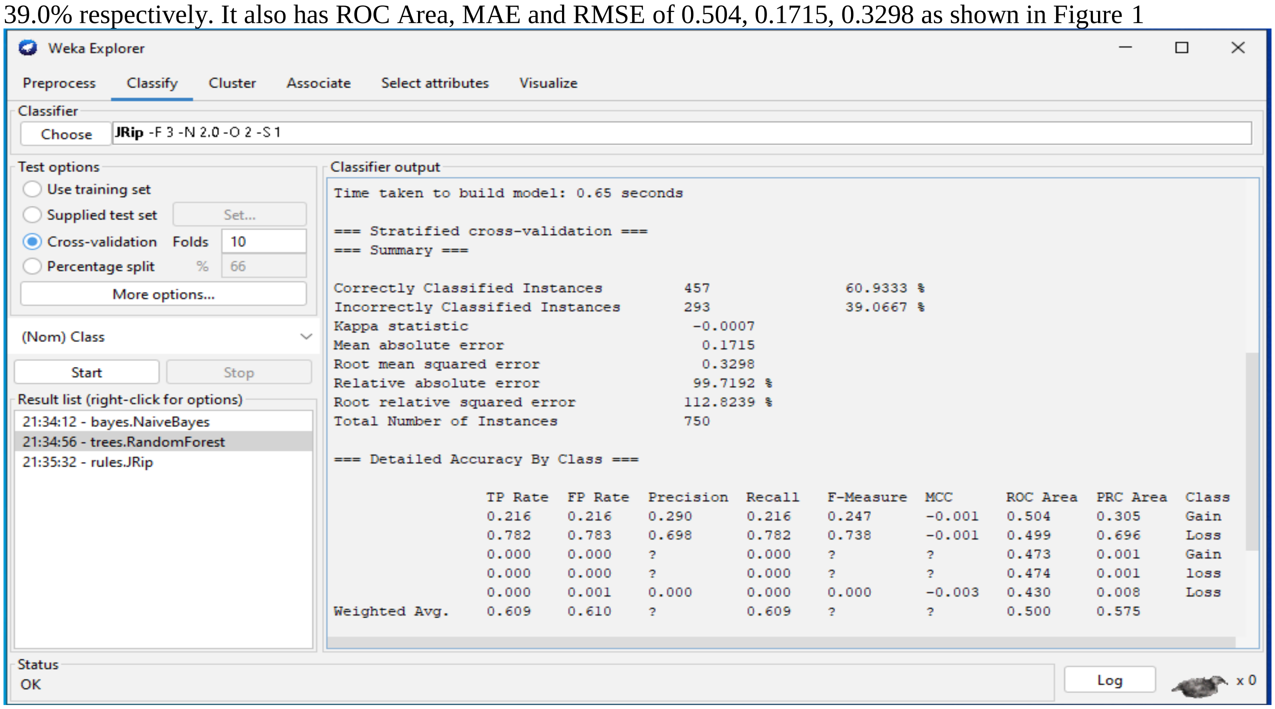


Figure 1: Random Forest Algorithm Model

```
=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
  47 171   0   0   0 |   a = Gain
 113 410   0   0   1 |   b = Loss
   1   0   0   0   0 |   c = Gain
   0   1   0   0   0 |   d = loss
   1   5   0   0   0 |   e = Loss
```

Figure 2: Confusion Matrix for Random Forest Algorithm Model

Cite this article as:
Taiwo, E. O., Ogunsanwo, G.O. & Alaba, O. B. (2023). Predictive modelling of business success using random forest, jrip and naïve Bayes algorithms. *FNAS Journal of Scientific Innovations, 5*(2), 30-37.

Table 2: Results obtained from the Random Forest Algorithm

Number of instances	750
Correctly Classified Instances	457
Incorrectly Classified Instances	293
Classification Accuracy (%)	60.9
Execution Time (Seconds)	0.68
Error Rate (%)	39.0
ROC Area	0.504
MAE	0.1715
RMSE	0.3298
Kappa Statistics	-0.0007

The accuracy of Naïve Bayes Algorithm is 69.3%. It has an execution time and error rate (%) of 0.03 secs and 30.6% respectively. It also has ROC Area, MAE and RMSE of 0.549, 0.1683, 0.2938 as shown in Figure 3

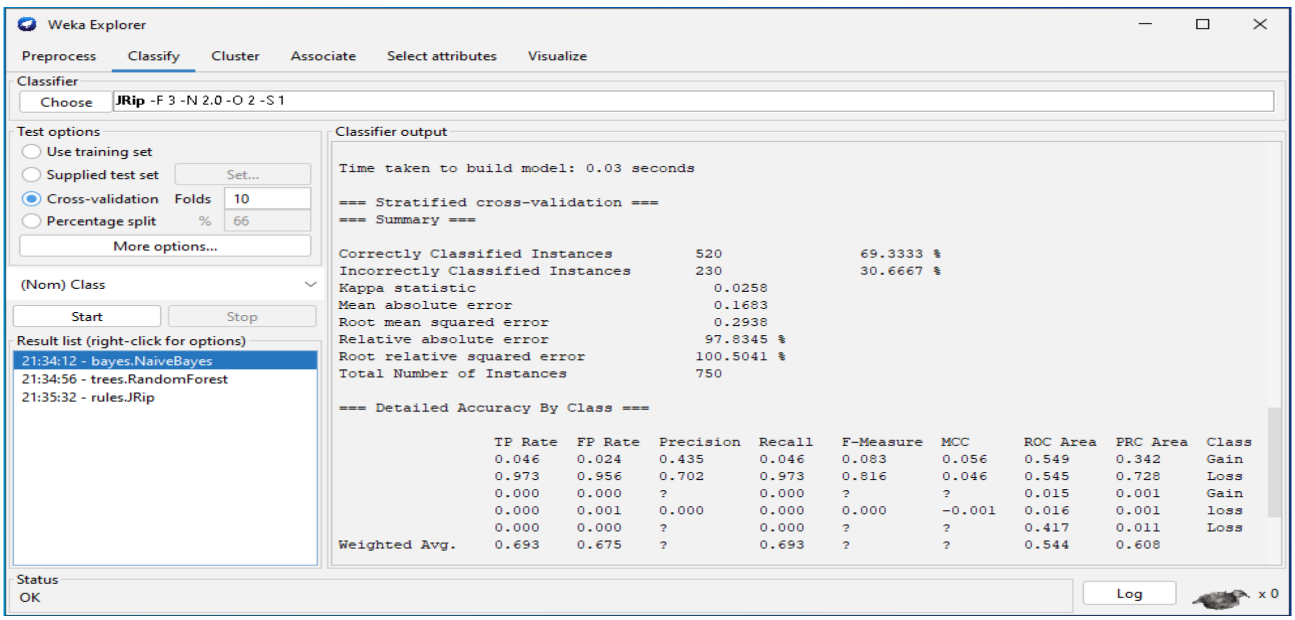
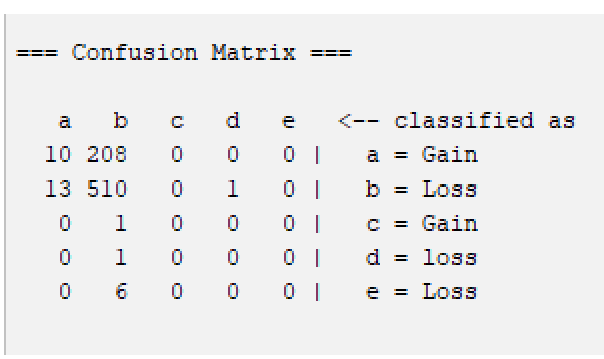


Figure 3: Naïve Bayes Algorithm Model



=== Confusion Matrix ===

a	b	c	d	e	<-- classified as
10	208	0	0	0	a = Gain
13	510	0	1	0	b = Loss
0	1	0	0	0	c = Gain
0	1	0	0	0	d = loss
0	6	0	0	0	e = Loss

Figure 4: Confusion Matrix for Naïve Bayes Algorithm Model

Cite this article as:
Taiwo, E. O., Ogunsanwo, G.O. & Alaba, O. B. (2023). Predictive modelling of business success using random forest, jrip and naïve Bayes algorithms. *FNAS Journal of Scientific Innovations, 5*(2), 30-37.

Table 3: Results obtained from Naïve Bayes Algorithm

Number of instances	750
Correctly Classified Instances	520
Incorrectly Classified Instances	230
Classification Accuracy (%)	69.3
Execution Time (Seconds)	0.03
Error Rate (%)	30.6
ROC Area	0.549
MAE	0.1683
RMSE	0.2938
Kappa Statistics	0.0258

The accuracy of the JRip Algorithm is 68.4%. It has an execution time and error rate (%) of 0.1 secs and 31.6% respectively. It also has ROC Area, MAE and RMSE of 0.488, 0.1717, 0.2953 as shown in Figure 5

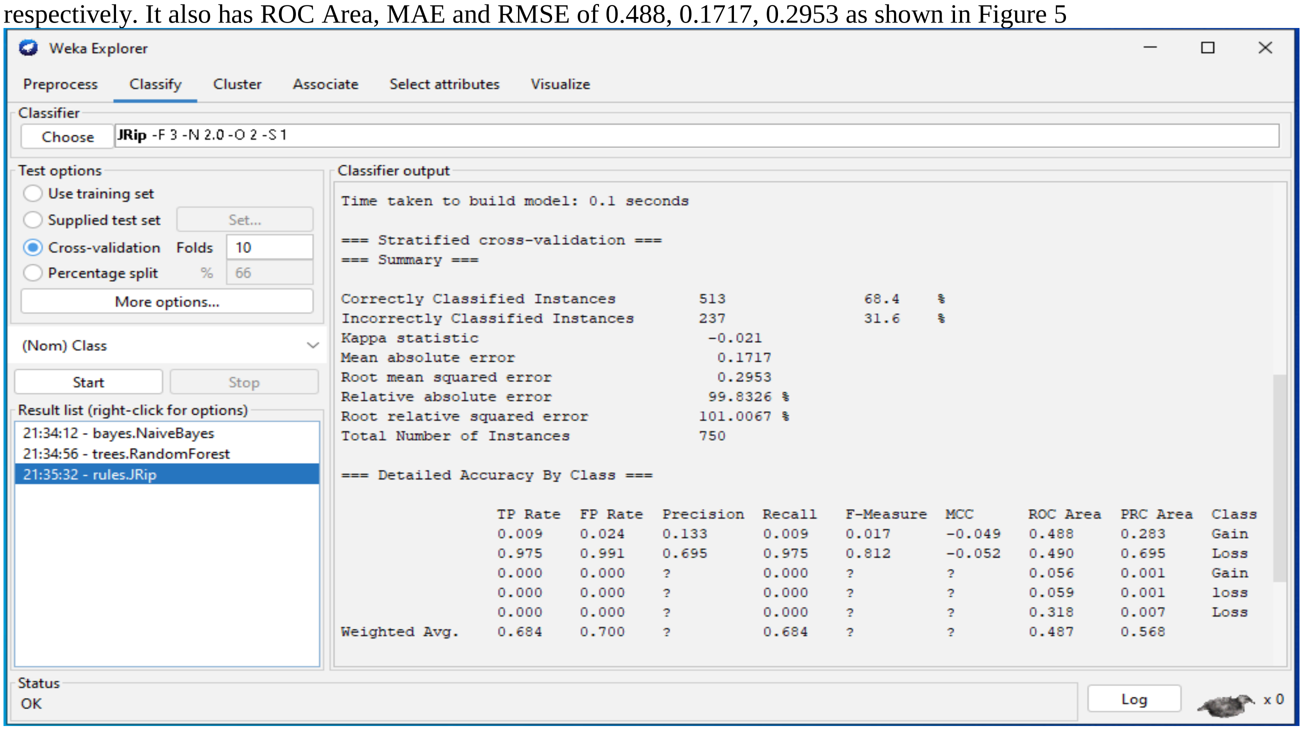


Figure 5: JRip Algorithm Model

```
=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
   2 216   0   0   0 |   a = Gain
  13 511   0   0   0 |   b = Loss
   0   1   0   0   0 |   c = Gain
   0   1   0   0   0 |   d = loss
   0   6   0   0   0 |   e = Loss
```

Figure 6: Confusion Matrix for JRip Algorithm Model

Table 4: Results obtained from the JRip Algorithm

Number of instances	750
Correctly Classified Instances	513
Incorrectly Classified Instances	237
Classification Accuracy (%)	68.4
Execution Time (Seconds)	0.1
Error Rate (%)	31.6
ROC Area	0.488
MAE	0.1717
RMSE	0.2953
Kappa Statistics	-0.021

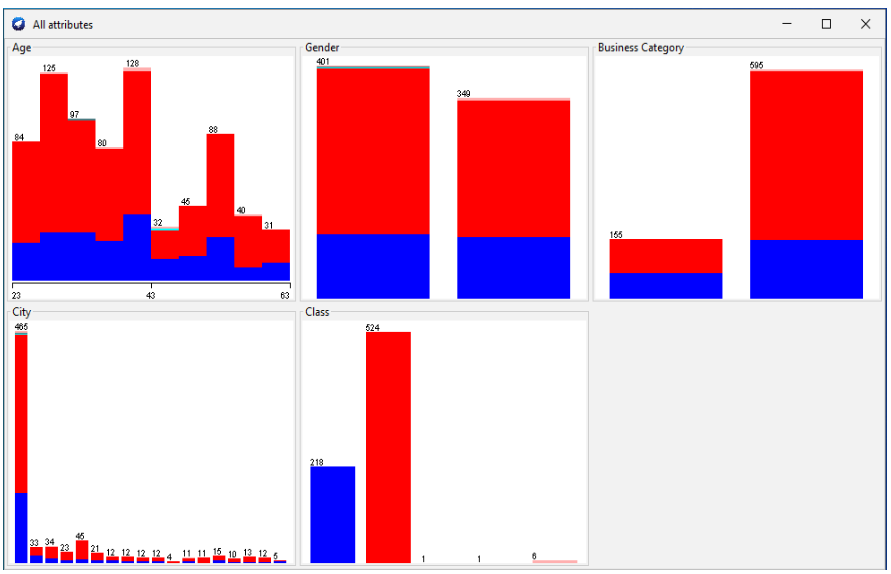


Figure 7: Data Visualization

Table 5: Results from the three Algorithms

Business Dataset	Random Forest Algorithm 80% 20%	Naïve Bayes Algorithm 80% 20%	JRip Algorithm 80% 20%
Number of instances	750	750	750
Correctly Classified Instances	457	520	513
Incorrectly Classified Instances	293	230	237
Classification Accuracy (%)	60.9	69.3	68.4
Execution Time (Seconds)	0.68	0.03	0.1
Error Rate (%)	39.0	30.6	31.6
ROC Area	0.504	0.549	0.488
MAE	0.1715	0.1683	0.1717
RMSE	0.3298	0.2938	0.2953
Kappa Statistics	-0.0007	0.0258	-0.021

Discussion

The accuracy of the Naïve Bayes algorithm is higher compared to the Random Forest algorithm and JRip algorithm for the business dataset split into 80% -20%. The execution time of the JRip algorithm is faster compared to the other two algorithms as shown in Table 5. Concerning the error rate, the Random Forest algorithm has a higher percentage of recorded errors compared to the Naïve Bayes algorithm and JRip algorithm as shown in Table 5. The Kappa statistic of Naïve Bayes is 0.0258 which is higher compared to the Random Forest algorithm which is -0.0007 and -0.021 for the JRip algorithm. The MAE is 0.1715, 0.1683, and 0.1717 for the Random Forest algorithm, Naive Bayes and JRip respectively. The RMSE is 0.3298, 0.2938, and 0.2953 for the Random Forest algorithm, Naive Bayes and JRip respectively. The ROC Area is 0.504, 0.549, and 0.488 for the Random Forest algorithm, Naive Bayes and JRip respectively.

Conclusion

In this paper, we build a model that could be used to classify and predict business as either gain or loss using three machine learning algorithms (Random Forest, Naïve Bayes and JRip). The business dataset was split into 80% training and 20% testing. The predictive model was implemented on the WEKA statistical tool. The Naïve Bayes algorithm recorded the highest prediction accuracy followed by JRip and Random Forest respectively. The model is recommended for Business evaluation and any other machine learning algorithms can be used for business success predictive model.

References

Alqudah, N., & Yaseen, Q. (2020). Machine Learning for Traffic Analysis: A Review. *International Workshop on Data-Driven Security (DDS 2020)* April 6-9, 2020, Warsaw, Poland. Procedia Computer Science 170 (2020) 911–916. Available online at www.sciencedirect.com

Andrew, M., & Parvathi, R. (2020). Vehicular Traffic analysis and prediction using Machine learning algorithms. *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE).* https://doi./10.1109@ic-ETITE47903.2020.279

Choi, T. M., Hui, C. L., Liu, N., Ng, S. F. & Yu, Y. (2014). Fast fashion sales forecasting with limited data and time. *Decision Support Systems*, 59, 84-92.

Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, *30(*4), 551-575.

Fan, Z. P., Che, Y. J., & Chen, Z. Y. (2017). Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis. *Journal of Business Research*, *74*, 90-100.

Islam, M. R., & Habib, A. (2015). Data mining approach to predict prospective business sectors for sending in retail banking using decision tree. IBM SPSS predictive analytics: Optimizing decision at the point impact.

Alzubi, J., Nayyar, A., & Kumar, A. (2018, November). Machine learning from theory to algorithms: an overview. In *Journal of physics: conference series* (Vol. 1142, p. 012012). IOP Publishing.

Kaneko, Y., Miyazaki, S. & Yada, K. (2017). The Influence of Customer Movement between Sales Areas on Sales Amount: A Dynamic Bayesian Model of the In-store Customer Movement and Sales Relationship. *Procedia Computer Science,* 112, 1845- 1854.

Lu, C. J. (2014). Sales forecasting of computer products based on variable selection scheme and support vector regression. *Neurocomputing, 128*, 491-499.

Ravi, A., Nandhini, R., Bhuvaneshwari, K., Divya, J. & Janani, K. (2021). Traffic management system using machine learning algorithm. *International Journal of Innovative Research in Technology (IJIRT), 7*(11), 303-308. ISSN: 2349-6002.

Schneider, M. J., & Gupta, S. (2016). Forecasting sales of new and existing products using consumer reviews: A random projections approach. *International Journal of Forecasting, 32*(2), 243-256.

Singh, D. A., Leavline, E. J., Muthukrishnan, S. & Yuvaraj, R. (2018). Machine learning-based business forecasting. *I.J. Information Engineering and Electronic Business*, *6,* 40-51. doi: 10.5815/ijieeb.2018.06.05

Singh, D. A., Leavline, E. J., Muthukrishnan, S., & Yuvaraj, R. (2017). Regression based sales data forecasting for predicting the business performance. *International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRSCE), 3*(11), 589-593.

Tomy, S., & Pardede, E. (2018). From uncertainties to successful start ups: A data analytic approach to predict success in technological entrepreneurship. *Sustainability*, *10*(3), 602.

Wang, P., Li, Y., & Reddy, C. K. (2019). Machine learning for survival analysis: A survey. ACM Computing Surveys (CSUR) 51(6), 110.

Yu, X., Qi, Z., & Zhao, Y. (2013). Support vector regression for newspaper/magazine sales forecasting. *Procedia Computer Science, 17,* 1055-1062.