Comparative Analysis of Logistic Regression and Decision Tree Models for Forecasting Users' Adoption of Electronic Payments

Dr. Farhad Ali Ahmed
farhad.ahmed@univsul.edu.iq

Dr. Suzan Saber Haydar
sozan.haider@univsul.edu.iq

1,2 Sulamani University, Collage of Administration and Economics, Department of Statistics and Informatics

Date of receipt of the research: 1/11/2025 **Date of approval for publication: 25/11/2025**

Abstract:

Over the next few years, the financial system will increasingly undergo a digital transformation where electronic payments will be required to enhance the efficiency of transactions and advance financial inclusion. In this research, we intend to use a decision tree and logistic regression statistical prediction models to evaluate how well these models can predict users' intention to electronic payments. Research data on such characteristics including accessibility, financial incentives, user experience, trust, and ease of use were collected from 200 respondents through a questionnaire. Both of the models were evaluated based on Accuracy and F1-Score indices. Results showed that the fit logistic regression predicts user behavior better. The findings indicate that through statistical models, financial behavior can be addressed along with encouraging and improving the quality of the electronic payment system.

Keywords: Financial Inclusion, Electronic Payment, Logistic Regression, Decision Tree, User Behavior Prediction

1.Introduction:

Digital transformation became a hot topic in banking and financial sector in modern years. With the introduction of more modern electronic payment systems and new financial technologies (FinTech), a completely new way appeared to improve and reconcile the financial system and banking services. Despite these developments, as many countries, such as Iraq, with little adoption and usage of electronic payment systems, and still, people use the cash to execute the transaction.

The emergence of new technologies, such as e-payment, is driven not only by their innovation but also by various influencing factors. Those factors contain user demographics, technology access, trust in security, experience, and monetary or bank-supported incentives. Understanding and quantifying such factors enriches user experience design and assists modeling user behavior and impacts banking rules pertaining to service delivery. The relevance can easily be explained by a digest of decision trees and logistic regression for analyzing and predicting user behaviors and providing better insights regarding factors associated with electronic payment adoption. Each of them has a common attribute. Features ranging from the probability of an event, as seen in logistic regression, to model construction decision making offer complex global patterns to uncover. The objective of present study is to identify the factors of electronic adoption by some statistical models. Study results could help in the improvement of digital initiatives from banks and legislatures, and ultimately improve the financial system. Lufunda (2024) examined uptake of digital payment platforms for Zambian consumers based on a consumer survey in Lusaka in Zambia. According to the paper, the key obstacles in the path of larger adoption relating to low financial literacy, high transaction costs, lack of regulatory frameworks and perceived risks. The research stated that changes in regulations is essential as well as user education so it could accelerate the adoption of this technology. [6] Methods for predicting the worst non-financial payment status in commercial credit: Evidence from over 11 million businesses by analyzing Equifax data Rudd and Priestley (2017) Their results also indicate that the Kolmogorov-Smirnov statistic and the TOC index-based decision tree model is as predictive as the logistic regression model, and is useful in predicting worst payment default events in commercial lending. [9] Wani et al. Chin et al (2025) used public transportation consumer data in many emerging cities to examine the adoption of digital payments. The research found the main factors to drive the digital payment adoption. This includes ease of use, trust in security, and prior experience using. Machine learning methods, particularly decision trees, random forest, and XGBoost were also very good at finding the underlying patterns of adoption. [12] In the second part of this study, research methodology is presented, including data collection methods, the study population and statistical sample, and methods for analyzing the data. In the third part, we build and analyze the data and compare logistic regression and decision tree models for predicting electronic payment acceptance. The quantitative results and model comparisons will be shown in 4th section and the fifth section will interpret the findings. The last of these analyses will examine the factors in electronic payment adoption and suggest ways to improve.

2.Methodology:

The second segment introduces the theoretical background of the study, which is related to the models under which the data are modeled. The two most common means of analyzing user behaviors and predicting adoption of electronic payments are logistic regression and decision tree models (LUO et al. These models can assist us in detecting complex patterns of categorization (decision tree) and the impact that independent variables influence on the likelihood of adoption (logistic regression). The outline of this theoretical framework constitutes the basis for data analysis as well an explanation of the statistics found throughout the research.

2.1. Decision Tree Model

CART is a decision tree model. The decision tree is a nonparametric classification instrument. This strategy uses simple patterns to form a classification pattern for the current observations. A highly interpretable and easy decision-making pattern arises from the proposed method. Interestingly, it has diagnostic and prediction accuracy on par with far more complex techniques (like neural networks) despite its simple methodology. A decision tree is an acyclic, tree-like graph that works by high-level recursive questions. This straightforward method is incredibly powerful for splitting a dataset into clean and uniform buckets. [2] Usually, each question is answered with respect to one variable. A root is decided to be an auxiliary variable that be partitioned in several internal nodes over a sequence of queries along with features. During this process, a decision tree graph is generated and it consists of three components: root, internal nodes and external nodes (leaves). While there are other algorithms for carrying out decision trees. The first three are the most important: C5.0, Chaid and Quest. Methods with a comparison method: in this investigation, the CART decision tree was used. The CART classification tree type has been proposed by Classical Associates in 1984. The model utilizes an undirected, tree structured graph with binary splits based in auxiliary variables to describe a classification and diagnostic pattern. The methodology is as follows: First, one of the variables is selected as a root auxiliary variable, then it is split to different internal nodes depending upon the purpose of the study. Bremen (and each internal node, e.g., root) is split into other nodes until each node gets assigned a response variable category.[8] We refer to these nodes as leaves. The Gini coefficient and the disorder function were used in this study to choose significant variables for the tree classification model. The awkwardness function for a node named t and a dependent variable with k ranks (C_1, C_2, C_K) is defined as follows:

$$i(t) = \Phi \begin{bmatrix} P(C = \frac{C_1}{t} \\ p(C = \frac{C_k}{t} \end{bmatrix}$$

The Gini index, which has the following definition, is frequently utilized in tree models that have binary splits at every node.

$$\begin{aligned} i(t) &= gini(t) \\ &= 1 - \textstyle\sum_{j=1}^{k} P^2 \left[P(C = \frac{C_1}{t}\right] \\ &= \textstyle\sum_{k=1} P(C = \frac{C_k}{t})\, P(C = \frac{C_1}{t}) \end{aligned}$$

When the data fall into only one category, the aforementioned relation is equal to zero; when the probabilities of each category are equal, it chooses the maximum probability. The disorder function, which is created using the Gini index as follows, will decrease when we take into account the auxiliary variable x, which is the basis for the division of the node (t) into (n) sub-branches, each of which is represented by (Tj) for (j) from (1) to (n).

$$gini\ gini = GG(T,X)$$
$$= gini(t)\ \sum_{j=1}^{n} P(\frac{t_j}{t})\ .gini(t_j)$$

The variable that provides the highest value for GG (T,X) among a number of variables is the appropriate one. This is the standard by which the best variable is chosen. Therefore, taking into account the disorder function and the Gini index, the value of the disorder function in the general case is first calculated for the response variable. The value of the disorder function in each of the two created subsets is then calculated, and their weighted average is deducted from the value of the total disorder functions. This process is repeated for all auxiliary variables, taking into account the best binary distributions for the response variable. In the first step of tree classification, the variable with the highest value for this relationship is chosen from among the auxiliary variables. We designate a point, like a, as the cut-off point when working with quantitative variables rather than using binary divisions. It is important to note that the index itself in this case, the Gini index specifies the cut-off point in many tree classification methods. Each level of the variable is regarded as a subbranch of the classification tree when working with qualitative variables.[3] In the CART model, a method called Complexity Cost is used to select the appropriate size of the classification tree. A tree model is suitable when, in addition to performing well for existing observations (the training sample), it is also suitable for new observations (the test sample). Although the classification accuracy for the training sample increases as the size of the classification tree increases, this accuracy decreases for the test sample of the dimension. The method in question actually establishes a balance between the accuracy of the classification tree and its size, and decides which node to remove from the tree, considering the line size and the number of tree nodes.[5]

2.2. Logistic Regression Model:

The dependent variable in many research may have two outcomes and is not continuous. For instance, it might only accept one of two values: zero or one, where zero indicates that the event did not occur and one indicates that it did. For instance, we might utilize a variety of factors, such as an individual's effort, intelligence, success, or failure, to assess whether a business is experiencing financial crisis. We employ logistic regression in these situations. With the exception of the different coefficient estimates, logistic regression is comparable to ordinary regression.[11] With the exception of the different coefficient estimates, logistic regression is comparable to ordinary regression.

SJH p-ISSN: 1813-0852, e-ISSN: 2617-3034

$$P(Y=1)=\frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_nX_n)}}$$

Where:

$P(Y=1)$ is the probability that the event occurs (e.g., user adopts electronic payment)

$X_1, X_2, \ldots, X_n$ are independent variables

β_0 is the intercept (constant term)

$\beta_0, \beta_2, \ldots, \beta_n$ are the coefficients for each independent variable

e is the base of the natural logarithm

Logistic regression maximizes the likelihood that an event will occur rather than minimizing the square of the errors that occur in ordinary regression.

2.3 Model Evaluation

Standard assessment scales will be used to gauge the effectiveness of the logistic regression and decision tree models; two of the most significant will be discussed below.

2.3.1 .Accuracy

A common indicator for assessing how well predictive models operate is accuracy, which shows what proportion of the model's predictions were accurate. This statistic enables a comparison of the models' overall performance and can be computed for both logistic regression and decision tree models. [1] The more accurately a model can categorize samples, the better its accuracy, which is represented as a percentage.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Where:

TP = True Positives (correctly predicted positive cases)

TN = True Negatives (correctly predicted negative cases)

FP = False Positives (negative cases incorrectly predicted as positive)

FN = False Negatives (positive cases incorrectly predicted as negative)

2.3.2. F1-Score

When the data is uneven, the F1-Score, a composite measure that condenses Precision and Recall into a single score, is very helpful. This metric measures the model's ability to correctly identify positive instances.[1] A higher value indicates a greater success rate in detecting these positive samples.

The F1-Score is calculated using the following formula:

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Where:

Precision $= \frac{TP}{TP+FP}$ (proportion of true positive predictions to the total number of positive predictions)

Recall $= \frac{TP}{TP+FN}$ (proportion of correctly positive predictions to the total number of actual positive cases)

3. Results and Data Analysis

This study will establish two predictive models' logistic regression and decision tree to predict user acceptance of electronic payment applications. After running the models, their performance will be compared based on standard metrics including Accuracy and F1-Score. Then, a comparison will be made to see which model demonstrates greater predictive ability and which approach demonstrates greater predictive ability given the data.

3.1 Logistic Model:

The logistic regression model used to predict e-payment acceptance is derived from a dataset collected from 200 samples. It showed a great ability to predict which users use the service, with the final model reaching an accuracy of 96.7% and an F1-Score of 96.3% in the test set. The table below gives a thorough performance breakdown including the count of correct and wrong predictions for positive and negative classes.

Table.1. Logistic Regression Model Performance

index	Accuracy	F1-Score	True Positives	True Negatives	False Positives	False Negatives
Value	96.70%	96.30%	26	32	0	2

These results show that the logistic regression model is able to predict customers at high-rate accuracy with accuracy and F1-Score scores of 96.7% and 96.3%, respectively. From 60 test samples, it has correctly identified as 32 negative and 26 positive samples, and misclassified only two positive samples as negative (figure (1)). This suggests that the logistic model had excellent discriminating ability between electronic payment adopters and non-adopters. The findings affirm the strength of this model in distinguishing user behavior and predicting e-payment-providers acceptance while laying the stage for logistic regression to remain a credible and useful analytical tool in the future. One reason for the impressive results is that this model only needs to identify relationships between independent variables (like in our case accessibility, trust, ease of use, and user experience) and a dependent variable (adoption decision). Our F1-score shows that our precision is balanced with our recall, and thus, our false positives and false negatives are relatively low, as well. Behaviorally, this finding indicates that users' choices to use e-payments are explainable through quantifiable psychological and functional variables. As with prior research, the strong predictors of adoption from the previous model still hold: trust and ease of use.[7]

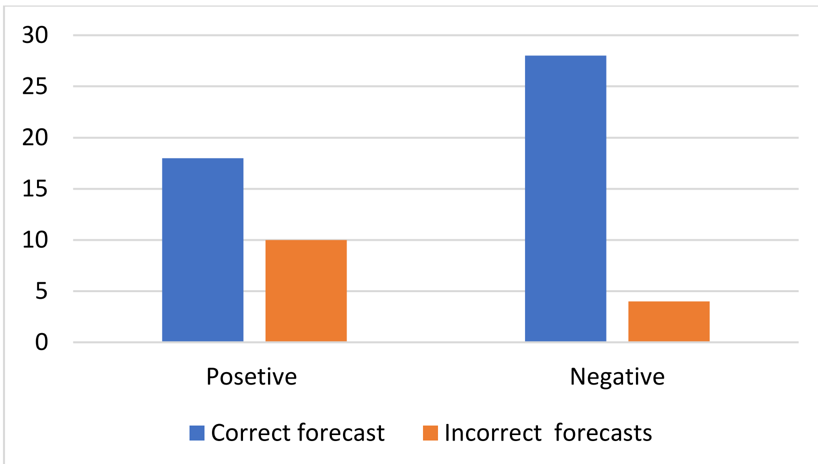

Figure.1. Confusion Matrix summary for Logistic Model

3.2 The Decision Tree Model:

Figure 6: Decision tree model developed based on the study dataset to predict e-payment acceptance That model was able to predict the user actions pretty accurately with accuracy of 76.7% and F1 score of 72.0% Its performance details including correct and wrong predictions for positive and negative classes in table (2).

Table.2. Decision Tree Model Performance

index	Accuracy	F1-Score	True Positives	True Negatives	False Positives	False Negatives
Value	76.70%	72.00%	28	18	4	10

Hence, the output illustrates that the decision tree model performed well in prediction with an accuracy score of 76.7% and an F1 score of 72.0%. In Figure (2), of 60 test samples, it found 28 negative samples and 28 positive samples, 4, and 10 negative samples were classified as false positive and false negative respectively. Related: This gives us a strong affirmation that the model

sjh@univsul.edu.iq

is able to detect user behaviors and predict e-payment acceptance rather well. Some of the reasons that could have contributed to the decision tree performing less than logistic regression are to begin with, decision trees are very volatile towards the size of the sample and the randomization of the data. Since there are only 200 observations, we could have overfitting or fragmentation of data on the model end, making it not generalize. Secondly, even if decision trees cope well with nonlinear interactions, the structure of the dataset seemed mostly linear, favoring the logistic model method. However, for interpretability and exploratory analysis, the decision tree still has its usefulness, as one can visually map paths taken by users while making decisions on the tree. The root node, for example, also deemed “trust in e-payment security”—and then subsequently ease of use and financial incentives—while valuable to adoption decisions as the most important variable in the process. These patterns are in line with findings by Sujatha & Kavitha (2023), who reported similar hierarchies in the context of technology acceptance modeling. [10]

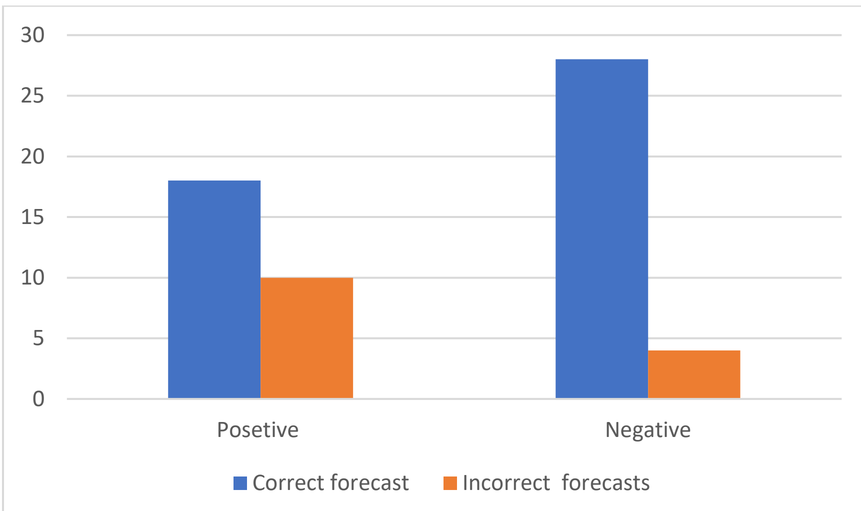

Figure.2. Confusion Matrix summary for Decision tree Model

3.3. Model Evaluation and Selection

Two logistic regression and decision tree models have been evaluated in order to more precisely identify whether model is more qualified to forecast user acceptance of electronic payments. According to the research results, the logistic regression model was more effective in forecasting the acceptance of electronic payments. This model's accuracy of approximately 96.7% and F1 of 96.3% permitted it to forecast the samples with a relatively low error. This model's primary features are its simplicity, high stability, and straightforward coefficient interpretation; nevertheless, it has limitations when it comes to simulating nonlinear interactions. With an accuracy of 76.7% and an F1 of 72%, the decision tree model, on the other hand, produced less reliable findings. Although this model is good at modeling nonlinear interactions between variables, its main drawbacks are its low accuracy and high reactivity to training data. The decision tree model's small sample size and the impact of data noise are most likely the causes of this performance difference. As can be seen in Figure (3), logistic regression typically represents the more accurate and stable model for the data in this study. However, decision trees or a combination of them with other models may be beneficial in situations when the data are more complicated and contain nonlinear interactions.

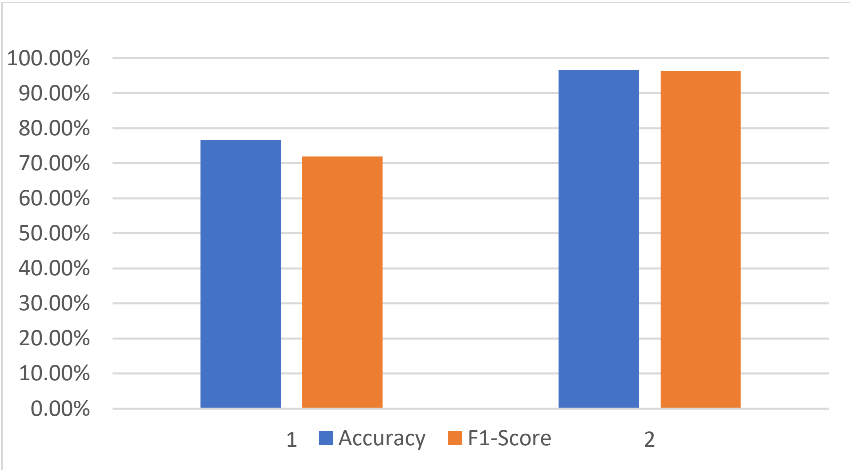

Figure.3. Comparison between models

4. Conclusion:

The results of this study indicate that machine learning and statistical approaches can be effective instruments for evaluating and foreseeing user behavior in the discipline of electronic payment use. Whenever two forecasting models' Logistic regression and decision tree were developed based on the questionnaire data, it emerged that the logistic regression model was more precise as well as reliable and could forecast user behavior with less error than the decision tree. The result suggests that classical statistical models, such logistic regression, outperform decision tree-based models where the data has a relatively linear pattern as well as are a few variables present. nevertheless, the decision tree model can be an intelligent choice in more complex data sets and also provides an opportunity to recognize nonlinear relationships and interactions between variables. Finally, it can be maintained that decision makers may establish policies and plans for the further development of the digital payment system through the use of predictive models in the fields of financial inclusion and electronic payments. To enhance prediction accuracy, it suggests that future studies utilize hybrid or artificial intelligence-based models (such Random Forest or XGBoost), larger sample sizes, and more realistic data or incorporate ensemble methods such as random forest or gradient boosting to enhance prediction accuracy.

References

1. Alhassan, M. D., Osei-Assibey, E., & Boateng, R. (2022), "*Predicting mobile money adoption using machine learning models: A comparative study*", *Information Systems Frontiers*, 24(3), 811–828.

2. Breiman, L., Friedman, J., Olshen, R., & Stone, C. (2017), "*Classification and Regression Trees",* CRC Press.
3. Digital payment adoption in public transportation: Mediating role of mode choice segments in developing cities
4. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013), "*Applied Logistic Regression"* (3rd ed.). Wiley.
5. Kumar, S., & Alok, S. (2023), "*Machine learning approaches for predicting digital payment adoption in emerging economies*", *Journal of Financial Innovation*, 9(2), 56–74.
6. Lufunda, B. L. (2024). DIPLOMA THESIS ASSIGNMENT.
7. Nguyen, T. T., & Le, H. P. (2021), "*Comparing machine learning models for predicting fintech adoption*", *Computers in Human Behavior Reports*, 4, 100120.
8. Quinlan, J. R. (1996), "*Improved use of continuous attributes in C4.5*", *Journal of Artificial Intelligence Research*, 4, 77–90.
9. Rudd, M. P. H., GStat, J. M., & Priestley, J. L. (2017). A comparison of decision tree with logistic regression model for prediction of worst non-financial payment status in commercial credit.
10. Sujatha, R., & Kavitha, D. (2023), "*A hybrid approach to predict e-wallet adoption using decision tree-based models*. *International Journal of Data Science and Analytics"*, 17(4), 415–427.
11. Venkatesh, V., Thong, J. Y., & Xu, X. (2012), "*Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology*", *MIS Quarterly*, 36(1), 157–178.
12. Wani, S. A., Pani, A., Mohan, R., & Bhowmik, B. (2025). Digital payment adoption in public transportation: Mediating role of mode choice segments in developing cities. *Transportation Research Part A: Policy and Practice*, *191*, 104319.

تحليل مقارنة لنماذج الانحدار اللوجستي وشجرة القرار للتنبؤ بتبني المستخدمين للمدفوعات الإلكترونية

الملخص:

على مدار السنوات القليلة القادمة، سيشهد النظام المالي تحولاً رقمياً متزايداً، حيث ستكون المدفوعات الإلكترونية مطلوبة لتعزيز كفاءة المعاملات وتعزيز الشمول المالي. في هذا البحث، نعتزم استخدام نموذجي التنبؤ الإحصائي لشجرة القرار والانحدار اللوجستي لتقييم مدى قدرة هذه النماذج على التنبؤ بنية المستخدمين لاستخدام الدفع الإلكتروني. تم جمع بيانات البحث المتعلقة بهذه الخصائص، بما في ذلك إمكانية الوصول والحوافز المالية وتجربة المستخدم والثقة وسهولة الاستخدام، من 200 مشارك من خلال استبيان. تم تقييم كلا النموذجين بناءً على مؤشرات الدقة ودرجة F1. أظهرت النتائج أن الانحدار اللوجستي الملائم يتنبأ بسلوك المستخدم بشكل أفضل. تشير النتائج إلى أنه من خلال النماذج الإحصائية، يمكن معالجة السلوك المالي إلى جانب تشجيع وتحسين جودة نظام الدفع الإلكتروني.

الكلمات المفتاحية :الشمول المالي، الدفع الإلكتروني، الانحدار اللوجستي، شجرة القرار، التنبؤ بسلوك المستخدمين.

هەڵسەنگاندنی بەراوردی مۆدێلەکانی ڕێگرێسیۆنی لۆجستیك و دارە بڕیار بۆ پێشبینی وەرگرتنی پارەدانی ئەلیکترۆنی لەلایەن بەکارهێنەرانەوە.

پۆختە:

لە چەند ساڵی داهاتوودا، سیستەمی دارایی زیاتر گۆڕانکارییەکی دیجیتاڵی بەسەردا دێت، و ئەمە وەک هۆکارێک دەبێت بۆ پەرەپێدانی پارەدانی ئەلیکترۆنی بۆ بەرزکردنەوەی کاریگەری مامەڵەکان و پێشخستنی گشتگیری داراییەکان. ئامانجی ئەم توێژینەوەیە بەکارهێنانی مۆدێلە ستاتستیکییەکانی پێشبینی، وەک دارە بڕیارەکان و ڕێگرێسیۆنی لۆجستیکە، بۆ هەڵسەنگاندنی ئاستی توانای ئەم مۆدێلانە لە پێشبینی مەبەستی بەکارهێنەران بۆ پارەدان بە شێوەی ئەلیکترۆنییە. داتاکانی توێژینەوە لەسەرئەو تایبەتمەندیی و فاکتەرەرانەی وەکو دەستڕاگەیشتن، هاندانە داراییەکان، ئەزموونی بەکارهێنەر، متمانە، و ئاسانکاری بەکارهێنان لە ڕێگەی پرسیارنامەیەکەوە کۆکراونەوە، کە 200 کەس وەڵامی داوەتەوە. هەردوو مۆدێلەکە لە سەر بنەمای پێوەرەکانی وردبینی (Precision) و F1-Score هەڵسەنگێندراون. ئەنجامەکان دەریانخستوە کە ڕێگرێسیۆنی لۆجستیکی لە پێشبینی ڕەفتاری بەکارهێنەران کاریگەرێکی باشتری نواندوە. ئەنجامەکان ئاماژە بەوە دەدەن کە بە بەکارهێنانی مۆدێلە ستاتستیکییەکان، دەتوانرێت چارەسەری ڕەفتاری دارایی بە شێوەیەکی زانستی بکرێت و لە هەمان کاتدا کوالیتی و هاندانەی سیستەمی پارەدانی ئەلیکترۆنی باشتر بکرێت.

وشە سەرەکییەکان: گەیاندنی دارایی، پارەدانی ئەلیکترۆنی، ڕێگرێسیۆنی لۆجستیکی، دارە بڕیار، پێشبینی ڕەفتاری بەکارهێنەران.