Nigerian Journal of Science and Environment 2024 Volume 22 (1) 57 – 67 ISSN: 1119 – 9008
https://doi.org/10.61448/njse221245

APPLICATION OF RFM MODEL ON CUSTOMER SEGMENTATION IN DIGITAL MARKETING

Akazue Maureen[1], Esiri Kesiena Henry*[2] and Clive Asuai[3]

[1,3]Department of Computer Science, Faculty of Science, Delta State University Abraka

[2] Department of Computer Science & Information Technology, Petroleum Training Institute, Effurun

*Corresponding author: esiri_kh@pti.edu.ng

ABSTRACT

Customers play a pivotal role in the success of any business. The ability to attract and retain the right clientele, who consistently engage with a company's products and services, hinges on a thorough understanding of their purchasing behavior. Successful businesses tailor their offerings to meet the unique requirements and preferences of their customers. Utilizing marketing analysis tools, such as the RFM model, facilitates the segmentation of customers based on distinct parameters, enabling the identification of high-value customers through factors like purchase behavior. In addressing this critical aspect of customer-centric strategies, our research introduces an innovative early purchase prediction framework. This framework aims to assess the likelihood of a potential customer purchasing soon. Leveraging a combination of a decision tree classifier and a gradient-enhancing classification approach, we tackled the classification challenge using a dataset obtained from statistadata.com, focusing on digital market dynamics. Our findings reveal that the gradient-enhancing classifier model outperformed others, demonstrating an impressive accuracy of 93% and an AUC (Area Under the Curve) of 0.98. This research not only contributes to the advancement of predictive modeling in the realm of customer behavior but also underscores the practical application of such models in digital marketing strategies. As businesses navigate the intricacies of customer engagement, our proposed framework offers valuable insights to enhance decision-making processes and optimize marketing efforts.

Keywords: RFM, Digital Marketing, Customer Segmentation, Customer Behaviour Prediction

INTRODUCTION

In recent years, the field of computing and information technologies has proven to be an invaluable resource across various domains due to its transformative and beneficial impact in enhancing and simplifying digital activities (Ojie *et al.*,2023). The progression of technology aims to elevate society to greater levels of sophistication with enhanced convenience (Ojugo *et al.*,2023). In the business world, marketers understand that individual customer needs and wants do vary. Hence, organizations deploy various criteria and methods to segment customer groups to better identify and understand customer uniqueness and provide products and services to them in order of preference and consequently satisfy their different needs and wants (Dogan *et al.*,2018).

The business environment is highly competitive and exponentially increases the value of customers to companies. Creating a sense of confidence within the e-commerce transaction ecosystem is essential to encourage individuals to embrace this platform for their financial interactions (Akazue *et al.*,2015). Thus, providing precise and practical service is important because companies understand that, retaining their existing customers is easier than attempting to seek and secure potential ones (Cuce *et al.*,2022). In this context, the RFM model, which is frequently used in segmentation and targeting, is a method that uses the parameters recency, frequency, and monetary to classify customers according to the date they last made a purchase, the frequency of their purchases, and the amount of expenditure they made (Cuce *et al.*,2022). The RFM variables are suitable for identifying particulars of customer purchase behaviors and can enhance customer segmentation. The benefit to the organization is that, by using a small number of parameters, services can be personalized and customers are most likely to respond to adverts and promotions identified (Kadir *et al.*,2019).

By analyzing data and using predictive models, business organizations that can successfully segment their customers into groups of unique identities and characteristics can predict customer behavior and adopt the right marketing approaches to retain them. Segmenting clients, or the act of breaking consumers up into homogenous and discrete groups, is seen to be an efficient way to manage customers while creating a variety of marketing techniques (Kadir *et al.*,2019).

Segments of consumers are created using the RFM approach. It classifies customers according to their past purchases, taking into account factors such Recency (R): Date of the final purchase made during a given session. Purchase count during the designated session is the frequency (F). Monetary (M): Purchase value within the designated period (Chavhan *et al.*,2022).

Marketers use a consumer buying model to study and unearth the different actions engaged in by customers from start to finish. There are five steps in this process: identifying the need, gathering information, weighing your options, making a choice, and acting after the purchase (Kotmi *et al.*,2019). According to (Schiffman *et al.*,2007) knowledge of customer behavior better equips businesses to understand and predict purchase behavior, and the drivers of such purchase behavior. A new age of communication strategy which allows for businesses to engage with their customers, and conduct market research by observing customers’ behavior and preferences over time online, has been established by digital marketing (ElisabetaIoanăs

2014). It is the platform where the launch and execution of marketing activities are driven using the internet and information technology (Füsunçizmeci 2015).

Digital marketing and e-marketing have been used interchangeably and described as vendors and buyers exchanging products and services via electronic processes and devices. Further, (Ponde *et al.*,2019) opines that e-marketing in this context is based on using electronic innovations to establish a medium of communication between organizations and existing or potential customers. In the realm of predictive modeling for customer segmentation and behavior analysis, the utilization of ensembles has become a pivotal strategy. Ensembles, comprising classes such as bagging and boosting, offer a robust approach by combining the strengths of multiple models to improve overall predictive performance. This choice of ensembles over single heuristics is motivated by their ability to handle diverse and complex patterns inherent in customer data. In this study, we specifically focus on the adoption of XGBoost and Decision trees within the ensemble framework. XGBoost, known for its efficiency and scalability, and Decision trees, valued for their interpretability, are selected for their complementary strengths in enhancing the accuracy and interpretability of customer behavior predictions.

MATERIALS AND METHODS

The methodology for this study involves applying ensemble learning techniques, namely Decision Tree and XGBoost, to improve the precision and comprehensibility of behaviour analysis and consumer segmentation. The XGBoost method, which is renowned for its efficiency and scalability, and the Decision Tree algorithm, which was selected for its interpretive powers and simplicity, are applied to the dataset in their respective implementations. This methodical process guarantees that the characteristics that are chosen complement the strengths of the ensemble models, which helps to provide accurate and comprehensible predictions in later stages of analysis.

In this study we looked at several models with diverse architectural designs in order to accurately capture the resulting feature matrices and outcomes. The customer purchase classification problem is resolved using the Decision Tree Classifier and Gradient Boosting Classifier models. A subset of the best characteristics from our dataset (Akazue *et al.*,2023, Ojie *et al.*,2023a) must be chosen, and features must then be added to our model in order to divide consumers into groups according to how valuable they are to the business. We used the RFM segmentation approach in the execution of this. RFM is an acronym denoting:

Nigerian Journal of Science and Environment 2024 Volume 22 (1) 57 – 67 ISSN: 1119 – 9008
https://doi.org/10.61448/njse221245

i. Recency: a measure of a customer's most recent acquisition.

ii. Frequency: The quantity or regularity of a client's transactions.

iii. Monetary Value/Revenue: The total sum of money a consumer pays at a certain moment on something they bought.

The study's dataset was obtained from http://www.statistadata.com

Recency, frequency, and monetary value/revenue are the three variables deployed to group the customer using an RFM score system. Essentially, the RFM score is used to help provide light on the likely future purchases that a consumer might make.

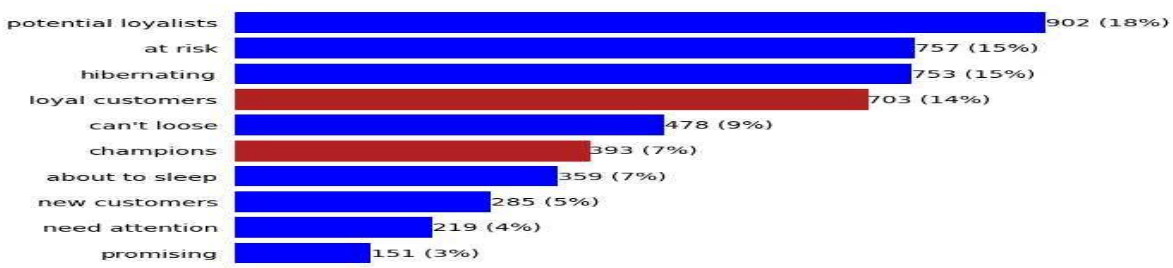

Figure 1. Market segmentation distribution of customers

From the figure depicted above we can notice the following:

Potential Loyalists: recent clients who have made many purchases and spend a significant sum of money.

i. At risk: made large purchases more frequently, but not in a lengthy period of time.

ii. Hibernating: These are thrifty consumers who haven't made many purchases.

iii. Can't lose: Recently made the largest purchases, yet it's been a while since I made a buy.

iv. Loyal consumers are attentive to promotions and make large purchases.

v. Champions: the most expensive, often purchased, and newly acquired.

vi. Regarding sleep: less than average in terms of frequency, recency, and monetary worth.

vii. New client: came in lately but fails to make frequent purchases.

viii. Attention Needed: higher than normal frequency, monetary worth, and recency.

ix. Promising: recent buyers who haven't made big purchases.

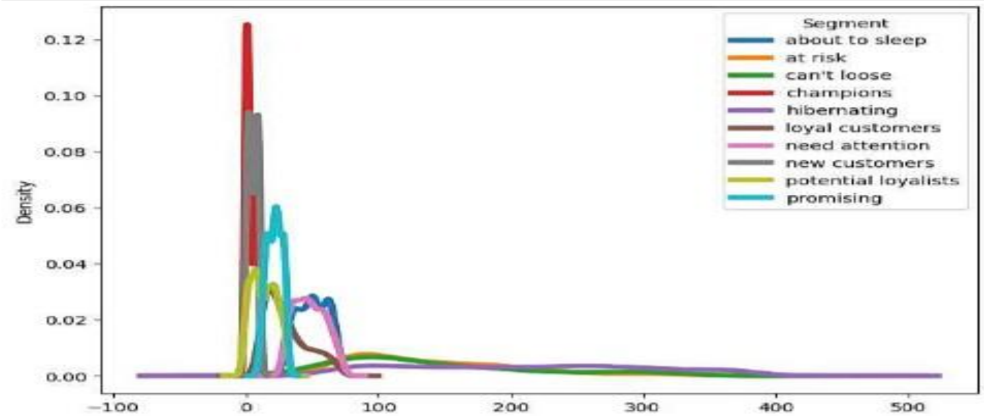

Figure 2. Destiny plot consumers according to spending habits

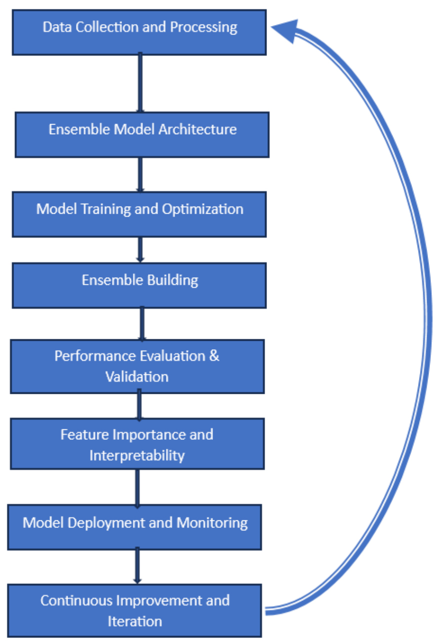

Figure 3. **Ensemble model Development for Customer Digital Marketing Behaviour Classification**

In this study, we aimed at classifying whether a consumer will make a purchase within a time frame. The table below gives the description.

Table 1. Description of the timeframe for a customer purchase decision

Next purchase (Binary)	Description
0	Customers who will purchase within 90 days
1	Customers who will purchase in more than 90 days.

The basic workflow for this study entails four key steps: data preparation, dividing the dataset into train and test sets, developing and training the models, and finally comparing the various models to see which one performs best and has the highest accuracy.

The purchase classification problem is resolved using the following models:

Decision Tree Classifier

Regression as well as classification applications can both benefit from the non-parametric supervised learning strategy of decision trees (Akazue *et al.*,2023e). It is organised like an organisation tree, with internalized, leaf, branch, and root nodes.

Gradient Boosting Classifier

A family of ensemble machine-learning methods called gradient boosting can be used for problems involving regression or classification predictive modelling (Akazue et al, 2023) . Gradient boosting is also known as gradient tree boosting, gradient boosting machines, or GBM, and stochastic gradient boosting (an extension).

Performance metrics

A few measures were examined in order to gauge the 2 model's efficacy. The model analysis was aided by these measures, which also revealed how effectively the selected machine learning algorithm predicts the future. For the measures, the following definitions apply:

i. Accuracy - Rate of data instances correctly classified by a model.
ii. Precision is calculated by dividing the total number of projected occurrences by the rate of successfully identified cases (Akazue *et al.*,2023).
iii. Recall: The percentage of cases properly categorized divided by the total number of cases.
iv. The accuracy of a model on a dataset is measured by the F1-score.
v. Sensitivity is a metric used to assess a machine learning model's ability to identify good examples.
vi. Specificity: the percentage of true negatives that the model accurately predicts.

RESULTS AND DISCUSSION

Figs. 3 and 4 show the prediction results of each classifierbased model as an area under the ROC curve (AUC) and receiver operating characteristic (ROC). Area under the ROC Curve, or AUC. An overall performance metric across all potential categorization criteria is provided by AUC; stands for the level or measurement of separability, and ROC is a probability curve. The evaluation outcomes of the used models are displayed below in terms of AUC scores. The Gradient Boosting Classifier, which has an AUC value of 0.98, is the best machine learning model, according to Figures 4.0 and 4.1, as well as the ROC curves shown below.

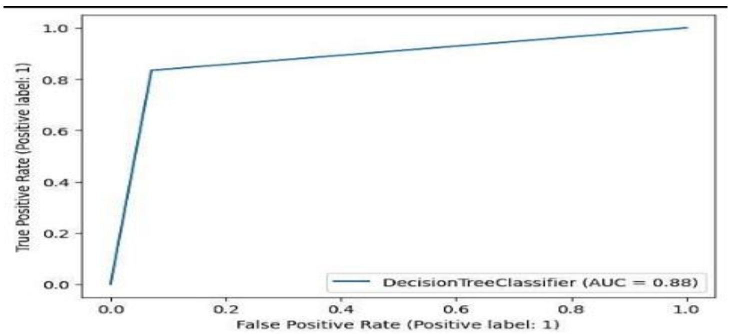

Figure 4. ROC curve for Decision Tree Classifier

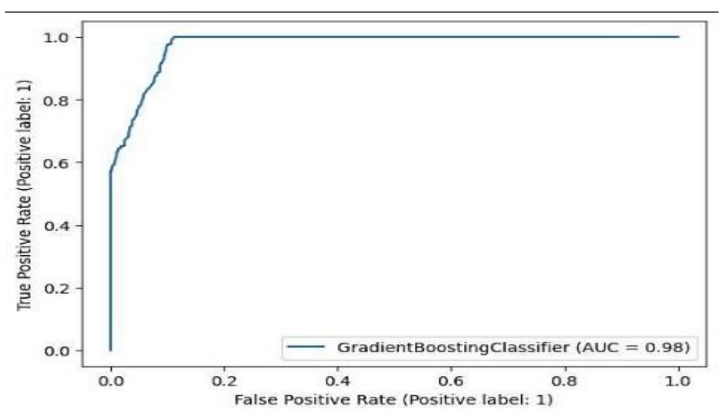

Figure 5. ROC curve for Gradient Boosting Classifier

The values generated from the user's preferences regarding the mentioned intelligent virtual assistant serve as the basis for the prediction. The machine learning model that has been implemented is given the final values for Trust, Neutral, and Untrust in order to estimate the

Nigerian Journal of Science and Environment 2024 Volume 22 (1) 57 – 67 ISSN: 1119 – 9008
https://doi.org/10.61448/njse221245

user's level of trust. The user believes the technology if the class after forecast is 1, and the likelihood of trust is shown. Additionally, the consumer will not trust the technology if the outputted class following forecast is 0. Additionally provided is the likelihood of trust that is less than 0.5 and near to 0.

The result from the Decision Tree Classifier

Table 2 displays the results obtained from the Decision Tree Classifier in terms of accuracy, precision, F1 score, Specificity, and sensitivity.

Table 2. Summary of Decision Tree Classifier results.

Model	Accuracy	Precision	F1 score	Sensitivity	Specificity
Decision Tree Classifier	0.91	0.93	0.93	0.94	0.83

The above table revealed that the Decision Tree Classifier tested an accuracy of 91% in predicting consumers' next purchase. Figure 5 shows the confusion matrix of the model.

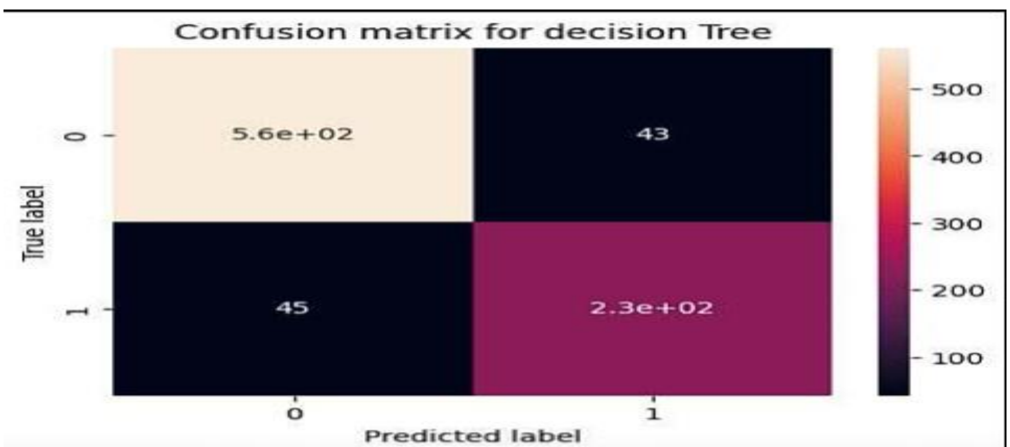

Figure 6. Confusion matrix for Decision Tree Classifier

Result from Gradient Boosting Classifier

Table 3 displays the results obtained from the Decision Tree Classifier in terms of accuracy, precision, F1 score, Specificity, and sensitivity.

Nigerian Journal of Science and Environment 2024 Volume 22 (1) 57 – 67 ISSN: 1119 – 9008
https://doi.org/10.61448/njse221245

Table 3. Summary of Gradient Boosting Classifier results.

Model	Accuracy	Precision	F1 score	Sensitivity	Specificity
Gradient Boosting Classifier	0.93	0.92	0.95	0.99	0.86

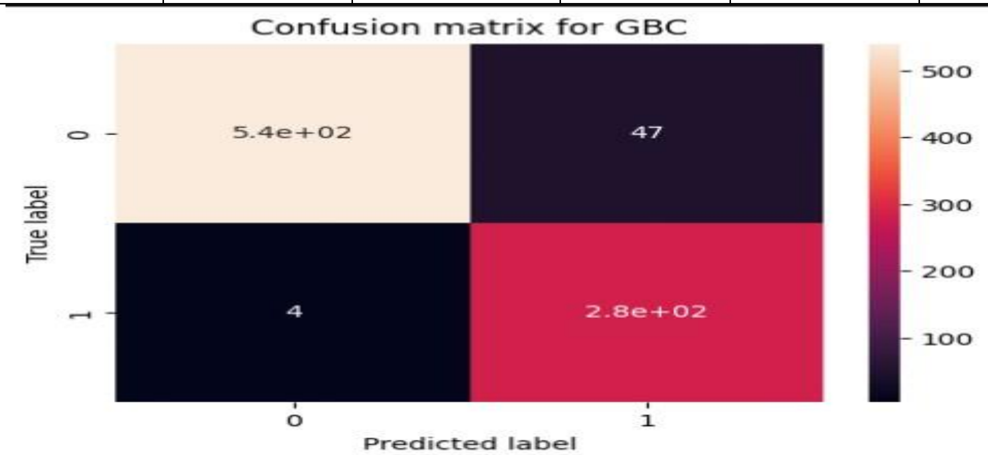

Figure 7. Confusion matrix for Decision Tree Classifier

CONCLUSION

This research presents an early purchase prediction framework that can tell us whether a potential consumer is likely to buy a product soon. To tackle the classification challenge, we used both Gradient Boosting Classifier and Decision Tree Classifier. Having an accuracy of 93% and an AUC of 0.98, the Gradient Boosting Classifier model performed the best. This work can be applied in business organizations seeking to attract and retain potential customers. It will help businesses and companies analyze, understand, and predict customer wants and needs. Hence, the company can improve the service or offer a suitable product to the customer, increase its sales, and be more successful in business.

The following are recommendations for more research:

a. Future research on the prediction of customer behaviour may build upon the findings of this study.
b. Deep learning and ensemble learning can be hybridised to do more research

COMPETING INTERESTS

Authors have declared no conflict of interest.

AUTHORS' CONTRIBUTIONS

Authors contributed collectively

REFERENCES

Akazue, M. I. (2015). A Survey of Ecommerce Transaction Fraud Prevention Models. In The Proceedings of the International Conference on Digital Information Processing, Data Mining, and Wireless Communications, Dubai, UAE.

Akazue, M. I., Aghaulor, A., & Ajenaghughrure, B. I. (2015). Customer's Protection in Ecommerce Transaction Through Identifying Fake Online Stores. In International Conference e-Learning, e-Bus., EIS, and e-Gov. IEEE'15 (pp. 52–54).

Akazue, M. I., Edoki, E. J., Ogeh, C. O., & Ufiofio, E. (2023). Application of Blockchain Technology Model in Food Palliative Distribution in Developing Countries. FUPRE Journal of Scientific and Industrial Research (FJSIR), 7(2), 81-90.

Akazue, M. I., Ekpewu, C., Omede, E., & Abel, E. E. (2023). Development of a Semantic Web Framework for the Blind. International Journal of Innovative Science and Research Technology, 8(1), 1781-1789.

Akazue, M. I., Izakpa, G. E., Ogeh, C. O., & Emmanuel, U. (2023). A secured computer based test system with resumption capability module. Kongzhi yu Juece/Control and Decision, 38(02), 1-9.

Akazue, M. I., Nwokolo, G. A., Ejaita, O. A., Ogeh, C. O., & Ufiofio, E. (2023). Machine Learning Survival Analysis Model for Diabetes Mellitus. International Journal of Innovative Science and Research Technology, 8(4), 754-760.

Akazue, M., & Augusta, A. (2015). Identification of Cloned Payment Page in Ecommerce Transaction. International Management Review, 11(2), 70-76.

Akazue, M., & Blessing, O. (2014). Building Data Mining For Phone Business. Oriental journal of computer science and technology: An international open access peer-reviewed research journal, 7(03), 316-322.

Akazue, M., Asuai, C., Edje, A., Omede, E., & Emmanuel, U. (2023). Cybershield: harnessing ensemble feature selection technique for robust distributed denial of service attacks detection. Kongzhi yu Juece/Control and Decision, Volume 38, Issue 03, July, pp 1211-1224.

Chavhan, S., Dharmik, R. C., Jain, S., & Kamble, K. (2022). RFM Analysis for Customer Segmentation using Machine Learning: A Survey of a Decade of Research. 3C TIC. Cuadernos de desarrollo aplicados a las TIC, 11(2), 166-173. Retrieved from https://doi.org/10.17993/3ctic.2022.112. 166-173.

Chiemeke, S. C., & Omede, E. U. (2014). Mal-typho diagnosis intelligent system (matdis): the auto-diagnostic rule generation algorithm. 5(4).

Cuce, A., & Tiryaki, E. (2022). Data Analytics in Customer Segmentation and RFM. Istanbul Technical University. doi:10.13140/RG.2.2.28067.94244.

Dogan, O., Aycin, E., & Bulut, Z. (2018). Customer Segmentation by using RFM model and clustering methods: A case study in retail industry. International Journal of Contemporary Economics and Administrative Sciences, 8(11), 1-19.

Efozia, N. F., Anigbogu, S. O., & Anigbogu, K. S. (2019). Development of a hybrid model for enhancing data integration process of business intelligence system. Journal of Basic Physical Research, 9(2), 1-16.

ElisabetaIoanăs, & IvonaStoica. (2014). Social media and its impact on consumers behavior. International Journal of Economic Practices and Theories, 4.

Füsunçizmeci, T. E. (2015). The effect of digital marketing communication tools in the creation brand awareness by housing companies.

Ihama, E. I., Akazue, M. I., Omede, E., & Ojie, D. (2023). A Framework for Smart City Model Enabled by Internet of Things (IoT). International Journal of Computer Applications, 185(6), 1-11.

Kadir, M. A., & Achyar, A. (2019). Customer Segmentation on Online Retail using RFM Analysis: Big Data Case of Bukku.id. The International Conference on Environmental Awareness for Sustainable Development (ICEAD). doi:10.4108/eai.1-4-2019.2287279. Retrieved from https://eudl.eu/pdf/10.4108/eai.1-4-2019.2287279.

Kotni, D. P., & Divya, S. (2019). A Study on Consumer Behaviour and Buying Patterns in Apparel Retail Environment. International Journal of Research, 6(11), 878-887.

Maureen, A., & Blessing, O. (2014). Building Data Mining For Phone Business. Oriental journal of computer science and technology: An international open access peer-reviewed research journal, 7(03), 316-322.

Ojie, D. V., Akazue, M., Omede, E. U., Oboh, E. O., & Imianvan, A. (May 2023b). Survival Prediction of Cervical Cancer Patients using Genetic Algorithm-Based Data Value Metric and Recurrent Neural Network. International Journal of Soft Computing and Engineering (IJSCE), 13(2), May 2023, ISSN: 2231-2307 (Online).

Ojie, D., Akazue, M., & Imianvan, A. (2023a). A Framework for Feature Selection using Data Value Metric and Genetic Algorithm. International Journal of Computer Applications, 184(43), 14-21.

Ojugo, A. A., Akazue, M. I., Ejeh, P. O., Odiakaose, C. C., & Emordi, F. U. (2023). DeGATraMoNN: Deep Learning Memetic Ensemble to Detect Spam Threats via Content-Based Processing. Kongzhi yu Juece/Control and Decision, 38(01), April 2023, 1-11. ISSN: 1001-0920.

Okofu, S. N., Akazue, M. I., & Ajenaghughrure, B. I. (2018). An Enhanced Speech-Based Airline Ticket Reservation Data Confirmation Module. Journal of Social and Management Sciences, 13(1), 11-21.

Okofu, S., Anazia, E. K., Akazue, M., Ogeh, C., & Ajenaghughrure, I. B. (2023). The Interplay Between Trust In Human-Like Technologies and Integral Emotions: Google Assistant. Kongzhi yu Juece/Control and Decision, 38(01), 1-11.

Ponde, S., & Jain, A. (2019). Digital Marketing: Concepts & Aspects. International Journal of Advanced Research, 7(2), 260-266.

Schiffman, L. G., & Kanuk, L. L. (2007). Purchasing Behavior (9th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.