Innovations

Customers' Loyalty Prediction Model for E-Hailing Mobility Service Companies in Nigeria

[1]Awogbemi Clement Adeyeye; [1]Ilori Adetunji Kolawole;
[2]Mabur Yaks Mafuyai; [3]Oyeyemi Gafar Matanmi; [4]Onyeagu Sydney Iheanyi
[1]Statistics Programme, National Mathematical Centre, Abuja, Nigeria
[2]Physics Department, University of Jos, Jos, Nigeria
[3]Statistics Department, University of Ilorin, Ilorin, Nigeria
[4]Statistics Department, Nnamdi Azikiwe University, Awka, Nigeria

Corresponding Author: **Awogbemi Clement Adeyeye**

DOI 10.5281/zenodo.16614174

Abstract: *The recent surge of inflation in Nigeria has significantly impacted the transportation sector. E-hailing mobility services, in particular, are likely to see a decline in customer loyalty, as they already had the highest fare rates even before the removal of subsidies on petroleum products. This research is aimed at developing a customer loyalty prediction model for e-hailing mobility services in Nigeria. Data used were collected through an online survey, with customer loyalty determined by frequency of use. A binary logistic regression model was fitted to analyze the data. The results indicated that the model has a good fit according to the Hosmer and Lemeshow Test of Goodness-of-fit, with a classification accuracy of 66%. Age and duration of service use were found to be statistically significant. Based on the odds ratio, students, self-employed individuals, low- and middle-income earners, civil and public servants, and daily and weekly users are more likely to be loyal compared to artisans, high-income earners, and yearly users. This study recommends targeted motivational service packages for these customer segments to attract and retain loyalty. Further research could explore the reason why these identified customer segments remain loyal despite rising transportation costs.*

Keywords*: Accuracy, Inflation, Transportation, Predictive Model, Ride-Hailing*

1.0. Introduction

Electronic hailing services, commonly known as e-hailing, allow users to request goods and services through internet-connected devices. This shift in purchasing behavior has also transformed the transportation sector. E-hailing mobility services provide commuters with on-demand transport options via online mobile applications, where service providers match passengers with the nearest available driver (Acheampong et al., 2020). In Nigeria, e-hailing mobility services began in Lagos when Uber launched its operations in 2014. Since then, these services have expanded rapidly, now covering most major cities and state capitals. Companies such as Uber, Bolt, and In-Drive currently dominate the Nigerian market (Ojekere et al., 2022).

The rapid growth of e-hailing services in Nigeria can be attributed to several factors, with projections estimating a market volume of US$477.10 million by 2029 (Statista, 2024). However, recent increases in the price of premium motor spirit (petrol) have sharply raised transportation costs nationwide, potentially affecting sector growth. Consequently, there is a need to understand how these price hikes impact customer loyalty in the e-hailing sector. Customer classification is a promising technique that enables companies to offer tailored services to various customer segments. This research aims to develop a prediction model using binary logistic regression to classify e-hailing mobility service customers. Specifically, the study objectives are to:

(i) Identify customer characteristics that significantly impact loyalty.

(ii) Estimate the likelihood of customer loyalty.

(iii) Classify customers based on the model.

This paper makes two main contributions: First, it presents a novel approach to determining customer loyalty through purchase history, which could be valuable for researchers utilizing secondary data. Secondly, it is the first study to classify e-hailing mobility service customers in Nigeria, offering insights into addressing customer disloyalty amid rising service costs.

2.0. Literature Review

Empirical studies on the frequency of use, adoption, and policy issues surrounding e-hailing mobility services in Africa are gaining the attention of researchers due to the growing market share of this transport sector. In Nigeria, Olawole (2022) examined the adoption and frequency of ride-hailing service usage, focusing on the impact of socio-demographic factors. The study found that the largest group of users was male, young, educated, and wealthy. Key challenges facing e-hailing services in Nigeria include a lack of regulatory framework, driver misconduct, technology barriers, unhealthy competition, multi-homing, payment gateway issues, and trust concerns (Ojekere et al., 2022).

Adenigbo et al. (2023) investigated the service quality of e-hailing taxis in Johannesburg, focusing on passenger satisfaction. The study surveyed 499 e-hailing taxi users, adding safety and affordability to the traditional five service quality dimensions. Exploratory factor analysis (EFA) identified reliability, tangibility, safety, and empathy as critical factors influencing satisfaction, with safety emerging as a primary concern. The authors recommended that e-hailing operators prioritize passenger safety through vigilant and effective measures. While this study underscores the value of classification techniques in understanding customer behavior, it differs from the current research, which is motivated by the impact of rising transportation costs in Nigeria due to increased fuel and spare parts prices.

In another related study in the rural town of Thohoyandou, South Africa, Masikhwa et al. (2024) evaluated the quality of Bolt's e-hailing services using the SERVQUAL model. The results showed that customers perceived service quality as falling short of their expectations across all five SERVQUAL dimensions. The study was driven by safety concerns in rural areas. Additionally, Acheampong et al. (2020) found that factors such as socio-demographics, perceived benefits and ease of use, perceived safety risks, and car dependency influence e-hailing adoption and usage.
Binary logistic regression, a classical statistical tool for modeling dichotomous outcomes, has been widely applied in classification studies (Cervantes et al., 2020; Olson et al., 2009). For instance, Olson et al. (2009) used logistic regression on recency, frequency, and monetary (RFM) variables, achieving an 84% correct classification rate. In a study on the prevalence of Diabetes Mellitus (DM) in Bangladesh, a two-level logistic regression model with individual and regional random intercepts was applied to identify DM risk factors (Talukder & Hossain, 2020).
Regularized logistic regression and artificial neural networks (ANNs) have also been employed for classification, such as in ozone level classification across El Paso County, Texas (Obunadike et al., 2023), with both techniques performing well. Similarly, support vector machines (SVM) have been used in various classification studies.

Chhajer et al. (2022) applied SVM for stock market prediction, noting that SVM can be faster and more efficient than ANN when handling thousands of data samples. Butar-Butar et al. (2023) used SVM to classify drought codes in North Sumatra, Indonesia. SVM is also popular in forecasting applications (Álvarez-Alvarado et al., 2021; Rashinkar & Ghonge, 2018) and is regarded as a top supervised learning algorithm (Boateng et al., 2020).

While empirical research on the e-hailing transport sector is gaining momentum, customer classification studies remain limited, particularly in Nigeria. Although many classification algorithms exist, this study focuses on logistic regression due to its popularity and robustness in classification research (Boateng et al., 2020; Chhajer et al., 2022; Ngai et al., 2009; Salminen et al., 2023).

3.0. Methodology

3.1. Data Collection

Since e-hailing mobility service companies have not made their data available to researchers (Acheampong et al., 2020), this study collected data through an online survey. The survey was administered via Google forms and distributed on social media platforms. The questionnaire included questions on demographic variables (age, gender, income, and occupation), behavioral variables (purchasing history, frequency, and recency), preference variables (service preference, service satisfaction), and company variables (company location). Responses were collected over five days, yielding a total of 160 participants.

3.2 Logistic regression.

Logit—the natural logarithm of an odds ratio is the central mathematical concept that underlies logistic regression. According to Peng et al.(2002), consider the simplest case of linear regression for one continuous predictor X (a child's reading score on a standardized test) and one dichotomous outcome variable Y (the child being recommended for remedial reading classes). The logit is the natural logarithm (ln) of odds of Y, and odds are ratios of probabilities (Π) of Y happening (i.e., a student is recommended for remedial reading instruction) to probabilities (1-Π) of Y not happening (i.e., a student is not recommended for remedial reading instruction). The simple logistic model has the form:

$$\mathrm{Logit}(Y) = \text{natural}\log(\text{odds}) = \ln\left(\frac{\Pi}{1-\Pi}\right) = \alpha + \beta X \qquad (1)$$

Taking the antilog of Equation 1 on both sides, one derives an equation to predict the probability of the occurrence of the outcome of interest as follows:

$$\Pi = \mathrm{Probability}(Y = \text{outcomeofinterest}/X = x, \text{aspecificvalueofX}) = \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}} \qquad (2)$$

Extending the logic of the simple logistic regression to multiple predictors (say X_1 = reading score and X_2 = gender), a complex logistic regression for Y (recommendation for remedial reading programs) as follows:

$$\mathrm{Logit}(Y) = \text{natural}\log(\text{odds}) = \ln\left(\frac{\Pi}{1-\Pi}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 \qquad (3)$$

And,

$$\Pi = \mathrm{Probability}(Y = \text{outcomeofinterest}/X_1 = x_1, X_2 = x_2) = \frac{e^{\alpha+\beta_1 X_1+\beta_2 X_2}}{1+e^{\alpha+\beta_1 X_1+\beta_2 X_2}} \qquad (4)$$

where Π is once again the probability of the event, α is the Y intercept, βs are regression coefficients, and Xs are a set of predictors. α and βs are estimated by the Maximum Likelihood (ML) method or weighted least squares approach. The null hypothesis underlying the overall model states that all βs equal zero. A rejection of this null hypothesis implies that at least one β does not equal zero in the population, which means that the logistic regression equation predicts the probability of the outcome better than the mean of the dependent variable Y. The interpretation of results is rendered using the odds ratio for both categorical and continuous predictors.

3.3. Model's Fit and Performance Evaluation

The model's goodness-of-fit was evaluated using the Hosmer-Lemeshow test, and its performance was assessed using a Confusion Matrix. According to the Hosmer-Lemeshow goodness-of-fit criteria, a model fit is considered good if the p-value is greater than 0.5 and the Chi-Square value is less than 20. A fair model fit occurs when the p-value is greater than 0.5 and the Chi-Square value is between 20 and 30, while a poor fit is indicated by a p-value less than 0.5 and a Chi-Square value above 30. Additionally, the Confusion Matrix evaluates classification accuracy, where correct classifications are represented by the diagonal elements, True Negative (TN) and True Positive (TP), as shown in Table 1.

Table1: Confusion Matrix Table

Classification		Observation	
		Negative	Positive
Prediction	Negative	True negative (TN)	False positive (FP)
	Positive	False negative FN	True positive (TP)

4.0. Data Analysis and Discussion of Results

4.1. Customer Loyalty Determination using Purchase History.

Customer loyalty is defined as the consistency in purchase volume, regardless of any events that impact the product or service purchased. It is measured as the difference between purchase volume or frequency after an event affecting the product or service and the volume or frequency before the event. If this difference is less than zero, the customer is classified as not loyal; if it is equal to or greater than zero, the customer is considered loyal.

In this study, customers were asked to report the number of times they used e-hailing mobility services per day before the transport fare increase and after the increase. The difference between the frequency of use before and after the price change was calculated, and customers were then classified as either loyal or not loyal based on this measure.

4.2. Data Preparation

A total of 159 responses were collected from the five-day online survey. The data was analyzed for missing values, and none was found. The independent variables were encoded using the dummy variable encoding method. Age was classified into three groups: young, middle-aged, and senior. Income level was categorized as low-income (0-100,000 Naira), middle-income (100,000-200,000 Naira), and high-income (200,000 Naira and above). Gender was classified as male and female, while duration of e-hailing service use was classified as 1 year, 2 years, 3 years, 4 years, and more than 4 years. Occupation was divided into student, civil or public servant, self-employed, and artisan, and user type was categorized as daily, weekly, monthly, and yearly. Lastly, use preference was classified as individual, group, or package delivery. The analysis was conducted using IBM SPSS statistical software, version 26.

4.3. Result of loyalty analysis

Respondents were assessed for loyalty using the method outlined in Section 4.1 and were classified as either loyal or non-loyal. Their classifications and corresponding counts are presented in Table 2.

Table 2: Classes of Customers based on Loyalty

Number of Loyal customers	Number of None-Loyal customers
82	77

4.4. Results of Model Developed

A predictive model was developed using binary logistic regression. The model summary and evaluation results are presented below

Table 3: Model Summary of Predictive Model Developed

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	193.178[a]	.157	.209
a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.			

Table 4: Hosmer and Lemeshow Test of Goodness-of-Fit

Step	Chi-square	df	Sig.
1	5.545	8	.698

Table 5: Summary of Contribution of each Independent Variable Relative to a Given
Reference Variable in the Model

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Age group 0-25 years	-2.901	1.366	4.509	1	.034	.055
	Age group 25-50 years	-2.227	1.175	3.593	1	.058	.108
	Male	-.549	.509	1.167	1	.280	.577
	Student	.610	1.365	.199	1	.655	1.840
	Civil Publicservant	.178	1.361	.017	1	.896	1.195
	Self employed	.574	1.329	.187	1	.666	1.776
	Income group earning between 0-100,000 Naira per month	.451	.527	.730	1	.393	1.569
	Income group earning between 100,000-200,000 naira per month	.506	.503	1.012	1	.314	1.659
	More than 4years	-1.740	.611	8.119	1	.004	.176
	4 years	-1.628	.775	4.415	1	.036	.196
	3 years	-1.340	.637	4.421	1	.035	.262
	2 years	-1.463	.674	4.706	1	.030	.232
	Daily Users	.012	.670	.000	1	.986	1.012
	Weekly Users	.033	.593	.003	1	.956	1.033
	Monthly Users	-.437	.585	.560	1	.454	.646
	Individually	-.641	.979	.429	1	.512	.527

	In_a_group	-.374	.989	.143	1	.705	.688
	Constant	3.933	2.190	3.227	1	.072	51.076
a. Variable(s) entered on step 1: Age group 0-25 years, Age group 25-50 years, Male, Student, Civil Public servant, Self employed, Income group earning between 0-100,000 naira per month, Income group earning between 100,000-200,000 naira per month, More than 4years, 4years, 3years, 2years, Daily Users, Weekly Users, Monthly Users, Individually, In a group.							

Table 6: Classification Rate Table[a]

Observed			Predicted		
			Customer_Loyalty		Percentage Correct
			Not Loyal	Loyal	
Step 1	Customer_Loyalty	Not Loyal	52	25	67.5
		Loyal	29	53	64.6
	Overall Percentage				66.0
a. The cut value is 0.500					

4.5. Discussion of Results

From Table 2, it is observed that 51.6% (82) of the customers are loyal, while 48.4% (77) are non-loyal. The slight difference between loyal and non-loyal customers indicates that e-hailing mobility service customers in Nigeria are sensitive to price changes.

Table 3 presents the summary of the binary logistic regression model for classifying e-hailing mobility service customers in Nigeria. The model's fit is considered good according to the Hosmer and Lemeshow Test of Goodness-of-Fit criteria (Table 4). The model's classification rate is 66% (see Table 6) and the model correctly classified 105 out of the 159 customers who participated in the study.

As shown in Table 5, only age and duration of use significantly affect the model's predictive accuracy. Some independent variables, such as age, gender, duration of use, and preference of use, have a negative impact on the model's predictions. Furthermore, from the odds ratio, students and self-employed customers are 84% and 77.6% more likely to be loyal compared to artisans. In contrast, civil and public servants are 19.5% more likely to be loyal. This result can be explained in two ways: (a) students and self-employed individuals may not own private vehicles and therefore depend on commercial vehicles, especially e-hailing rides, for daily commuting;(b) the erosion of purchasing power for civil and public servants, caused by the ongoing inflation, may be responsible for this group continuing to patronize e-hailing services.

Similarly, low and middle-income earners are 56.9% and 65.9% more likely to be loyal compared to high-income earners. This could be due to their lack of personal vehicles or the inability to maintain their vehicles due to increased fuel and maintenance costs. Likewise, daily and weekly users are 1.2% and 3.3% more likely to be loyal compared to yearly users, likely due to the necessity of mobility for work, school, or business purposes.

5.0. Conclusion and Recommendation

A predictive model for e-hailing mobility service customers' loyalty has been developed, with a good fit and a classification rate of 66%. The model identifies age and duration of use as the only statistically significant ($p<0.05$) customer characteristics. The results indicate that students, self-employed individuals, low and middle-income earners and daily or weekly users are more likely to be loyal customers. Based on these findings, this study recommends that service providers develop motivational packages targeted at these customer groups to attract and retain their loyalty. Further research could explore the reason why these identified customer segments remain loyal despite the increase in transportation fares.

Acknowledgement

The authors wish to express their gratitude to the Management of National Mathematical Centre (NMC), Abuja, Nigeria for the opportunity given to them to participate in the Research Oriented Course (ROC) in Multivariate Analysis & Applications.

References

1. *Acheampong, R. A., Siiba, A., Okyere, D. K., & Tuffour, J. P. (2020). Mobility-on-demand: An empirical study of internet-based ride-hailing adoption factors, travel characteristics and mode substitution effects. Transportation Research Part C: Emerging Technologies, 115, 102638.*
2. *Adenigbo, A. J., Mageto, J., Makan, H., & Luke, R. (2023). Service quality of e-hailing taxi services in Johannesburg. A l Acta lActa l Acta Logistica Ogisticaog, 10(4), 537–548.*
3. *Álvarez-Alvarado, J. M., Ríos-Moreno, J. G., Obregón-Biosca, S. A., Ronquillo-Lomelí, G., Ventura-Ramos, Jr., E., & Trejo-Perea, M. (2021). Hybrid Techniques to Predict Solar Radiation Using Support Vector Machine and Search Optimization Algorithms: A Review. Appl. Sci., 11(1044), 2–17.*
4. *Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and*

Neural Network: A Review. Journal of Data Analysis and Information Processing, 08(04), 341–357.

5. *Butar-Butar, K. D., Sihombing, P., & Tulus. (2023). Application of Support Vector Machine on Drought Code Classification in North Sumatra Indonesia. Indonesian Physical Review, 6(1), 51–56.*
6. *Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408, 189–215.*
7. *Chhajer, P., Shah, M., & Kshirsagar, A. (2022). The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction. Decision Analytics Journal, 2, 100015.*
8. *Lengua, M. A. C., & Quiroz, E. A. P. (2020). A Systematic Literature Review on Support Vector Machines Applied to Classification. IEEE Xplore.*
9. *Masikhwa, M., Nyamwanza, S., & Chakwizira, J. (2024). A Quality Assessment of Bolt E-hailing Service Delivery in Rural Towns. A Case Study of Thohoyandou, South Africa. Journal of Psychiatry and Neurochemistry Research, 2(4), 01-12.*
10. *Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications, 36(2), 2592–2602.*
11. *Obunadike, C., Adefabi, A., Olisah, S., Abimbola, D., & Oloyede, K. (2023). Application of Regularized Logistic Regression and Artificial Neural Network Model for Ozone Classification across El Paso County, Texas, United States. Journal of Data Analysis and Information Processing, 11(03), 217–239.*
12. *Ojekere, S., Ojo, O., & Mkpandiok, D. (2022). Sustainable E-Hailing Mobility Services in Nigerian Cities: Issues and Policy Direction. Iconic Research and Engineering, 6(5), 90–96.*
13. *Olawole, M. O. (2022). Ride-Hailing Services in Nigeria: Adoption, Insights and Implications. In Transport technology and Innovations in Nigeria: Policy Guides to Sustanable Future (1st ed., Vol. 1, pp. 121–147). Nigerian Institute of Transport Technology.*
14. *Olson, D. L., Cao, Q., Gu, C., & Lee, D. (2009). Comparison of customer response models. Service Business, 3(2), 117–130.*
15. *Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An Introduction to Logistic Regression Analysis and Reporting. The Journal of Educational Research, 3–14.*
16. *Rashinkar, Yogesh. D., & Ghonge, P. A. (2018). Application of Support Vector Machine for Wind Speed Forecasting. International Journal of Research & Review, 5(7), 161–165.*
17. *Salminen, J., Mustak, M., Sufyan, M., & Jansen, B. J. (2023). How can algorithms help in segmenting users and customers? A systematic review and research*

agenda for algorithmic customer segmentation. Journal of Marketing Analytics, 11(4), 677–692.

18. *Talukder, A., & Hossain, Md. Z. (2020). Prevalence of Diabetes Mellitus and Its Associated Factors in Bangladesh: Application of Two-level Logistic Regression Model. Scientific Reports, 10(1), 10237.*