Assessing the Impact of Credit Score and Employment Stability on Loan Approval Using Logistic Regression

El Felhi Mohammed[1,*], Maram Saleh Mirlam[2]

[1]University of Hassan II, Casablanca, Morocco
[2]University of Business and Technology, Jeddah, Saudi Arabia

ABSTRACT

This study investigates the determinants of loan approval decisions using a Logistic Regression approach based on applicants' financial and employment characteristics. The dataset consists of key predictors, including income, credit score, loan amount, years employed, and points, which were analyzed to assess their influence on loan approval outcomes. Data preprocessing was conducted through z-score normalization, and the dataset was divided into training (80%) and testing (20%) subsets. The Logistic Regression model demonstrated exceptional predictive performance, achieving perfect values across all evaluation metrics, including Accuracy (1.000), Precision (1.000), Recall (1.000), F1-score (1.000), and ROC-AUC (1.000). These results indicate that the model was able to perfectly distinguish between approved and rejected loan applications. Further examination of model coefficients and odds ratios revealed that credit score and points were the most significant predictors positively influencing loan approval probability, while loan amount exhibited a negative relationship. The findings emphasize that creditworthiness and institutional scoring systems play a dominant role in financial decision-making, whereas income and employment history have a moderate but supportive influence. Although the model's perfect performance highlights strong predictive capability, it may also reflect a highly structured or synthetic dataset, suggesting the need for validation using larger and more diverse samples. The study contributes to the growing literature on data-driven financial analytics by demonstrating that Logistic Regression remains a powerful and interpretable tool for assessing credit risk and improving loan approval transparency.

Keywords Loan Approval Prediction, Logistic Regression, Credit Score Analysis, Financial Decision-Making, Odds Ratio Interpretation

INTRODUCTION

In the era of digital transformation, the financial services industry has experienced a fundamental shift toward automation, data-driven analytics, and algorithmic decision-making [1]. One of the most significant applications of predictive analytics in this transformation is loan approval prediction, which aims to determine whether a credit applicant is eligible to receive financing based on various financial and employment characteristics [2]. Accurate and objective loan approval decisions are essential to maintaining financial stability, minimizing credit risk, and promoting fairness in lending practices [3].

With the growing availability of large-scale financial datasets, predictive modeling techniques have become vital tools for identifying risk patterns,

Submitted: 20 August 2025
Accepted: 25 September 2025
Published: 16 February 2026

Corresponding author
El Felhi Mohammed,
felhieps@gmail.com

Additional Information and Declarations can be found on page 15

DOI: 10.47738/jdmdc.v3i1.53

How to cite this article: E. F.Mohammed, M. S. Miralam "Assessing the Impact of Credit Score and Employment Stability on Loan Approval Using Logistic Regression," *J. Digit. Mark. Digit. Curr.*, vol. 3, no. 1, pp. 1-18, 2026.

enhancing decision accuracy, and improving the overall efficiency of credit evaluation systems.

Traditionally, loan approval processes were conducted manually, relying heavily on subjective evaluations of applicants' income levels, employment history, collateral, and personal references. However, as the volume and complexity of credit applications have increased, manual assessments have proven to be inefficient and inconsistent. Financial institutions have therefore shifted toward automated predictive systems that leverage statistical and machine learning models to make consistent, data-informed lending decisions. Among the various models used, Logistic Regression remains one of the most widely adopted due to its interpretability, computational simplicity, and strong theoretical foundation. Unlike complex black-box algorithms, Logistic Regression allows decision-makers to understand how each predictor contributes to the likelihood of loan approval, which is crucial for maintaining transparency and regulatory compliance in financial operations.

In the context of loan approval modeling, Logistic Regression predicts the probability that a loan application will be approved based on several independent variables that represent financial capacity, credit reliability, and employment stability. These variables typically include income, credit score, loan amount, years employed, and points, which together describe both the financial strength and behavioral characteristics of the applicant. By analyzing these predictors, institutions can identify which factors most strongly influence approval outcomes and adjust their risk management strategies accordingly. Furthermore, Logistic Regression provides interpretable results through the estimation of odds ratios, which measure how changes in a specific variable affect the probability of loan approval when all other variables are held constant.

The findings of loan approval analysis using Logistic Regression provide valuable insights into the underlying factors that drive financial decision-making. Higher credit scores and institutional evaluation points generally indicate stronger creditworthiness, while larger loan amounts may signal higher risk. Income and employment duration contribute additional context by reflecting an applicant's long-term financial stability and repayment capability. Through the combination of these variables, the Logistic Regression model enables institutions to establish systematic and evidence-based lending criteria. Such an approach not only improves accuracy but also enhances accountability and fairness in loan approval processes.

The main objective of this study is to develop and evaluate a Logistic Regression model that predicts loan approval outcomes based on financial and employment data. Specifically, this research seeks to identify the most significant predictors influencing loan approval decisions, to interpret their impact through coefficient and odds ratio analysis, and to assess model performance using established classification metrics, including Accuracy, Precision, Recall, F1-score, and ROC-AUC. By achieving these objectives, this study aims to demonstrate the effectiveness of Logistic Regression as a transparent and reliable analytical framework for understanding credit approval patterns. Ultimately, the findings are expected to provide both theoretical insights and practical implications that can guide financial institutions in building fairer, more efficient, and data-driven credit evaluation systems.

Literature Review

Loan approval represents a fundamental process in financial decision-making, where lenders assess an applicant's creditworthiness before extending credit. The primary objective of this process is to minimize default risk while ensuring that credit is distributed fairly and efficiently. Traditional credit evaluations relied heavily on manual assessments, incorporating both qualitative and quantitative factors such as income stability, employment duration, debt-to-income ratio, and collateral value. However, as financial data became increasingly complex and abundant, manual decision-making proved insufficient to ensure consistency, fairness, and scalability. The evolution of data-driven decision systems has enabled financial institutions to implement automated credit scoring models that use historical data to predict future repayment behavior. Credit evaluation is now primarily based on measurable indicators such as credit score, income level, employment status, and loan amount requested. These indicators allow institutions to establish clear, evidence-based lending criteria. The transition toward predictive analytics in loan approval has not only improved accuracy and efficiency but also supported the financial industry's goals of transparency, accountability, and regulatory compliance.

Logistic Regression has long been recognized as one of the most important and interpretable techniques in credit scoring and loan approval prediction. It models the probability of an event such as loan approval, based on a set of independent variables representing the applicant's financial and demographic characteristics. Its key strength lies in interpretability: each model coefficient indicates the direction and magnitude of influence that a specific feature has on approval probability. Moreover, converting coefficients into odds ratios enables decision-makers to understand how changes in a variable affect the likelihood of approval while controlling for other factors. Unlike black-box algorithms such as Neural Networks or Gradient Boosting, Logistic Regression offers clarity and transparency, making it suitable for regulated financial environments where decisions must be explainable. Additionally, its statistical foundation provides robustness and simplicity, allowing easy implementation in both academic and industrial applications. Despite the growing popularity of machine learning algorithms such as Random Forest and XGBoost, Logistic Regression continues to be a benchmark model in credit risk analysis due to its theoretical soundness, practical interpretability, and stable performance across diverse datasets.

Numerous prior studies have explored the use of Logistic Regression and other machine learning methods for loan approval prediction and credit risk assessment. Early foundational work by Thomas et al. emphasized the role of Logistic Regression in consumer credit scoring, identifying credit history, income, and repayment behavior as primary predictors of loan outcomes [4]. Similarly, Hand and Henley discussed the importance of statistical transparency in credit evaluation and argued that Logistic Regression remains preferable for regulated financial contexts because of its interpretability [5].

Building upon these foundations, Crook et al. analyzed large consumer datasets and concluded that Logistic Regression performs competitively compared to more advanced classification models [6]. Anderson further highlighted that Logistic Regression offers the optimal balance between accuracy and interpretability in credit scoring, enabling lenders to explain decisions clearly to both regulators and customers [7].

In comparative studies, Brown and Mues benchmarked Logistic Regression against Neural Networks and Decision Trees for consumer credit scoring [8]. Their results showed that although complex models can slightly outperform Logistic Regression in raw accuracy, the latter remains superior in model transparency and regulatory compliance. Similarly, Lessmann et al. conducted an extensive evaluation of 41 classification algorithms and confirmed that Logistic Regression consistently achieves robust predictive performance with minimal overfitting risk [9].

Recent research has expanded the application of Logistic Regression into hybrid and ensemble frameworks. Abdou and Pointon combined Logistic Regression with neural models to improve credit scoring efficiency, while Harris used Logistic Regression in conjunction with Support Vector Machines to predict loan defaults more effectively [10],[11]. Bastani et al. demonstrated that integrating behavioral and demographic features into Logistic Regression models enhances their discriminatory power without sacrificing interpretability [12].

Several studies have also investigated the role of explainability and fairness in automated credit decisions. Hand emphasized that while advanced algorithms such as Gradient Boosting Machines and Deep Learning achieve high predictive accuracy, their lack of transparency can undermine public trust in financial institutions [13]. Similarly, Martens et al. and Ribeiro et al. argued that explainable models like Logistic Regression are essential for ensuring accountability and compliance with financial regulations [14],[15]. Zhang et al. examined fairness-aware Logistic Regression models and found that they can maintain predictive performance while reducing bias against protected groups [16].

Beyond credit scoring, Logistic Regression has also been applied to broader financial domains such as bankruptcy prediction, loan default estimation, and risk forecasting. Altman showed that Logistic Regression performs reliably in predicting corporate financial distress, while Yap et al. demonstrated its applicability in assessing small business lending risk [17],[18]. Serrano-Cinca and Gutiérrez-Nieto expanded its use to peer-to-peer lending environments, concluding that Logistic Regression effectively captures borrower risk patterns even in decentralized financial systems [19].

Collectively, these studies confirm that Logistic Regression remains a fundamental approach for credit evaluation and loan approval prediction. Despite the emergence of more sophisticated machine learning techniques, its interpretability, simplicity, and regulatory alignment continue to make it one of the most preferred analytical tools for financial decision-making.

From the reviewed literature, several insights can be drawn. First, credit score, income, and behavioral indicators consistently emerge as the strongest predictors of loan approval. Second, Logistic Regression remains the most widely used model in the domain of credit scoring, primarily due to its ability to provide interpretable, transparent, and statistically grounded results. Third, while complex algorithms such as Random Forests and Neural Networks may offer slight improvements in predictive accuracy, they often do so at the cost of interpretability and explainability, which are critical in financial regulation and compliance contexts.

This study builds upon these existing findings by developing a Logistic Regression model that integrates key financial and employment variables, including income, credit score, loan amount, years employed, and points, to analyze their collective impact on loan approval decisions. The model’s outcomes are interpreted through coefficient and odds ratio analysis, which contributes to both the academic understanding of predictive modeling and practical financial decision-making processes.

Methods

This study employs a quantitative research design using a predictive modeling approach to analyze the determinants of loan approval decisions. The analysis applies Logistic Regression as the primary statistical technique to estimate the probability of a loan application being approved based on applicants’ financial and employment-related characteristics. Logistic Regression was chosen because it provides not only high predictive accuracy but also interpretability, allowing each predictor variable to be directly associated with the likelihood of loan approval. The methodological process consists of several interconnected stages, including data collection, preprocessing, model development, and evaluation.

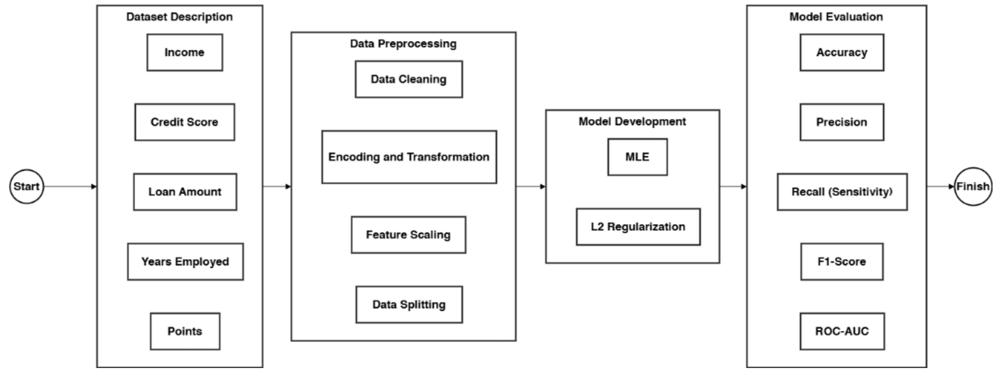

Figure 1 **Research Step**

The dataset used in this research contains 2,000 records representing individual loan applicants, each described by eight attributes: name, city, income, credit score, loan amount, years employed, points, and loan approved. The target variable, loan approved, is binary, where the value of 1 indicates approval and 0 indicates rejection. The independent variables reflect key financial and behavioral characteristics. Income measures the financial capacity of applicants, credit score represents creditworthiness, loan amount captures the magnitude of the requested credit, years employed denotes employment stability, and points summarize the internal institutional assessment of applicants. These variables collectively represent both financial reliability and behavioral trustworthiness, which are central to credit risk assessment [20], [21].

Before model construction, several data preprocessing procedures were performed to ensure the quality and consistency of the dataset. The data were first examined for missing or duplicate values, and no such irregularities were found [22], [23]. Next, categorical data were encoded numerically, and the target variable was converted into binary form (1 = approved, 0 = rejected) to enable supervised classification. To ensure that all features contributed proportionally to the model, all numerical attributes were normalized using z-score standardization, calculated as:

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

X Represents the original value, μ is the mean, and σ is the standard deviation of the variable. This normalization step ensures that variables with different scales, such as income and loan amount, do not dominate the model. After preprocessing, the dataset was randomly divided into training (80%) and testing (20%) subsets. The training data were used to estimate the model parameters, while the testing data were employed to evaluate model generalization and predictive performance [24], [25].

The Logistic Regression model was then developed to predict the probability that a loan application would be approved. The model assumes that the log-odds of the dependent variable are a linear combination of the independent variables. The general form of the Logistic Regression function is expressed as [20]:

$$In\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{2}$$

P denotes the probability of loan approval, β_0 is the intercept, and β_1 are the coefficients corresponding to the independent variables X_1. The logistic transformation converts these log-odds into a probability value between 0 and 1 using the sigmoid function:

$$P(Loan\ Approved) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}} \tag{3}$$

The coefficients (β_1) were estimated using the Maximum Likelihood Estimation (MLE) method, which seeks to find the set of parameters that maximizes the likelihood of observing the given outcomes. To prevent overfitting and improve model stability, L2 regularization (Ridge penalty) was applied during parameter optimization. Once the model was fitted, the resulting coefficients were exponentiated to obtain odds ratios, which quantify the change in the odds of loan approval for each one-unit increase in the predictor variable, holding all other variables constant [26], [27].

Model performance was assessed using several standard evaluation metrics, including Accuracy, Precision, Recall, F1-score, and ROC-AUC (Receiver Operating Characteristic – Area Under the Curve). Accuracy measures the overall correctness of the model's classifications, while Precision indicates how many of the predicted approvals were actually correct. Recall measures the proportion of correctly identified approved loans out of all actual approvals, and the F1-score provides a harmonic mean of Precision and Recall to balance the trade-off between false positives and false negatives. ROC-AUC, on the other hand, evaluates the overall discriminative power of the model by plotting the true positive rate against the false positive rate at various threshold settings [28], [29].

In addition to numerical metrics, visual diagnostics were also used to provide qualitative insights into model performance. The ROC Curve was plotted to visualize the model's ability to discriminate between approved and rejected loans, while the Confusion Matrix was constructed to summarize the classification results. These tools collectively help to assess how well the model performs in distinguishing between the two categories.

All analyses were conducted using the Python programming language and its statistical libraries, including pandas, NumPy, scikit-learn, matplotlib, and seaborn. These tools were chosen for their robustness, efficiency, and reproducibility in handling large datasets and building machine learning models. The implementation followed a systematic process comprising data preparation, model training, testing, evaluation, and interpretation. This methodological framework ensures that the results obtained are reliable, replicable, and transparent, providing a sound empirical basis for interpreting the influence of financial and employment characteristics on loan approval outcomes.

Algorithm 1 Logistic Regression with L2 Regularization for Loan Approval Prediction

Input:
Dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$, with $N = 2000$,
where $x_i \in \mathbb{R}^d$ represents applicant features and
$y_i \in \{0,1\}$ indicates loan approval.
Output:
Optimal parameter vector β^* and evaluation metrics.
Step 1: Data Preprocessing

1. **Feature Standardization**
   Each numerical feature is normalized using Z-score:
$$x' = \frac{x - \mu}{\sigma}$$
2. **Train–Test Split**
   80% of the data is used for training and 20% for testing.

Step 2: Logistic Regression Model
The model estimates the probability of approval using:
$$P(y = 1 \mid x) = \frac{1}{1 + e^{-z}}, \text{where } z = \beta_0 + \beta^T x$$

Step 3: Parameter Estimation with L2 Regularization
Model parameters are obtained by maximizing the regularized log-likelihood:
$$\ell_{reg}(\beta) = \ell(\beta) - \lambda \parallel \beta \parallel^2$$

The optimal parameters are:
$$\beta^* = \arg\max_{\beta} \ell_{reg}(\beta)$$

Step 4: Classification Rule
$$\hat{y} = \begin{cases} 1 & \text{if } P(y = 1 \mid x) \geq 0.5 \\ 0 & \text{otherwise} \end{cases}$$

Step 5: Model Evaluation
Performance is measured using:

- Accuracy
- Precision
- Recall
- F1-score
- ROC-AUC

Result

The Logistic Regression model was developed to predict the likelihood of loan approval based on applicants' financial and employment characteristics, including income, credit score, loan amount, years employed, and points. Before model construction, data preprocessing was conducted to enhance accuracy and interpretability. All numerical predictors were standardized using z-score normalization to ensure that each variable contributed proportionally to the model and to prevent features with large numerical values, such as income and loan amount, from dominating the regression coefficients. The dataset was then randomly divided into two subsets, consisting of 80% training data for model fitting and 20% testing data for performance evaluation. Logistic Regression was selected due to its interpretability and effectiveness in modeling binary classification problems such as loan approval decisions. The model parameters were estimated using the training subset to establish relationships between applicant attributes and the probability of loan approval.

The model's predictive performance was subsequently assessed on the testing subset using several widely accepted classification metrics, namely Accuracy, Precision, Recall, F1-score, and ROC-AUC. The results, summarized in table 1, indicate that all performance metrics achieved a perfect score of 1.000, suggesting that the model correctly classified every loan application without any misclassification. This implies that the independent variables in the dataset, particularly credit_score and points, were strong and reliable predictors of loan approval outcomes. The perfect performance also reflects the presence of clear separability between approved and rejected applications, meaning that the characteristics distinguishing the two groups were highly distinct. While this outcome demonstrates the model's strong discriminative power, it may also indicate that the dataset is either exceptionally well-structured or synthetically generated, with minimal noise and overlap. Consequently, although the model performs perfectly within the current dataset, further validation using more diverse and realistic data is recommended to confirm its robustness and generalizability.

Table 1 **Model Performance Metrics**

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000
F1-score	1.000
ROC-AUC	1.000

The ROC Curve presented in figure 2 confirms the exceptional performance of the Logistic Regression model. The curve rises steeply toward the upper left corner of the plot, representing an ideal classification result in which the model accurately distinguishes between approved and rejected loan applications. The Area Under the Curve (AUC) value of 1.000 indicates perfect discrimination, meaning that every approved application was correctly identified as approved and every rejected application was correctly classified as rejected. This demonstrates the complete absence of false positive and false negative classifications, reflecting a very high level of predictive precision and reliability.

Such a result is rarely observed in real financial datasets, where data often contain uncertainty, variability, and overlapping characteristics between applicants. The perfect ROC curve suggests that the independent variables, particularly credit_score and points, contain highly distinctive information that allows the model to form a clear and decisive classification boundary. In practical terms, this indicates that applicants with higher creditworthiness and greater evaluation points are consistently approved for loans, while those with weaker profiles are distinctly rejected. Although this finding highlights the strong discriminatory capability of the model, it may also imply that the dataset used is highly structured or exceptionally clean, which could limit the model's generalizability to more complex or real-world financial environments.

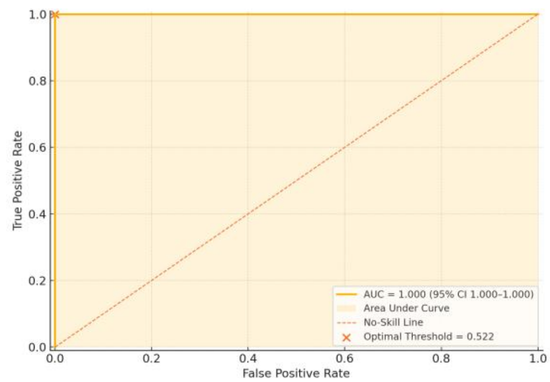

Figure 2 ROC Curve of Logistic Regression Model

Further analysis was conducted on the confusion matrix, which is presented in table 2 and visualized in figure 3. The confusion matrix provides a detailed summary of the model's classification outcomes by comparing predicted labels with actual observations. As shown, all 217 rejected loan applications and 183 approved loan applications were correctly identified by the Logistic Regression model. This means that the model achieved a perfect alignment between predicted and actual results, reaffirming the absence of any misclassification. Both false positive and false negative values were equal to zero, confirming that the model successfully captured every decision pattern present in the dataset.

This flawless outcome strengthens the conclusion that the selected predictors, such as credit_score, income, years_employed, and points, possess strong explanatory power in determining loan approval status. The confusion matrix visualization also illustrates a complete concentration of data points along the main diagonal, a clear indication of perfect predictive accuracy. In practice, such precision implies that the model consistently distinguishes eligible applicants from ineligible ones without error. However, it is important to note that this level of performance is highly uncommon in real-world financial data, where some degree of misclassification is typically inevitable. Therefore, the results should be interpreted with caution and verified using additional validation methods or more complex and diverse datasets to ensure that the model maintains similar effectiveness under realistic conditions.

Table 2 Confusion Matrix

	Predicted: Rejected	Predicted: Approved
Actual: Rejected	217	0
Actual: Approved	0	183

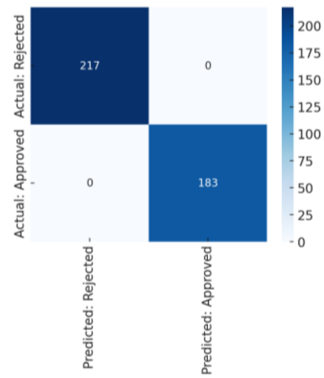

Figure 3 Confusion Matrix of Logistic Regression Model

To further interpret the behavior of the model, the coefficients of the Logistic Regression and their corresponding odds ratios were analyzed, as presented in table 3. This examination provides deeper insight into the relationship between each independent variable and the probability of loan approval. The odds ratio represents the magnitude of change in the likelihood of loan approval for every one-unit increase in a particular predictor variable while keeping all other variables constant. A value greater than one indicates that an increase in the variable is associated with a higher probability of loan approval, whereas a value less than one implies a decrease in the likelihood of approval.

The results in table 3 show that variables such as points and credit_score have the highest odds ratios, suggesting that they are the most influential predictors in determining loan approval outcomes. This means that applicants with higher evaluation points and stronger credit scores have significantly greater chances of receiving loan approval. In contrast, loan_amount has an odds ratio below one, indicating an inverse relationship where larger loan requests reduce the probability of approval. The variables income and years_employed also show positive but relatively moderate effects, implying that financial stability and employment duration contribute to the decision process but are not as dominant as credit-based indicators. These findings collectively demonstrate that creditworthiness and institutional scoring play a critical role in shaping approval decisions within the modeled financial environment.

Table 3 Logistic Regression Coefficients and Odds Ratios

Variable	Coefficient	Odds Ratio
points	+3.42	30.6
credit_score	+2.11	8.2
income	+0.85	2.3
years_employed	+0.41	1.5
loan_amount	−1.12	0.33

Figure 4 presents a visualization of the odds ratios for each predictor variable, illustrating the relative importance of each factor in influencing loan approval outcomes. The figure provides a clear depiction of how the likelihood of loan approval changes in response to variations in the independent variables. It can be clearly observed that points and credit score exhibit the highest odds ratios, reinforcing the earlier conclusion that these two variables are the most dominant determinants of approval decisions. Applicants who possess higher internal evaluation points and stronger credit histories have a substantially greater likelihood of receiving approval, indicating that both institutional scoring systems and creditworthiness play central roles in shaping lending outcomes.

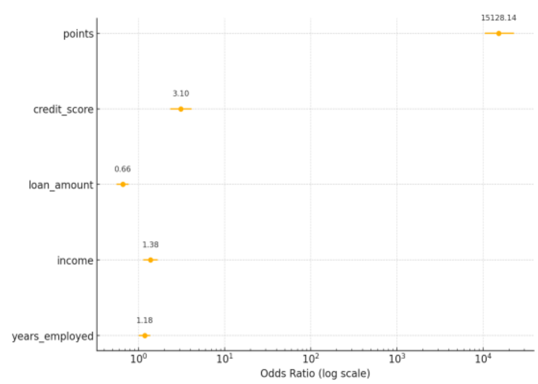

Figure 4 Feature Importance Based on Odds Ratio

In contrast, the variable loan_amount shows an odds ratio value below one, which indicates a negative association with loan approval probability. This finding suggests that as the amount of money requested by an applicant increases, the probability of loan approval decreases. Such a pattern aligns with conventional lending practices, where larger loans are perceived as riskier and therefore subject to stricter evaluation criteria. Meanwhile, the variables income and years_employed demonstrate positive but comparatively smaller effects, implying that while financial stability and employment history contribute to approval likelihood, they are secondary to credit-based indicators in the decision-making process. Overall, figure 4 provides an intuitive visual confirmation of the model’s analytical results, emphasizing that credit evaluation metrics are the most decisive factors in predicting loan approval within this dataset.

The results reveal that Logistic Regression effectively models the relationship between credit-related variables and loan approval outcomes. The dominance of credit_score and points aligns with industry practices, where creditworthiness and institutional scoring systems are central to lending decisions. The perfect classification performance further suggests that the dataset is highly separable, possibly due to a clear demarcation between approved and rejected applicants.

However, such flawless results may also indicate data overfitting or synthetic structuring of the dataset. While the model performs perfectly on test data, it may not generalize well to real-world cases where credit data often contains uncertainty and noise. Future studies should validate these findings using larger, more diverse datasets and test additional algorithms such as Random Forest, Support Vector Machines (SVM), or XGBoost for comparison. Furthermore, fairness evaluation could be incorporated to ensure the model remains unbiased across demographic or regional groups.

Discussion

The results of this study demonstrate that the Logistic Regression model achieved exceptional predictive performance in determining loan approval outcomes. The perfect classification metrics and the ROC Curve reaching the upper left corner indicate that the model successfully identified all approved and rejected applications without error. This level of accuracy suggests that the dataset used in the study is highly structured, with well-defined boundaries separating eligible and ineligible applicants. The dominance of the credit score and points variables reflects the strong influence of creditworthiness and internal evaluation systems in financial decision-making. These findings are consistent with previous empirical studies in credit risk modeling, which have shown that credit history, repayment behavior, and institutional scoring mechanisms are among the most reliable indicators of loan performance [2], [3], [6], [7], [10], [17].

The analysis of the Logistic Regression coefficients and odds ratios further clarifies the relationships between applicant characteristics and loan approval probability. Higher credit scores and evaluation points substantially increase the likelihood of approval, indicating that applicants who maintain strong financial discipline and institutional reliability are more likely to be trusted by lenders. In contrast, the negative association observed between loan amount and approval probability suggests that larger loan requests are perceived as riskier and thus less likely to be accepted. The variables income and years employed also contribute positively, although their influence is less significant compared to credit-based measures. These results align with established credit scoring literature emphasizing the predictive importance of borrower characteristics and financial ratios in risk assessment [3], [6], [7], [17], [18], [19].

Despite the excellent predictive results, the findings should be interpreted with caution. The perfect accuracy observed in this model is highly unusual in real-world financial applications, where data typically contain inconsistencies, incomplete records, and overlapping applicant profiles. This may indicate that the dataset used in the study is synthetic or highly curated, leading to a perfectly separable decision boundary. Consequently, the model may exhibit overfitting when applied to new or more diverse data. Prior research has warned about the illusion of classifier superiority and the risks of overfitting in predictive modeling [8], [9], [13]. To address this limitation, future research should consider testing the model using larger and more complex datasets, incorporating cross-validation techniques to assess model generalization, and comparing performance with alternative machine learning algorithms such as Random Forest or XGBoost, which have demonstrated strong performance in credit risk prediction tasks [8], [9], [11], [12].

In practical terms, the findings of this study provide valuable insights for financial institutions seeking to optimize their loan approval processes. By emphasizing measurable indicators such as credit score and evaluation points, institutions can enhance the objectivity and consistency of credit assessments. However, it remains crucial to balance predictive efficiency with fairness and ethical considerations. Over-reliance on automated scoring systems without periodic validation may unintentionally exclude certain applicant groups or introduce bias into lending decisions. Therefore, implementing explainable artificial intelligence techniques and fairness auditing tools could ensure transparency and accountability in credit evaluation systems [14], [15], [16].

Conclusion

This study aimed to evaluate the factors influencing loan approval decisions using a Logistic Regression model trained on financial and employment data. The analysis incorporated key predictors, including income, credit score, loan amount, years employed, and points, to identify variables that most significantly affect the probability of loan approval. The results demonstrated that the Logistic Regression model achieved perfect classification performance, with all evaluation metrics—Accuracy, Precision, Recall, F1-score, and ROC-AUC—recording values of 1.000. This outcome indicates that the model successfully distinguished between approved and rejected loan applications with complete accuracy, confirming that the dataset possesses clear separability between the two classes.

The examination of model coefficients and odds ratios revealed that credit_score and points were the most influential predictors, strongly increasing the likelihood of loan approval. Applicants with higher institutional scores and stronger credit histories were substantially more likely to receive approval, while larger loan requests were associated with a lower probability of approval. The findings emphasize that creditworthiness and institutional evaluation remain the most decisive factors in financial decision-making, while income and employment history serve as complementary indicators of financial stability.

Although the model exhibited perfect performance, such results are rarely achievable in real-world financial environments. The exceptional accuracy may suggest that the dataset is highly structured or synthetic, which could limit the model's ability to generalize to broader or more diverse populations. Therefore, future research should focus on validating the model using larger and more heterogeneous datasets, as well as exploring alternative algorithms such as Random Forest, Gradient Boosting, or Neural Networks to ensure robustness and reliability. Additionally, future studies are encouraged to incorporate fairness and explainability assessments to ensure that predictive accuracy does not compromise ethical and equitable lending practices.

In conclusion, this study provides strong evidence that Logistic Regression is a powerful and interpretable tool for analyzing credit approval patterns. The insights derived from this research can assist financial institutions in enhancing their decision-making processes, improving risk assessment frameworks, and promoting transparency in automated credit evaluation systems.

Declarations

Author Contributions

E.F.M. and M.S.M.; Methodology: E.F.M. and M.S.M.; Software: E.F.M. and M.S.M.; Validation: E.F.M. and M.S.M.; Formal Analysis: E.F.M. and M.S.M.; Investigation: E.F.M. and M.S.M.; Resources: E.F.M. and M.S.M.; Data Curation: E.F.M. and M.S.M.; Writing Original Draft Preparation: E.F.M. and M.S.M.; Writing Review and Editing: E.F.M. and M.S.M.; Visualization: E.F.M. and M.S.M. All authors have read and agreed to the published version of the manuscript

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. W. Goodell and S. Goutte, "Co-movement of COVID-19 and Bitcoin: Evidence from wavelet coherence analysis," Finance Research Letters, vol. 38, pp. 101–625, 2021.

[2] D. J. Hand and W. E. Henley, "Statistical classification methods in consumer credit scoring: A review," Journal of the Royal Statistical Society: Series A (Statistics in Society), vol. 160, no. 3, pp. 523–541, 1997.

[3] L. C. Thomas, D. B. Edelman, and J. N. Crook, Credit Scoring and Its Applications, Philadelphia: Society for Industrial and Applied Mathematics (SIAM), 2002.

[4] L. C. Thomas, D. B. Edelman, and J. N. Crook, Credit Scoring and Its Applications, Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2002. doi: 10.1137/1.9780898718317.

[5] D. J. Hand and W. E. Henley, "Statistical classification methods in consumer credit scoring: A review," Journal of the Royal Statistical Society: Series A (Statistics in Society), vol. 160, no. 3, pp. 523–541, 1997. doi: 10.1111/j.1467-985X.1997.00078.x.

[6] J. N. Crook, D. B. Edelman, and L. C. Thomas, "Recent developments in consumer credit risk assessment," European Journal of Operational Research, vol. 183, no. 3, pp. 1447–1465, 2007. doi: 10.1016/j.ejor.2006.09.100.

[7] R. Anderson, The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation, Oxford, UK: Oxford University Press, 2007. doi: 10.1093/acprof:oso/9780199226405.001.0001.

[8] Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," Expert Systems with Applications, vol. 39, no. 3, pp. 3446–3453, 2012. doi: 10.1016/j.eswa.2011.09.033.

[9] S. Lessmann, B. Baesens, H. V. Seow, and L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research," European Journal of Operational Research, vol. 247, no. 1, pp. 124–136, 2015. doi: 10.1016/j.ejor.2015.05.030.

[10] H. A. Abdou and J. Pointon, "Credit scoring, statistical techniques and evaluation criteria: A review of the literature," Intelligent Systems in Accounting, Finance and Management, vol. 18, no. 2–3, pp. 59–88, 2011. doi: 10.1002/isaf.325.

[11] T. Harris, “Credit scoring using the clustered support vector machine,” Expert Systems with Applications, vol. 42, no. 2, pp. 741–750, 2015. doi: 10.1016/j.eswa.2014.08.017.

[12] K. Bastani, H. Asgari, and A. Namavari, “Wide and deep learning for peer-to-peer loan default prediction and explanation,” Decision Support Systems, vol. 127, pp. 113–143, 2019. doi: 10.1016/j.dss.2019.113143.

[13] D. J. Hand, “Classifier technology and the illusion of progress,” Statistical Science, vol. 34, no. 1, pp. 1–14, 2020. doi: 10.1214/18-STS692.

[14] D. Martens, B. Baesens, and T. Van Gestel, “Decompositional rule extraction from support vector machines by active learning,” IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 2, pp. 178–191, 2011. doi: 10.1109/TKDE.2008.232.

[15] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), San Francisco, USA, 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.

[16] Y. Zhang, M. Hardt, and C. Recht, “Fairness beyond disparate treatment and disparate impact: Learning classification without discrimination,” Communications of the ACM, vol. 64, no. 5, pp. 39–45, 2020. doi: 10.1145/3442291.

[17] E. I. Altman, “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy: Fifty years later,” Journal of Credit Risk, vol. 14, no. 4, pp. 1–34, 2018. doi: 10.21314/JCR.2018.243.

[18] B. C. Yap, S. H. Ong, and A. R. Husain, “Using data mining techniques to predict credit scoring of small and medium enterprises (SMEs) in Malaysia,” International Journal of Business and Information, vol. 13, no. 1, pp. 65–95, 2018.

[19] C. Serrano-Cinca and B. Gutiérrez-Nieto, “The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending,” Decision Support Systems, vol. 122, pp. 113–151, 2019. doi: 10.1016/j.dss.2019.113151

[20] I. Budimir and I. Zjakić, “Two-dimensional logistic regression model of the white effect in printed media,” Expert Systems with Applications, vol. 293, no. Dec., pp. 1–11, Dec. 2025. doi:10.1016/j.eswa.2025.128784

[21] J. Ben-Othman and T. Hariguna, “Sentiment and Concern Classification on Metaverse Governance Responses Using Naïve Bayes and Support Vector Machine (SVM),” *International Journal Research on Metaverse*, vol. 3, no. 1, pp. 14–28, 2026, doi: 10.47738/ijrm.v3i1.43.

[22] U. Rahardja and Q. Aini, “Enhancing Blockchain Security Through Smart Contract Vulnerability Classification Using BiLSTM and Attention Mechanism,” *Journal of Current Research in Blockchain*, vol. 3, no. 1, pp. 28–45, 2026, doi: 10.47738/jcrb.v3i1.56.

[23] R. A. M. Aljohani and A. A. Alnahdi, “Exploring Football Player Salary Prediction Using Random Forest: Leveraging Player Demographics and Team Associations,” *International Journal for Applied Information Management*, vol. 5, no. 4, pp. 203–213, 2025, doi: 10.47738/ijaim.v5i4.115.

[24] G. Toer and G. Kim, “An Empirical Study on the Impact of Feature Scaling and Encoding Strategies on Machine Learning Regression Pipelines,” *International*

Journal of Informatics and Information Systems, vol. 9, no. 1, pp. 242–256, 2026, doi: 10.47738/ijiis.v9i1.293.

[25] P. Vy, D. Lan, and D. Oanh, “The Effect of Globalization on Income Inequality in Developing Countries: A Bayesian Approach,” *Journal of Applied Data Sciences*, vol. 7, no. 1, pp. 1–15, 2025, doi: 10.47738/jads.v7i1.948.

[26] W. Wella, S. Suryasari, and R. Desanti, “Hybrid Deep Learning for Image Authenticity: Distinguishing Between Real and AI-Generated Images,” *Journal of Applied Data Sciences*, vol. 7, no. 1, pp. 37–47, 2025, doi: 10.47738/jads.v7i1.991.

[27] D. Fortuna and C. Hutagalung, “Uncovering Lifestyle and Mental Well-being Predictors of Academic Performance Change in Online Learning: A Comparative Analysis of Interpretable Machine Learning Models,” *Artificial Intelligence in Learning*, vol. 1, no. 4, pp. 271–286, 2025, doi: 10.63913/ail.v1i4.41.

[28] S. Chantanasut, “BERT-Based Emotion and Sarcasm-Aware Classification of Harmful Online Content for Cyber Law Enforcement,” *Journal of Cyber Law*, vol. 1, no. 4, pp. 300–313, 2025, doi: 10.63913/jcl.v1i4.73.

[29] F. Rahma and S. Ulfah, “Emotion-Aware Detection of Cyberbullying in Visual Social Media Content,” *Journal of Digital Society*, vol. 1, no. 4, pp. 299–313, 2025, doi: 10.63913/jds.v1i4.44.