International Journal of Engineering Applied Sciences and Technology, 2025
Vol. 9, Issue 11, ISSN No. 2455-2143, Pages 15-22
Published Online March 2025 in IJEAST (http://www.ijeast.com)

CREDIT CARD DEFAULT PREDICTION USING MACHINE LEARNING: A LOGISTIC REGRESSION MODEL APPROACH

Quang Dang Minh
Student, HUS High School for Gifted Students,
VNUHCM, Vietnam

Hieu Le Duc Minh
Student, Tran Hung Dao High School, Vietnam

Linh Nguyen Hoang Anh
Faculty of Interdisciplinary Sciences,
VNUHCM - University Of Science, Vietnam

Abstract— **This comprehensive analysis explores the application of logistic regression models in predicting credit card defaults, a critical concern for financial institutions worldwide. Current research demonstrates that logistic regression provides an effective framework for identifying potential defaulters based on demographic, financial, and behavioral features. Studies across various datasets reveal prediction accuracies ranging from 85-90%, with key predictors including payment history, credit utilization, income levels, and demographic factors. While more complex algorithms may offer marginal performance improvements, logistic regression remains valuable for its interpretability and practical implementation advantages in financial risk management systems.**

I. INTRODUCTION

Credit card default prediction represents a crucial application of statistical and machine learning techniques in the financial industry. As consumer credit usage continues to expand globally, financial institutions face increasing challenges in managing default risk effectively. The ability to accurately predict which customers might default on their credit card payments enables banks to make informed lending decisions, implement appropriate risk management strategies, and maintain financial stability. Logistic regression has emerged as a particularly valuable modeling approach due to its interpretability and effectiveness in binary classification problems like default prediction.

The problem of credit card default has significant economic implications for both financial institutions and consumers. When consumers lack rational assessment of their repayment capabilities, they often overestimate their ability to fulfill financial obligations to banks in a timely manner. This not only increases the loan risk exposure for banks but also creates potential credit crises for the consumers themselves[1]. With the proliferation of credit cards in the market, instances of default have become increasingly common, underscoring the importance of reliable prediction models.

Financial institutions traditionally relied on credit scoring models based on demographic information and historical financial behavior. However, the advent of sophisticated machine learning techniques has transformed this field, allowing for more nuanced and accurate predictions. Among these techniques, logistic regression remains a cornerstone approach due to its combination of statistical rigor and practical utility. The model's interpretability provides particular value in the financial sector, where regulatory requirements often necessitate transparent decision-making processes.

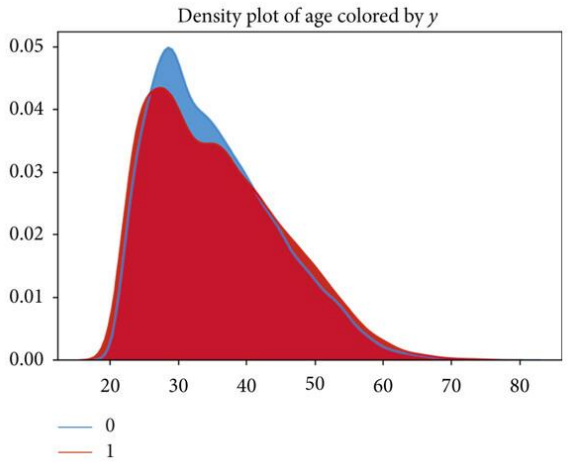

Fig.1.Density diagram of age

International Journal of Engineering Applied Sciences and Technology, 2025
Vol. 9, Issue 11, ISSN No. 2455-2143, Pages 15-22
Published Online March 2025 in IJEAST (http://www.ijeast.com)

II. THE VALUE OF PREDICTIVE MODELING IN FINANCIAL RISK MANAGEMENT

Predictive modeling serves as a critical component in the financial industry's risk management framework. By identifying potential defaulters before they miss payments, institutions can implement proactive measures such as adjusting credit limits, offering payment plans, or providing financial counseling. These interventions not only reduce financial losses but also potentially help consumers avoid damaging their credit histories. The economic significance of accurate default prediction becomes even more apparent when considering the scale of the credit card industry, where even small improvements in prediction accuracy can translate to substantial financial benefits.

III. UNDERSTANDING CREDIT CARD DEFAULT DATA

To develop effective prediction models, researchers and financial analysts work with various datasets containing information on credit card users and their payment behaviors. These datasets typically include a range of features that might influence default probability, from demographic characteristics to detailed payment histories.

(c) (d)

Fig.2.Samples in the data set

IV. COMMON DATASETS FOR DEFAULT PREDICTION

Several standardized datasets have become benchmarks in credit card default research. The UCI Credit Card dataset represents one of the most commonly used resources, containing information on default payments, demographic factors, credit data, payment history, and bill statements from credit card clients in Taiwan[2]. This dataset provides a comprehensive view of consumer behavior and has been extensively used to develop and validate prediction models.

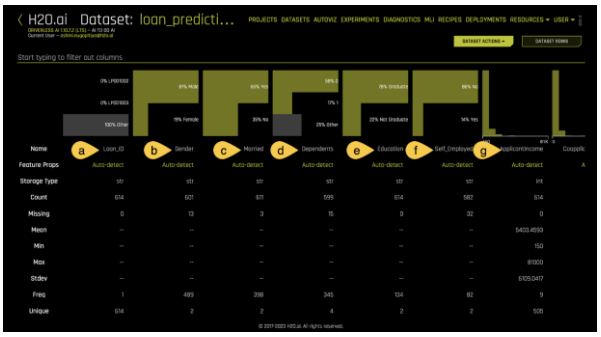

Fig 3. Example of UCI Credit card dataset - Taiwan

Another frequently used dataset is the 'Default' dataset from the ISLR package in R programming language. This dataset contains information on 10,000 credit card users and includes variables such as student status, income, balance, and a binary indicator of whether each user has defaulted on their credit card. The structured nature of this dataset makes it particularly suitable for educational purposes and algorithm development.

In practical applications, financial institutions typically work with proprietary datasets that include detailed customer information collected over time. These proprietary datasets may contain more comprehensive feature sets than publicly available alternatives, potentially improving prediction accuracy in real-world scenarios. For example, a study using credit scoring data from a Portuguese financial institution demonstrated the effectiveness of logistic regression in predicting consumer loan defaults with high accuracy[4].

Coefficients	Estimate	Std. Error	*z* value	*p*-value
(Intercept)	−4.293	0.951	−4.513	6.40e−06***
Spread	0.352	0.103	3.427	0.001***
Term	0.042	0.013	3.121	0.002**
Age	0.043	0.018	3.360	0.001***
Credit Cards	−1.550	0.225	−6.884	5.84e−12***
factor(Salary)1	−0.842	0.154	−5.473	4.42e−08***
factor(Tax Echelon)2	−3.235	1.010	−3.203	0.001**
factor(Tax Echelon)3	−3.367	0.716	−4.700	2.61e−06***
factor(Tax Echelon)4	−2.556	0.514	−4.975	6.54e−07***
factor(Tax Echelon)5	−4.636	1.004	−4.619	3.86e−06***
Null deviance:	1654.7 on 2576 degrees of freedom			
Residual deviance:	1183.2 on 2567 degrees of freedom			
AIC:	1203.2			

Fig 4. logistic regression obtained with Equation

V. KEY FEATURES IN DEFAULT PREDICTION MODELS

The predictive power of default models largely depends on the quality and relevance of the features included. Common variables found in credit card default datasets include:

Demographic information stands as a fundamental component in default prediction models, encompassing factors such as age, gender, education level, and marital status. Research indicates that age correlates positively with default risk,

International Journal of Engineering Applied Sciences and Technology, 2025
Vol. 9, Issue 11, ISSN No. 2455-2143, Pages 15-22
Published Online March 2025 in IJEAST (http://www.ijeast.com)

meaning older customers may exhibit higher propensity to default under certain conditions[4]. Education level typically appears as a categorical variable with designations like graduate school, university, high school, and others, providing insights into the relationship between educational attainment and financial responsibility. Marital status, generally categorized as married, single, or other, can reflect life stability and financial obligations that might influence repayment behavior[5].

Financial indicators provide direct measures of a customer's economic situation and credit utilization patterns. Income represents a primary consideration, as it fundamentally determines a person's ability to service debt. Credit limit allocation, which financial institutions determine based on their assessment of a customer's creditworthiness, offers insights into how the institution itself has evaluated the individual's financial reliability. Outstanding balance amounts reveal actual credit utilization and potential financial strain when they approach or exceed recommended thresholds. Research has demonstrated that clients in lower income tax brackets exhibit greater propensity to default, highlighting the importance of income assessment in prediction models[4].

Payment history variables typically incorporate the most valuable predictive information for default models. Repayment status records over multiple months indicate whether payments have been made on time, and if not, the severity of delinquency. Bill statement amounts across several months reveal spending patterns and potential changes in financial behavior. Previous payment amounts demonstrate a customer's history of debt management and ability to reduce outstanding balances. The temporal nature of these variables adds particular value, as trends and patterns over time often provide stronger signals than static snapshots of a customer's financial position[5].

VI. LOGISTIC REGRESSION FRAMEWORK FOR DEFAULT PREDICTION

Logistic regression represents a statistical modeling approach specifically designed for binary classification problems, making it naturally suited for credit card default prediction where the outcome variable has two possible states: default or no default. The technique belongs to the family of generalized linear models and has become widely adopted in the financial industry due to its interpretability and effectiveness.

6.1. Mathematical Foundation of Logistic Regression

Logistic regression models the probability of default as a function of various predictor variables[3]. Unlike linear regression, which produces continuous outputs that can extend beyond the range, logistic regression employs a logistic function (sigmoid) that constrains predictions between 0 and 1, making them interpretable as probabilities. The model predicts the probability of default using the following mathematical formulation:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n)}}$$

Where P(Y1|X) represents the probability of default given the feature set X, β_0 is the intercept, and β_1 through β_n are the coefficients corresponding to each predictor variable. The model is typically trained using maximum likelihood estimation, which finds the coefficient values that maximize the likelihood of observing the actual default outcomes in the training data.

The coefficients in a logistic regression model have direct interpretations in terms of odds ratios. Specifically, the exponentiated coefficient $\exp(\beta_i)$ represents the multiplicative change in the odds of default associated with a one-unit increase in the corresponding predictor variable, holding all other variables constant. This interpretability represents a significant advantage for financial institutions that need to explain their decision-making processes to regulators and customers[3].

6.2. Advantages of Logistic Regression in Default Prediction

Logistic regression offers several compelling advantages for credit card default prediction. Its interpretability allows financial analysts to understand and explain which factors contribute most significantly to default risk. This transparency proves invaluable in regulatory environments that require justifiable and non-discriminatory lending practices. Additionally, logistic regression performs well even with relatively modest sample sizes and provides probability estimates rather than just classifications, allowing for more nuanced risk assessment.

The computational efficiency of logistic regression also makes it practical for large-scale applications. Unlike more complex algorithms that may require substantial computational resources, logistic regression models can be trained quickly and deployed in real-time decision making systems. This efficiency becomes particularly important when models need frequent retraining to adapt to changing economic conditions or consumer behaviors[4].

VII. BUILDING AN EFFECTIVE LOGISTIC REGRESSION MODEL

The development of an effective logistic regression model for credit card default prediction involves several critical steps, from data preprocessing to model validation. Each stage requires careful consideration to ensure the resulting model provides accurate and reliable predictions.

7.1. Data Preprocessing and Feature Engineering

Data preprocessing represents a foundational step in model development, encompassing activities such as handling missing values, addressing data inconsistencies, and ensuring

proper data types across variables. Credit card datasets often contain missing or erroneous entries that require appropriate treatment before modeling. Common approaches include imputing missing values based on statistical measures like mean or median, removing observations with excessive missing data, or employing more sophisticated imputation techniques that preserve the relationships between variables[5].

Feature engineering involves creating new variables or transforming existing ones to improve model performance. For credit card default prediction, potentially valuable derivations include creating ratios between different financial indicators (such as payment amount to bill amount), calculating aggregate statistics across time periods, or generating categorical variables that capture specific risk thresholds. These engineered features often provide additional predictive power beyond the raw variables available in the dataset.

The process also includes encoding categorical variables appropriately for use in logistic regression. Common approaches include one-hot encoding for nominal variables with no inherent ordering (such as marital status) and ordinal encoding for variables with natural hierarchies (such as education level). Proper encoding ensures the model can effectively utilize the information contained in categorical predictors[5].

7.2. Model Training, Validation, and Evaluation

Model training involves fitting the logistic regression to a subset of the available data, typically 70-80% of the observations. During this process, the algorithm determines the optimal coefficients for each predictor variable by maximizing the likelihood function. Various regularization techniques, such as L1 (Lasso) or L2 (Ridge) regularization, may be employed to prevent overfitting, particularly when dealing with many predictor variables or limited sample sizes.

Validation requires evaluating the model's performance on data not used during training. Common validation approaches include k-fold cross-validation, which divides the data into k subsets and iteratively uses k-1 subsets for training and the remaining subset for validation. This approach provides a robust estimate of how the model will perform on new, unseen data.

Several metrics serve to evaluate classification model performance, each emphasizing different aspects of predictive capability. Accuracy measures the overall percentage of correct predictions but may provide misleading results when dealing with imbalanced classes—a common situation in default prediction where defaulters typically represent a minority class. Precision quantifies the proportion of predicted defaults that actually defaulted, while recall measures the proportion of actual defaults that were correctly identified. The F1-score combines precision and recall into a single metric, providing a balanced assessment for imbalanced datasets.

Area Under the Receiver Operating Characteristic Curve (AUC-ROC) represents another valuable metric that evaluates model performance across various classification thresholds. A study using logistic regression for consumer loan default prediction achieved 89.79% accuracy, demonstrating the effectiveness of this approach when properly implemented[4].

VIII. KEY PREDICTORS IN CREDIT CARD DEFAULT MODELS

Research across various credit card default studies has identified several variables that consistently demonstrate significant predictive power. Understanding these key predictors helps both in model development and in gaining insights into the factors that influence default behavior.

8.1. Demographic and Financial Indicators

Age has emerged as a significant predictor in multiple studies, with research indicating that default risk tends to increase with the customer's age[4]. This somewhat counterintuitive finding might reflect changing financial responsibilities and obligations across different life stages. Education level also shows correlation with default probability, potentially serving as a proxy for long-term earning potential and financial literacy.

Income represents a fundamental predictor, with lower income brackets generally associated with higher default probabilities. A study of Portuguese credit data specifically identified customers in the lowest income tax echelon as having elevated default risk[4]. Credit limit allocation, typically determined by the financial institution based on perceived creditworthiness, also provides predictive information beyond other financial indicators.

The relationship between marital status and default behavior varies across different cultural and economic contexts, making it important to evaluate this factor within the specific population being modeled. Some studies suggest single individuals may have different default patterns compared to married counterparts, possibly reflecting differences in financial responsibilities and stability.

8.2. Payment History and Behavioral Patterns

Payment history variables consistently rank among the strongest predictors of future default. The timeliness of previous payments provides direct evidence of a customer's willingness and ability to meet financial obligations. Repayment status indicators across multiple months enable the identification of patterns such as occasional delinquency versus chronic payment problems.

Outstanding balance represents another critical predictor, with higher balances generally associated with increased default risk. The ratio of balance to credit limit (utilization ratio) provides particularly valuable information, as high utilization may indicate financial strain regardless of the absolute balance amount. A study using the ISLR Default dataset identified

balance as one of the primary predictors of credit card default[3].

Credit card ownership patterns also demonstrate predictive value, with research indicating that customers owning more credit cards tend to have lower default probabilities. This seemingly counterintuitive finding might reflect greater financial sophistication or access to multiple credit sources that can help manage temporary financial challenges.

Behavioral patterns revealed through longitudinal data provide additional predictive power. Changes in spending or payment patterns, seasonal variations in credit utilization, and responses to economic events all contribute information beyond static financial snapshots. These temporal dynamics often require specialized features that capture trends and changes over time rather than point-in-time measurements.

IX. ADDRESSING CLASS IMBALANCE IN DEFAULT PREDICTION

Credit card default datasets typically exhibit significant class imbalance, with defaulting customers representing a small minority of the overall population. This imbalance poses challenges for model training and evaluation, as algorithms naturally tend to favor the majority class (non-defaulters) without appropriate adjustments.

9.1. Impact of Class Imbalance on Model Performance

Class imbalance creates several problems in prediction models. A naive model that simply predicts no default for all cases might achieve high accuracy (e.g., 95% if only 5% of customers default) despite providing no actual predictive value. Standard training procedures may produce models that rarely predict the minority class, resulting in poor recall for defaults even when overall accuracy appears acceptable.

Evaluation metrics can also be misleading under imbalance conditions. Accuracy alone fails to capture whether the model correctly identifies the critical minority class of defaulters. For financial institutions, false negatives (failing to identify potential defaulters) often carry significantly higher costs than false positives, making balanced prediction particularly important despite the natural imbalance in the data.

9.2. Techniques for Handling Imbalanced Data

Several approaches help address class imbalance in default prediction models. Sampling techniques modify the training data distribution to create more balanced classes. Undersampling reduces the majority class (non-defaulters) to create a more balanced dataset, though potentially discarding valuable information. Oversampling duplicates or creates synthetic instances of the minority class (defaulters) to achieve better balance. The Synthetic Minority Over-sampling Technique (SMOTE) represents a sophisticated approach that creates synthetic examples of the minority class based on feature similarities.

A study on credit card default prediction proposed using k-means SMOTE, which combines clustering with synthetic sample generation, to address imbalance issues. This approach significantly improved prediction performance, increasing the AUC value from 0.765 to 0.929 compared to standard methods[4]. The technique first identifies meaningful clusters within the minority class and then generates synthetic samples within these clusters, creating more representative and diverse synthetic examples.

Alternative approaches include algorithm-level methods that modify the learning process to account for class imbalance. Cost-sensitive learning assigns higher misclassification costs to the minority class, effectively forcing the algorithm to pay more attention to defaulters during training. Ensemble methods like balanced random forests or ensemble combinations of undersampled datasets can also effectively handle imbalanced data while maintaining predictive performance.

X. COMPARATIVE PERFORMANCE WITH OTHER MACHINE LEARNING MODELS

While logistic regression provides a solid foundation for default prediction, comparing its performance with other machine learning approaches offers valuable insights into potential tradeoffs between interpretability and predictive power.

10.1. Performance Metrics Across Different Algorithms

Research comparing various machine learning algorithms for credit card default prediction shows interesting performance patterns. A study evaluating multiple models found that while more complex algorithms sometimes achieve marginally better predictive performance, logistic regression remains competitive across most evaluation metrics[1]. The study reported logistic regression achieving respectable AUC values, though slightly lower than ensemble methods like random forests or gradient boosting machines.

In terms of specific metrics, logistic regression typically demonstrates balanced performance across precision and recall, making it well-suited for default prediction where both false positives and false negatives carry significant costs. Ensemble methods like random forests often achieve higher overall accuracy and AUC, but may require more careful tuning to balance precision and recall appropriately for the business context.

Neural network approaches, including the BP neural network methodology proposed in one study, can achieve superior performance when properly configured and trained on sufficient data. The research demonstrated that a BP neural network with features weighted according to their importance calculated by random forest achieved excellent predictive performance, outperforming other common machine learning models including KNN, logistic regression, SVM,and decision trees[1].

10.2. Interpretability vs. Performance Tradeoffs

The choice between logistic regression and more complex models often involves tradeoffs between interpretability and pure predictive performance. Logistic regression produces coefficients with clear interpretations regarding how each variable influences default probability. These interpretations serve crucial business purposes, including regulatory compliance, customer communication, and strategic decision-making about credit policies.

More complex models like neural networks or ensemble methods may achieve marginally better predictive performance but function essentially as "black boxes" where the relationship between inputs and outputs becomes difficult to interpret. This lack of transparency can create challenges in regulated financial environments where decisions affecting consumers must be explainable and justifiable.

Financial institutions often adopt a pragmatic approach that balances these considerations. Logistic regression might serve as the primary model for decision-making due to its interpretability, while more complex models provide supplementary insights or handle specific segments where additional predictive power justifies the reduced interpretability. This hybrid approach leverages the strengths of different modeling paradigms while mitigating their respective limitations.

XI. IMPLEMENTATION IN FINANCIAL RISK MANAGEMENT SYSTEMS

The practical deployment of logistic regression models for credit card default prediction involves integrating them into broader risk management frameworks and operational systems within financial institutions.

11.1. Integration with Credit Scoring Systems

Default prediction models typically function as components within comprehensive credit scoring systems that inform lending decisions and account management. These integrated systems combine multiple risk assessments, including application scores (evaluating new applicants), behavior scores (monitoring existing customers), and collection scores (managing delinquent accounts). Logistic regression models can serve various roles within this ecosystem, from initial credit approval to ongoing portfolio management.

The outputs from logistic regression models—default probabilities—are often converted into credit scores using appropriate scaling and transformation functions. These scores provide standardized risk measures that business users can easily interpret and apply in decision making processes. Score cutoffs establish thresholds for different actions, such as approval, rejection, or manual review of applications, with these thresholds calibrated to align with the institution's risk appetite and business objectives.

Integration also requires establishing feedback loops that continuously validate and improve model performance. As new default data becomes available, models undergo periodic recalibration and validation to ensure they maintain predictive power despite changing economic conditions or customer behaviors. This ongoing refinement process ensures the models remain effective tools for risk management over time.

11.2. Regulatory Considerations and Model Governance

Financial institutions must navigate various regulatory requirements when implementing default prediction models. Regulations like the Equal Credit Opportunity Act (ECOA) in the United States prohibit discrimination in lending based on protected characteristics such as race, gender, or marital status. Logistic regression models require careful evaluation to ensure they don't inadvertently create disparate impacts on protected groups, even when these characteristics aren't explicitly included as predictor variables.

Model governance frameworks establish processes for model development, validation, and monitoring throughout the model lifecycle. These frameworks ensure that models meet statistical quality standards, comply with regulatory requirements, and align with business objectives. Documentation requirements typically include detailed explanations of model methodology, variable selection rationale, performance metrics, and limitations, with logistic regression's interpretability making these explanations more straightforward compared to complex algorithms.

Independent model validation serves as a critical component of governance, with separate teams evaluating models to confirm their statistical validity, implementation accuracy, and ongoing performance. This validation provides an additional safeguard against potential issues that might affect model reliability or regulatory compliance. The transparent nature of logistic regression facilitates this validation process, as reviewers can directly assess the reasonableness of coefficients and their business implications.

XII. FUTURE DIRECTIONS IN CREDIT CARD DEFAULT PREDICTION

The field of credit card default prediction continues to evolve with advances in data availability, algorithmic approaches, and computational capabilities. Several emerging trends point to future directions that may enhance predictive capabilities while addressing current limitations.

12.1. Incorporating Alternative Data Sources

Traditional default prediction models rely primarily on demographic information, financial indicators, and payment history. However, the increasing availability of alternative data sources offers opportunities to enhance predictive power. Digital footprints, including online behavior patterns, device usage, and application interaction data, may provide signals about financial responsibility not captured in conventional data. Mobile phone data, including call patterns and payment consistency for telecommunication services, have

demonstrated predictive value in markets where traditional credit data remains limited.

Psychometric assessments represent another innovative data source, measuring traits like conscientiousness, risk perception, and time preferences that correlate with repayment behavior. While adoption of such alternative data requires careful ethical and regulatory consideration, these sources may prove particularly valuable for evaluating "thin-file" customers with limited traditional credit histories.

12.2. Methodological Innovations and Hybrid Approaches

Advanced methodological approaches continue to emerge, potentially offering improvements over traditional logistic regression. Deep learning techniques can automatically extract complex patterns from raw data, potentially capturing subtle interactions and non-linear relationships that parametric models might miss. Explainable AI methods seek to combine the predictive power of complex algorithms with the interpretability necessary for financial applications, offering a potential "best of both worlds" solution.

Hybrid modeling approaches that combine multiple techniques have shown promise in recent research. For example, using machine learning for feature selection or transformation before applying logistic regression maintains interpretability while potentially improving predictive performance. The study on k-means SMOTE and BP neural networks demonstrated how combining clustering, synthetic data generation, and neural networks can achieve superior results compared to individual methods[1].

Ensemble approaches that integrate predictions from multiple model types also continue to gain traction. These approaches leverage the strengths of different algorithms while mitigating their individual weaknesses, potentially achieving both higher performance and greater robustness across different data segments or economic conditions.

XIII. CONCLUSION: THE ENDURING VALUE OF LOGISTIC REGRESSION

Despite rapid advances in machine learning and artificial intelligence, logistic regression remains a foundational approach for credit card default prediction with enduring relevance in the financial industry. Its combination of solid predictive performance, statistical rigor, and interpretability continues to provide value in real-world applications where decisions must be both accurate and explainable.

The studies examined in this report demonstrate logistic regression's effectiveness across various datasets and financial contexts. Research using Portuguese credit scoring data[4] achieved 89.79% accuracy in default prediction using logistic regression, while comparative studies show the model performing competitively with more complex algorithms across most evaluation metrics. These results confirm that logistic regression provides reliable predictive power for default risk assessment when properly implemented.

Looking forward, logistic regression will likely continue evolving within hybrid frameworks that preserve its interpretability advantages while incorporating elements from newer methodological approaches. As financial institutions navigate increasingly complex regulatory environments while seeking to maintain competitive risk management capabilities, the transparent yet effective nature of logistic regression ensures its continued relevance in credit card default prediction and broader financial risk management applications.

The practical value of logistic regression ultimately stems from its ability to transform complex customer data into actionable insights that inform sound financial decisions. By identifying key predictors of default risk and quantifying their impacts through interpretable coefficients, these models enable institutions to make more informed lending decisions, implement appropriate risk mitigation strategies, and maintain financial stability—a valuable contribution to both institutional success and broader economic health.

REFERENCES

[1] Chen, 2021, “Research on credit card default prediction based on K-means smote and BP neural network”, https://onlinelibrary.wiley.com/doi/10.1155/2021/6618841.

[2] F. N. Khan, A. H. Khan, and Lamiah Israt, “Credit Card Fraud Prediction and Classification using Deep Neural Network and Ensemble Learning,” 2017 IEEE Region 10 Symposium (TENSYMP), pp. 114–119, Jan. 2020, doi: https://doi.org/10.1109/tensymp50017.2020.9231001.

[3] O. Zhang, “Utilizing a Logistic Regression Model to Predict Credit Card Default,” Ou Zhang, Oct. 27, 2020. https://ouzhang.rbind.io/2020/10/27/logistic-regression-model-predict-credit-card-default/.

[4] E. Costa e Silva, I. C. Lopes, A. Correia, and S. Faria, “A logistic regression model for consumer default risk,” Journal of Applied Statistics, vol. 47, no. 13–15, pp. 2879–2894, May 2020, doi: https://doi.org/10.1080/02664763.2020.1759030.

[5] iambitttu, “GitHub - iambitttu/Credit-Card-Default-Prediction: This project involved data preprocessing, model building, and deployment of a machine learning model to predict credit card default.,” GitHub, 2023. https://github.com/iambitttu/Credit-Card-Default-Prediction.

[6] J. Gao, W. Sun, and X. Sui, “Research on Default Prediction for Credit Card Users Based on XGBoost-LSTM Model,” Discrete Dynamics in Nature and Society, vol. 2021, pp. 1–13, Dec. 2021, doi: https://doi.org/10.1155/2021/5080472.

[7] I-Cheng. Yeh, “UCI Machine Learning Repository,” archive.ics.uci.edu, Jan. 25, 2016.

https://archive.ics.uci.edu/dataset/350/default+of+credit+card+clients

[8] [8] A. Arram et al., "Credit card score prediction using machine learning models: A new dataset." Available: https://arxiv.org/pdf/2310.02956

[9] R. Bhandary and B. K. Ghosh, "Credit Card Default Prediction: An Empirical Analysis on Predictive Performance Using Statistical and Machine Learning Methods," Journal of Risk and Financial Management, vol. 18, no. 1, p. 23, Jan. 2025, doi: https://doi.org/10.3390/jrfm18010023.

[10] C. Egan, "Improving Credit Default Prediction Using Explainable AI MSc Research Project Data Analytics Improving Credit Default Prediction Using Explainable AI." Available: https://norma.ncirl.ie/5146/1/ciaranegan.pdf