Indonesian Journal of Electrical Engineering and Computer Science
Vol. 43, No. 2, August 2026, pp. 507~521
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v43.i2.pp507-521

Adaptive feature selection for credit scoring models

Mostafa Mohamed SeifElnasr, Yasser Omar, Saleh Mesbah
Department of Information Systems, College of Computing and Information Technology,
Arab Academy for Science, Technology and Maritime Transport (AASTMT), Cairo, Egypt

Article Info

Article history:

Received Jan 31, 2026
Revised Jun 12, 2026
Accepted Jun 27, 2026

Keywords:

Credit scoring
Delinquency prediction
Feature selection
Financial risk modelling
Q-learning
Reinforcement learning
Wrapper methods

ABSTRACT

Accurate credit scoring and delinquency prediction are critical for financial institutions, particularly when evaluating both banked and unbanked clients. Traditional feature selection approaches such as wrapper, filter, and embedded methods often require repeated retraining and may not efficiently explore optimal feature subsets in multi-class credit risk settings. This study proposes a reinforcement learning (RL)–based adaptive feature selection framework using tabular Q-learning to dynamically identify informative feature subsets for credit bucket classification and delinquency prediction. The framework was evaluated on two datasets, including proprietary credit datasets (24,533 records with 18–19 features). Using stratified 10-fold cross-validation, the proposed approach achieved 62.84% accuracy (F1=0.63) for 8-class credit bucket prediction, outperforming traditional wrapper-based methods by up to 4.2% while reducing computational cost compared to exhaustive subset evaluation. For delinquency prediction, the model achieved approximately 70% F1-score, demonstrating improved minority-class sensitivity. Compared to filter and embedded methods, the RL-based framework produced more compact feature subsets while maintaining competitive computational efficiency. These findings demonstrate that adaptive RL driven feature selection provides a practical and scalable mechanism for enhancing predictive performance in credit risk modelling, supporting automated and consistent decision making in financial institutions.

Corresponding Author:

Mostafa Mohamed SeifElnasr
Department of Information Systems, College of Computing and Information Technology
Arab Academy for Science, Technology and Maritime Transport (AASTMT)
Cairo, Egypt
Email: mmseif87@yahoo.com

1. INTRODUCTION

Credit scoring refers to the evaluation of an individual client's creditworthiness within a financial institution. The rapid development of financial services and the emergence of fintech have made companies that leverage information technology to provide financial solutions, and many fintech startups are now offering diverse services, such as consumer finance, microfinance, and treasury and trade finance applications. Before providing these financial services, companies must assess the risk of their clients as mandated by central bank regulations to ensure borrowers are capable of repaying their liabilities. The advent of big data has significantly impacted the financial sector, especially in the areas of credit scoring and risk assessment. Credit scores, which result from comprehensive risk evaluations, play a crucial role in differentiating clients and managing risk. Traditionally, banks have relied on credit scoring systems that use statistical techniques to analyse client data, which includes historical credit information. However, such conventional approaches typically require substantial manual intervention and incur significant processing

time. Data science has revolutionized this space by introducing advanced predictive tools and techniques. These modern approaches automate the risk assessment process by evaluating client data to generate credit scores using both historical credit records and demographic attributes. However, a major challenge remains millions of individuals in low-income economies, referred to as "unbanked" clients, lack sufficient credit history to generate a credit score or access financial services. To address this issue, data science models must incorporate machine learning techniques that can predict credit scores for both "banked" and "unbanked" clients. A crucial component of this process is feature selection, which forms the foundation of these models. Effective feature selection techniques are essential for managing credit scoring datasets that contain a large number of variables. Traditional approaches to credit scoring: these techniques involve extensive manual work and are time-consuming. Data science addresses these challenges by introducing predictive tools and techniques that analyse client data to perform risk assessments and provide clients with adjusted credit scores based on their credit history and other demographic information.

The need for machine learning (ML) in credit scoring: with the increasing number of individuals without credit histories, traditional credit scoring models are inadequate in evaluating the risk for unbanked or non-banking clients. These clients often lack the formal financial background necessary for conventional assessments. Machine learning techniques address this challenge by identifying complex relationships within large and heterogeneous datasets to produce data-driven predictions. These models can incorporate alternative information sources, including transaction records, behavioural indicators, and other non-traditional attributes, to evaluate client creditworthiness. The flexibility and adaptability of ML models are particularly useful in expanding credit access to a wider population, including those who do not have formal credit histories [1].

The role of feature selection in machine learning: feature selection enhances model generalization by removing irrelevant, noisy, or redundant attributes while retaining each feature's original information unlike feature extraction. The feature space is reduced according to a defined optimization objective. Most research in this area focuses on credit scoring using machine learning or deep learning algorithms [2]. Feature selection methods fall into two main categories: i) Statistical methods, which measure the correlation between independent variables and class labels, typically applying least-squares estimation alongside subset selection, shrinkage, and dimensionality reduction to improve predictive performance [3]. ii) Wrapper methods, which evaluate feature subsets by iteratively adding or removing predictors using a machine learning algorithm to score performance, with common techniques including forward selection, backward elimination, and stepwise selection [4] a key drawback of traditional statistical techniques is their inability to capture dependencies among features. Wrapper-based methods are more expressive but incur high computational costs due to repeated model training across feature subsets and are prone to overfitting, as selected features may not generalize beyond the training data.

Reinforcement learning (RL) for feature selection: RL introduces a more dynamic and adaptive approach to feature selection compared to traditional methods. In RL, an agent improves its behavior through continuous interaction with the environment, using reward-based feedback to refine its actions toward maximizing long-term cumulative returns. When applied to feature selection, an RL agent can intelligently navigate through the feature space, selecting the most informative features based on their contribution to the model's predictive accuracy. Unlike wrapper methods, which rely on predefined rules and exhaustive searches, RL agents can autonomously adapt to different datasets and continuously improve feature selection. This flexibility makes RL particularly powerful when dealing with large-scale datasets or non-banking clients with limited historical data [5].

This paper will focus on how to optimize the credit scoring results by using RL to optimize the feature selection methods to enhance the result of credit score prediction; also, we will go to use the credit scoring prediction result to predict the delinquency behaviour for this client after using this credit limit. The contribution will help financial organizations to make a good risk assessment and add value to feature selection methods.

Statement of problem, the rapid growth of FinTech and the increasing availability of large-scale financial data have significantly elevated the importance of credit scoring in banking and consumer financing. Traditional credit scoring approaches, which rely heavily on historical credit records and classical statistical analysis, are increasingly inadequate for modern risk assessment. Consequently, data-driven and machine learning based models have become essential for building robust decision support systems that can accurately assess creditworthiness and manage financial risk. Current limitations in credit scoring methods can be broadly categorized into business related and technical challenges. From a business perspective, one major issue is the treatment of unbanked clients. Most existing credit scoring models are trained exclusively on historical credit data from banked customers, rendering them unsuitable for clients without prior credit histories. As a result, credit limit determination for unbanked individuals often depends on manual, time-consuming risk assessment processes. Incorporating alternative data sources such as demographic and

socio-economic attributes offer a practical solution for automating credit limit assignment and improving decision consistency. Another business-related challenge concerns delinquency behaviour. While many studies focus on predicting credit scores or eligibility, they often neglect the likelihood of delinquency after credit is granted. Delinquency prediction is critical, as it directly influences risk exposure and may alter the assigned credit bucket or lending decision. From a technical perspective, feature selection remains a key challenge. Conventional feature selection techniques including filter, wrapper, and embedded methods are typically static and unable to explore the full combinatorial feature space efficiently. Deep learning-based selection approaches, while powerful, require large datasets, incur high computational cost, and lack adaptability to evolving data distributions or newly introduced features. Moreover, these methods generally fail to incorporate online feedback mechanisms, limiting their effectiveness in dynamic credit scoring environments [6].

Background, Credit loans and consumer financing constitute a core component of the banking industry, and the effectiveness of credit risk assessment directly impacts a bank's profitability and financial stability. Accurate evaluation of a customer's financial background and creditworthiness is therefore essential prior to any lending decision and plays a crucial role in mitigating credit risk [7]. Ineffective credit scoring practices contribute to an increase in non-performing loans (NPLs) and default rates, posing significant challenges to banking institutions. As illustrated in Figure 1, quarterly NPL statistics from 2022 to 2024 reflect persistent levels of credit default. Although credit decisions are generally based on available financial information, external factors such as economic volatility, interest rate fluctuations, and regulatory or tax changes can adversely affect repayment behaviour. Consequently, credit risk remains the primary cause of bank failure and represents one of the most critical risks faced by banking management.

On the other hand, the need for microfinancing and small loans increased these days with high inflations and bad economy status and most people going to the concept of buy now pay later "BNPL" and the number of consumer financing Fintech's increased to fulfil these demands from the clients because all clients going to instalments plans instead of cash payment. According to statistics published by the Finance and Leasing Association (FLA) [8], consumer finance activity increased by 14% in August 2022 relative to August 2021, with cumulative growth reaching 21% over the first eight months of 2022 compared to the corresponding period in the previous year (Figure 2).

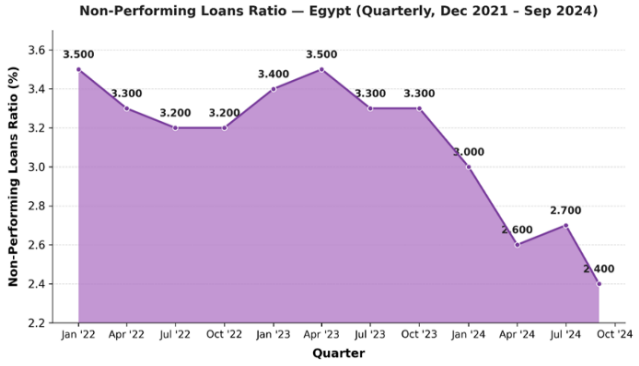

Figure 1. NPL ratio quarterly: Egypt

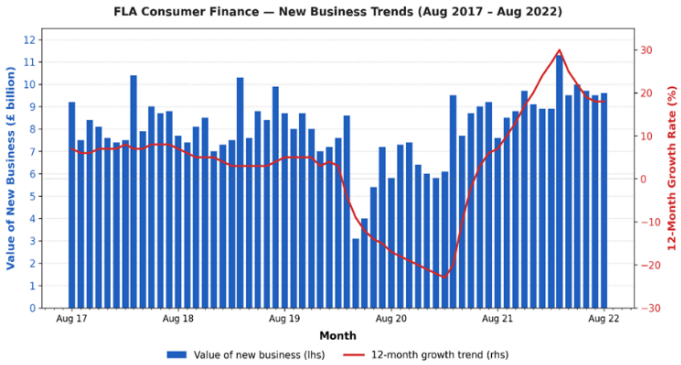

Figure 2. Consumer financing growth analysis

Banks and Fintech's need a quick way to evaluate the client from credit perspective to give them a peed loan without time consuming. on the same time these financial institutions need to guarantee credit risk of the clients.

Credit scoring techniques have been developed to address credit risk assessment by leveraging data-driven models to estimate appropriate credit scores based on demographic and financial attributes [9]. Early foundational work [10] introduced statistical measurements to distinguish between good and bad loans through analysis of applicant characteristics, forming the basis of modern credit scoring research. Since then, credit scoring methodologies have evolved into three principal categories: expert-based models, statistical models, and artificial intelligence (AI)-based approaches.

Statistical techniques such as linear discriminant analysis (LDA) and logistic regression (LR) have been widely adopted, with LDA employing linear discriminant functions to separate classes and LR estimating default probabilities while identifying behaviour-related variables. More recently, AI-based methodsincluding artificial neural networks (ANN), k-nearest neighbour (KNN), support vector machines (SVM), and genetic algorithms (GA) have demonstrated superior predictive performance in complex credit scoring tasks [11], highlighting the growing role of advanced machine learning in financial risk assessment.

Related work and literature survey, many studies have investigated the application of data mining and machine learning techniques including statistical models and neural networks in business and financial decision-making. Most of this work focuses on predicting credit scores or credit eligibility using historical data from clients with established credit histories. Feature selection has been addressed using traditional machine learning or deep learning approaches to improve predictive accuracy; however, alternative feature selection strategies beyond stepwise or deep learning-based methods remain under explored. Furthermore, the majority of existing research overlooks unbanked clients who lack prior credit records, with only limited studies incorporating alternative data sources. One notable exception leverages mobile and demographic data to enhance credit scoring performance for clients without credit history.

Recent studies have also examined RL for improving credit scoring decisions. Herasymovych *et al.* [12] proposed a RL framework to dynamically optimize acceptance thresholds in consumer credit scoring, addressing challenges such as selection bias, population drift, and misaligned business objectives. Using historical data from an international credit company, the authors combined RL with Monte Carlo simulations and gaussian radial basis functions for value approximation. Their results demonstrated that static, cost-sensitive threshold optimization approaches are insufficient and may lead to significant financial losses, whereas the proposed adaptive system achieved substantially higher profitability in both simulated and real-world settings. The study was limited to clients with existing credit histories and focused solely on threshold optimization, without addressing feature selection for improving the underlying credit scoring model.

A significant study on client evaluation in credit scoring is presented in [13], where the authors investigate optimal combinations of feature selection methods and classification algorithms for predicting credit scores. The study evaluates filter, wrapper, and embedded feature selection technique specifically forward selection and backward elimination together with multiple classifiers, including decision trees (DT), random forest (RF), Naïve Bayes, LR, and KNNs. Experiments conducted on a dataset of 91,759 records with 272 features (70/30 train–test split) showed that correlation-based filter selection achieved the best feature selection performance, while RF yielded the highest predictive accuracy. Although the study provides a comprehensive comparison of feature selection and classification methods, it is limited to clients with existing credit histories, does not explore exhaustive feature combinations, and lacks mechanisms for incorporating real-time feedback or adapting to newly introduced features. Kumar *et al.* [14], the authors propose a hybrid credit scoring framework combining deep learning and K-means clustering. The model follows a three-stage pipeline consisting of data preprocessing, dimensionality reduction via feature selection, and deep neural network–based prediction. Credit scoring is categorized into behavioural scoring and collection scoring, with K-means used to cluster customers based on behavioural patterns. Using the Kaggle Home Credit Default Risk dataset, the proposed approach achieved approximately 87% accuracy. Despite its strong predictive performance and extensive use of evaluation metrics including accuracy, RMSE, probability of default, and AUC the framework remains constrained to banked clients, does not explore all possible feature subsets, and lacks adaptability to real-time data changes.A comparative deep learning study is presented in [15], where deep sequential neural networks (DSNN) and convolutional neural networks (CNN) are evaluated against traditional classifiers such as KNN, classification and regression tree (CART), Naïve Bayes, and SVM. Experiments were conducted using 10-fold cross-validation on three datasets: German Credit Data, Australian Credit Approval, and Kaggle Credit Data. The deep learning models consistently outperformed traditional classifiers, achieving accuracies of up to 93.63% on the Kaggle dataset. While the study demonstrates the effectiveness of deep learning in credit scoring across multiple countries, its reliance on large datasets limits generalization to smaller datasets, and it does not address class imbalance or integrate feature selection with traditional classifiers. An alternative direction for credit scoring in

unbanked populations is introduced in MobiScore [16], which leverages mobile phone data as a substitute for conventional financial records. The study extracts consumption, mobility, and social network features from call detail records and evaluates multiple classifiers, including LR, SVM, and gradient boosted trees, using five-fold cross-validation. Results show that incorporating mobile data improves prediction accuracy to 71.5%, compared with 57.7% when using traditional credit bureau data alone. Although this approach significantly advances credit scoring for unbanked clients, it does not employ feature selection techniques and lacks a clear comparative results table, limiting interpretability and optimization.

Group-based feature selection for credit scoring is explored in [17], where Group Lasso regression is proposed as an alternative to stepwise selection methods. Using a UCI credit dataset with 1,000 records and 21 features, the authors compare Group Lasso with LR models optimized via AIC, BIC, cross-validation error, and backward elimination. The results demonstrate comparable accuracies across methods, with Group Lasso offering improved interpretability through grouped variable selection. However, the analysis is limited by the relatively small dataset, the use of a restricted set of classifiers, and insufficient presentation of experimental results. Rasoul *et al.* [15], the authors further explore RL based feature selection by modelling the problem as a markov decision process (MDP) and employing a temporal difference (TD) learning strategy. The selected feature subsets were evaluated using SVM across three datasets: Australian Credit, Breast Cancer Wisconsin (Prognostic), and Connectionist Bench. The reported performance results demonstrate the effectiveness of RL in adaptively identifying informative feature subsets Nevertheless, the study lacks benchmark comparisons with traditional feature selection methods, provides limited dataset and feature descriptions, and evaluates performance using a single classifier, restricting the generalizability of the findings. Table 1 Summarizes the reviewed studies across their core approach, dataset, target population, feature selection strategy, contributions, and limitations.

Table 1. Literature survey summary

Ref	Core approach	Dataset type	Target population	Feature selection strategy	Key contribution	Identified limitation
Herasymovych *et al.* [12]	RL for dynamic acceptance threshold optimization	Historical consumer credit data	Banked clients	Not feature-focused	Adaptive thresholds; higher profitability	No adaptive selection; limited to banked clients
Ziemba *et al.* [13]	RL for dynamic acceptance threshold optimization	91,759 records, 272 features	Banked clients	Forward, Backward, Correlation filter	Comprehensive FS and classifier comparison	No adaptive selection; no real-time feedback; no unbanked focus
Kumar *et al.* [14]	Hybrid Deep Learning + K-Means clustering	Kaggle Home Credit dataset	Banked clients	Dimensionality reduction + DL	High predictive accuracy (~87%)	No exhaustive FS; no adaptive mechanism; banked only
Rasoul *et al.* [15]	Hybrid Deep Learning + K-Means clustering	German, Australian, Kaggle datasets	Banked clients	No explicit FS integration	Strong deep learning performance (up to 93%)	Requires large datasets; no imbalance handling; no FS integration
Pedro *et al.* [16]	Deep Sequential NN and CNN vs traditional ML	Mobile + financial records	UnBanked clients	No explicit FS integration	Alternative data for unbanked scoring	No feature optimization; limited interpretability
Chen an Xiang [17]	MobiScore (mobile phone data-based credit scoring)	UCI dataset (1,000 records)	Banked clients	Embedded (Group Lasso)	Improved grouped feature interpretability	Small dataset; limited classifiers; no adaptive exploration
Rasoul *et al.* [15]	RL-based feature selection (TD learning)	Multiple structured datasets	General	RL-based	Improved grouped feature interpretability	No benchmark comparison; single classifier; limited dataset details

Research gap, despite the growing adoption of machine learning and deep learning techniques in credit scoring, several limitations remain unresolved. First, most existing studies focus on optimizing predictive models using static feature selection methods, such as filter, wrapper, or embedded approaches, which rely on predefined evaluation criteria and lack adaptability to evolving data distributions. Second, prior RL applications in credit risk modelling have primarily concentrated on threshold optimization or policy decisions, rather than adaptive feature subset discovery. Third, existing RL-based feature selection studies are generally evaluated on limited binary datasets and do not address multi-class credit bucket prediction combined with downstream delinquency modelling. Furthermore, the scalability and stability of RL for structured financial datasets remain insufficiently examined.

Consequently, there is a clear need for a rigorously validated framework that applies RL for adaptive feature selection in multi-class credit scoring while explicitly evaluating its impact on delinquency prediction, computational efficiency, and subset stability under controlled experimental settings.

Contribution of this Study, in light of the limitations identified in prior research, this study makes the following contributions:

- Unified RL-based framework for credit scoring and delinquency prediction: this study is among the first to use RL–based adaptive feature selection in a single framework that covers both credit scoring and delinquency prediction, making it more relevant for practical financial risk assessment.
- Comparison with conventional feature selection methods: the proposed method is compared with wrapper, filter, and embedded feature selection techniques under the same preprocessing and evaluation settings, providing a fair assessment of its predictive performance and computational cost.
- Adaptive feature selection instead of static selection: the proposed method treats feature selection as a sequential decision process, allowing the model to explore feature subsets dynamically rather than depending only on static filter, wrapper, or embedded methods.
- Evaluation on multi-class credit problems: the framework is tested on credit bucket prediction and delinquency prediction tasks, which makes it suitable for multi-class credit risk modelling and more realistic financial datasets.
- Stability and robustness analysis: the consistency of the selected features is examined through repeated runs and similarity analysis, providing additional evidence of robustness and interpretability.
- Fair and reproducible evaluation: the study applies stratified cross-validation, class imbalance handling, and the same preprocessing steps across all baseline methods to ensure a fair and reliable comparison.

The remainder of this paper is structured as follows. In section 2 presents the proposed RL-based feature selection framework. In section 3 describes the Method including dataset description, data preprocessing, evaluation setup, and evaluation matrix while section 4 reports the results and discussion. In section 5 concludes and outlines future research directions.

2. PROPOSED REINFORCEMENT LEARNING FRAMEWORK

RL techniques are generally classified into value-based and policy-based categories. Value-based approaches approximate a value function that reflects the expected long-term reward and infer the optimal policy by choosing actions that maximize this value, with Q-learning serving as a typical example. Policy-based methods, by contrast, focus on directly learning a parameterized policy that optimizes cumulative reward. This work adopts Q-learning, a model-free, value-based RL algorithm that does not require prior knowledge of the environment dynamics. Q-learning derives an optimal policy through iterative refinement of a state–action value function, referred to as the Q-function, which associates each state–action pair with its expected long-term return. The algorithm operates off-policy, updating Q-values based on observed transitions while balancing exploration and exploitation through an exploration strategy. At initialization, Q-values are uniformly assigned zero values, indicating that the agent has no prior information about the environment.

For feature selection, the RL environment is defined such that each state represents a subset of selected features, each action corresponds to adding a new feature to the subset, and the reward is determined by the predictive performance of a machine learning model trained using the selected features. A Q-table is maintained to store Q-values for all state-action pairs. At each step, the agent selects an action, updates the feature subset, evaluates model performance, and updates the corresponding Q-value using the standard Q-learning update rule. In this study, a RF classifier is employed to compute the reward metric, which serves as feedback for the agent. As a value-based RL method, Q-learning follows a three-stage workflow: estimating the value function, deriving the optimal policy by maximizing the estimated values, and executing this policy to maximize cumulative rewards. Two forms of value functions are relevant in this context: the state value function V(s), which quantifies the expected reward of being in a given state, and the state-action value function Q(s,a), which estimates the expected reward of taking a specific action in a given state [18].

In discrete environments, Q-values are typically stored in a Q-table. The Q-table is structured as a two-dimensional matrix in which rows correspond to environment states, columns denote available actions, and each entry Q(s,a) stores the estimated value of executing action (a) in state (s). Over time, as the Q-learning algorithm updates the Q-values using the temporal-difference rule, the Q-table converges toward the optimal action-value function. Optimal action selection is achieved by choosing the action associated with the maximum Q-value for the current state.

2.1. MDP formulation

The feature selection task is formulated as a MDP, defined by the tuple (S, A, T, R, γ):

- State (s): a subset of features selected from the full feature set Ψ.
- Action (a): selecting a feature f ∈ Ψ that is not yet included in the current subset.
- Transition $(s \rightarrow s')$: deterministic update of the subset by adding the selected feature.
- Reward (r): a predictive performance metric (e.g., accuracy, F1, or AUC) obtained by training a classifier on the resulting feature subset.
- Policy (π): the feature selection strategy that maximizes the expected long-term predictive performance.
- Discount factor (γ): controls the weighting of future rewards relative to immediate rewards $(0 \leq \gamma \leq 1)$.

Within this framework, Q-learning is applied to approximate the optimal state–action value function $Q^*(s, a)$, which represents the expected cumulative reward of taking action "a" in state "s" and acting optimally thereafter. The Q-table is initialized to zero for all state–action pairs. At each step, the agent selects an action using an ε-greedy policy, choosing a random action with probability ε to encourage exploration, and the action with the highest current Q-value with probability $(1 - \varepsilon)$.

After taking action "a" in state "s" and observing reward "r" and next state " s′ ", the agent computes the TD target, derived from the Bellman optimality equation, as shown in (1):

$$TD\ Target = r + \gamma\ a' \max Q(s', a') \quad (1)$$

- r: the immediate reward obtained after transitioning to s′.
- γ: the discount factor.
- a'maxQ(s′, a′): the maximum estimated value over all possible actions in the next state s′, retrieved directly from the Q-table.

The TD error measures the gap between this newly computed target and the current Q-value stored in the table, as shown in (2).

$$TD\ Error = TD\ Target - Q(s, a) \quad (2)$$

Finally, the Q-value is updated by moving the current estimate toward the TD target by a fraction α (the learning rate), as shown in (3).

$$Q(s, a) \leftarrow Q(s, a) + a\ TD\ Error \quad (3)$$

- α (learning rate): controls how strongly new information influences the existing Q-value $(0 < \alpha \leq 1)$.

Substituting (1)-(3) yields the standard Q-learning update in (4):

$$Q(s, a) \leftarrow Q(s, a) + a\ [r + \gamma\ a' max\ Q(S', a') - Q(s, a)] \quad (4)$$

This process is repeated across many episodes until the Q-values converge, at which point the learned policy corresponds to selecting the action with the highest Q-value in each state.

2.2. Workflow

The proposed framework applies Q-learning–based RL to adaptively select informative features for credit scoring and delinquency prediction, improving over traditional greedy methods Figure 3. The proposed framework is implemented through the following sequential steps:

- Step 1 – RL-based feature selection: a RL agent explores the feature space where each state represents a subset of features. The agent evaluates each subset using a classification model and receives a reward based on predictive performance. Through iterative exploration, the subset with the highest cumulative reward is selected as the optimal feature set for credit scoring.
- Step 2 – Credit scoring classification: the selected feature subset is used to train a classification model that predicts credit scoring buckets.
- Step 3 – Model validation: multiple supervised learning algorithms are applied to evaluate the selected feature subset and identify the classifier that achieves the best performance.
- Step 4 – Delinquency modelling: the predicted credit scoring results are combined with additional features to construct a dataset for delinquency prediction.
- Step 5 – Delinquency feature selection: the RL agent is applied again to identify the most relevant features for delinquency classification and to build a predictive model for client risk behaviour.
- Step 6 – Model assessment: the proposed approach is compared with traditional feature selection techniques, including wrapper methods (forward and backward selection) and embedded methods (tree-based importance), using standard evaluation metrics.

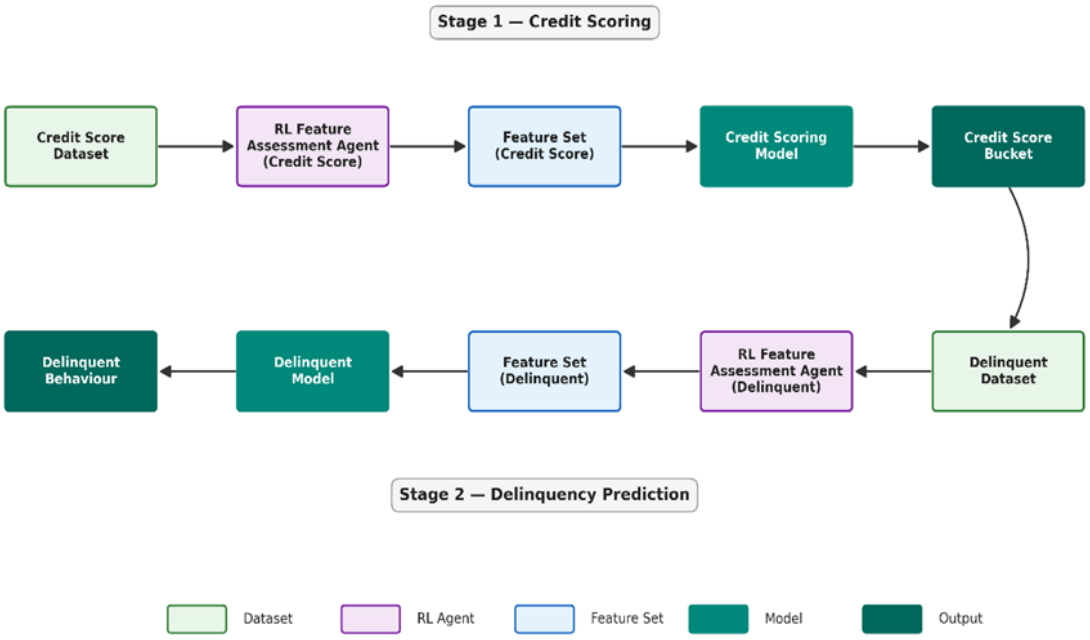

Figure 1. Feature selection agent workflow

Benefits of the proposed approach: the use of RL in feature selection allows for dynamic and adaptive optimization of the feature set, potentially leading to higher accuracy and more robust credit scoring models. Furthermore, this approach addresses the challenge of handling clients with no prior credit history by incorporating alternative data sources and predictive techniques.

3. METHOD

3.1. Datasets description

Dataset 1 – Credit bucket: this dataset Table 2 is employed to predict the credit bucket, which is classified into eight categories representing the client's overall creditworthiness and corresponding credit limit eligibility. The categories range from less than 5K up to above 100K as in Table 3.

Dataset 2 – Delinquency prediction: this dataset Table 2 is employed to predict the delinquency behaviour of clients. The predicted eligible credit bucket serves as an additional input feature, enabling the model to assess the likelihood of delinquency if the bank grants a credit facility. The delinquency behaviour is classified into two categories as in Table 3.

The prediction tasks are based primarily on demographic attributes. An extended formulation incorporates both demographic information and the assigned credit bucket to enhance delinquency prediction. The dataset further includes features capturing demographic, financial, and socio-economic characteristics of both banked and unbanked clients.

Table 2. Dataset 1, 2 features

Dataset	Features
Dataset 1	Gender, Age, Marital status, No. of dependents, Residential status, Time at current address, Profession, Employer type, Industry, Income bracket, Net monthly income, Net monthly expenses, Region, Governate, Actual saved money, Car brand, Club membership, eligible limit bucket
Dataset 2	Gender, Age, Marital status, No. of dependents, Residential status, Time at current address, Profession, Employer type, Industry, Income bracket, Net monthly income, Net monthly expenses, Region, Governate, Actual saved money, Car brand, Club membership, eligible limit bucket, delinquency behaviour

Table 3. Dataset 1, 2 labels

Dataset	Label	Values
Dataset 1	Eligible limit bucket	Less than 5K, 5K–10K, 10K–20K, 20K–30K, 30K–50K, 50K–70K, 70K–100K, Above 100K (8 classes)
Dataset 2	Delinquency behaviour	Good, Bad (2 classes)

3.1.1. Balancing dataset

One of the critical challenges in training machine learning models on financial datasets is the presence of imbalanced class distributions. In the eligible credit bucket dataset, the target variable is composed of 24,533 records distributed across eight categories, as shown in Table 4.

Table 4. Class distribution of credit limit bucket

Credit bucket category	Number of records	Percentage of total
Less than 5K	337	1.37%
5K–10K	1,470	6.00%
10K–20K	7,088	28.89%
20K–30K	7,647	31.17%
30K–50K	4,024	16.40%
50K–70K	1,713	6.98%
70K–100K	1,094	4.46%
Above 100K	1,130	4.61%
Total	24,533	100%

As shown in Figure 4, the distribution is heavily skewed toward the middle ranges (10K–20K and 20K–30K), which together account for over 60% of the dataset. Conversely, categories such as less than 5K and above 100K are severely underrepresented.

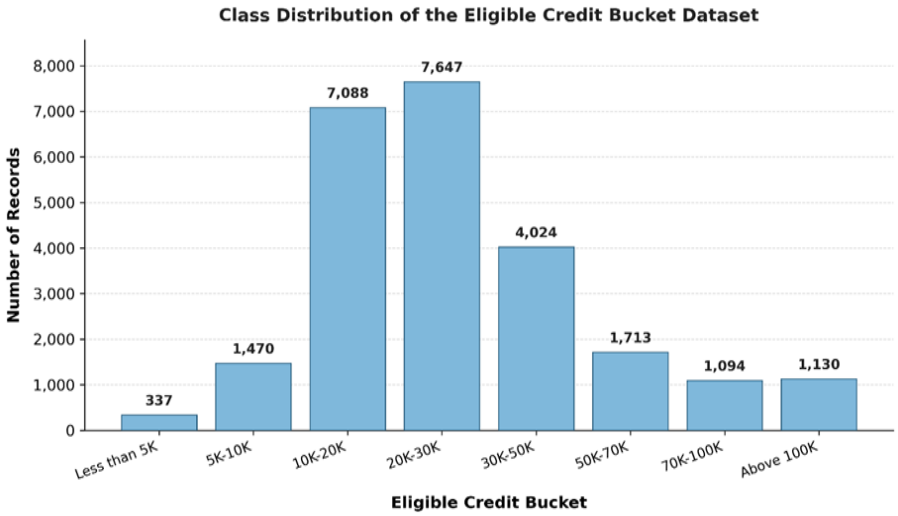

Figure 4. Class distribution of credit limit bucket

Class imbalance may cause learning algorithms to favor majority classes, leading to degraded predictive performance for minority classes. This often results in misleadingly high accuracy values accompanied by reduced Recall and F1-score, as well as poor generalization for underrepresented categories, such as clients belonging to low-frequency credit groups.

Solving imbalance Issue: To address the issue of class imbalance, the synthetic minority oversampling technique (SMOTE) was applied [19]. This method synthesizes new minority class samples through interpolation in the feature space, thereby enhancing data diversity compared to simple duplication strategies. To avoid data leakage, oversampling was restricted to the training data, while the test set retained its original class distribution [20]. After SMOTE, the dataset was more balanced across categories, enabling the model to learn robust decision boundaries. The same balancing methodology applied on the other other dataset [21].

3.2. Data preprocessing

Prior to RL and model construction, multiple feature engineering procedures were applied to improve predictive performance. These transformations ensure that categorical, numerical, and missing data are consistently represented and suitable for Q-learning–based feature selection [22].

Feature encoding: categorical variables, such as gender, profession, employer type, industry, region, governorate, marital status, residential status, car brand, and club membership, are transformed into binary indicator variables using one-hot encoding. This representation avoids ordinal assumptions and enables the RL agent to evaluate the predictive contribution of each category independently [23].

Binning of continuous variables: continuous attributes, including age, time at current address, net monthly income, and net monthly expenses, are discretized into fixed bins. Binning captures potential non-linear effects, improves interpretability, and reduces sensitivity to outliers. Equal-width or quantile-based binning strategies are considered, depending on variable distributions.

Feature scaling: numerical features are standardized or normalized to ensure comparability across different ranges. Feature scaling ensures that variables with larger numerical ranges, such as income, do not dominate the training model and improves the stability of distance-based algorithms where applicable.

Data imputation: missing data is handled systematically through tailored imputation strategies based on the feature type. For categorical variables with a low proportion of missing values, missing entries are imputed using the most frequent category. For numerical variables, mean or median imputation is applied as an initial approach.

Target variable transformation: the dependent variable in this study differs across the two datasets. For dataset 1, the dependent variable is the eligible limit bucket, which represents the client's eligible credit limit. This variable is discretized into ordinal categories, or buckets, reflecting different levels of creditworthiness and credit limit eligibility. For dataset 2, the dependent variable is Delinquency behaviour, which indicates whether the client is likely to demonstrate good or bad repayment behaviour. This variable is represented as a binary classification outcome.

In the case of the dataset 1, each state can be interpreted as a subset of selected features (e.g., Age group, Income range, and Marital status), while each action corresponds to the decision of whether to include or exclude a specific feature in the feature subset during the selection process. The Q-table thus stores the learned Q-values for all possible state–action pairs, reflecting the expected contribution of including a given feature in improving model performance. Table 5 provides a simplified example with three states (S1, S2, S3) representing different client profiles and three actions (A1, A2, A3) representing candidate features (e.g., Age, Income, Marital Status).

- (Income, young, low-income, single client), the Q-learning agent assigns the highest Q-value to Income (0.7), meaning that including Income as a feature contributes most to improving prediction accuracy for this profile.
- For State S2 Income (0.9) is again the most valuable feature, indicating strong predictive power for clients with medium income.
- For State S3 (senior, high-income, married clients), Marital status (0.8) emerges as the most useful feature.

This mapping illustrates how the Q-table helps the RL agent to iteratively identify and prioritize the most informative features for predicting the client's eligible credit bucket. By selecting the feature with the highest Q-value in each state, the agent progressively converges toward an optimal feature subset that maximizes classification performance.

Table 5. Q-learning table sample for dataset 1

State	Action (Age)	Action (Income)	Action (Marital status)
S1(Young-low income, single)	0.5	0.7	0.2
S2(Middle-Age, Medium, Married)	0.1	0.9	0.4
S3(Senior, High income, Married)	0.3	0.6	0.8

3.3. Cross validation

To ensure a fair and robust evaluation, 10-fold cross-validation procedure was employed. The full dataset was divided into ten approximately equal folds while preserving the class distribution in each fold. In each iteration, nine folds were used for training and one-fold was reserved for validation, and the process was repeated until every fold had been used once as validation data. The final performance was reported as the mean value across the ten folds. For imbalanced datasets, SMOTE was applied only to the training portion within each fold to avoid data leakage. The validation fold remained untouched and retained the original class distribution. This protocol ensures that performance estimates are unbiased and more representative of real-world generalization. In addition, the RL procedure was repeated across multiple runs to reduce the effect of randomness in exploration and subset selection. Where applicable, results are reported as mean ± standard deviation.

3.4. Exprimental setup

The Q-learning agent was configured with the following parameters:

- Learning rate (α=0.1): a relatively small value was selected to ensure stable and incremental updates of Q-values, avoiding large oscillations caused by noisy reward signals (classification accuracy from cross-validation).
- Discount factor (γ=0.9): this value emphasizes long-term reward maximization, which is essential in feature selection since the contribution of a feature may only become evident after several selections have been made.
- Exploration probability (ϵ=0.1): the choice reflects a balance between exploration of new feature subsets and exploitation of already promising subsets. A low but non zero exploration rate was used to avoid convergence to suboptimal feature combinations.
- Number of episodes (100): the episode count was constrained to control computational cost, given that each episode requires multiple classifier evaluations. Preliminary experiments showed that 100 episodes were sufficient for convergence in the feature space of 19–20 features

These values were chosen based on prior work in RL and empirical testing on the present datasets, balancing learning stability, convergence speed, and computational cost. The objective is to identify an optimal subset of features that maximizes classification performance across three prediction tasks: eligible limit bucket (Dataset 1), delinquency behaviour (Dataset 2).

3.5. Evaluation matrix

The quality of each selected feature subset was assessed using a RF model, and the resulting validation accuracy was used as the reward signal for the Q-learning agent. For benchmarking purposes, two additional classification algorithms KNN and LR were also included in the evaluation This configuration allowed for a systematic comparison of how different learning algorithms interact with the RL-based feature selection mechanism.

The proposed method and the baseline approaches were evaluated using several standard classification metrics. Accuracy was used to measure the overall proportion of correctly classified instances. Precision, Recall, and F1-score were used to provide a more detailed assessment of predictive performance, especially in the presence of class imbalance. Since the datasets involve multi-class classification, macro-averaged or weighted versions of these metrics were considered where appropriate. In addition, the area under the ROC curve (AUC-ROC) was used to evaluate the ranking ability of the classifiers, particularly in distinguishing between classes under imbalanced conditions. Beyond predictive performance, computational efficiency was assessed in terms of training time and relative memory demand, while feature subset size was considered to evaluate the compactness of the selected representation. To assess the consistency of the selected features across repeated runs, Jaccard similarity was used as a stability measure [24].

To assess the effectiveness of the proposed RL-based feature selection framework, comparative experiments were performed using conventional wrapper, filter, and embedded methods on the two datasets. Performance was evaluated in terms of classification accuracy, computational cost, and the size of the selected features.

4. RESULT AND DISCUSSION

The results indicate that the proposed RL–based feature selection framework provides practical value for credit risk management. For dataset 1, the model achieved moderate accuracy (about 62%), which is still useful for providing an early estimate of client eligibility and supporting credit allocation decisions. For dataset 2, the higher recall and F1-score (about 70% with RF) are more important from a risk perspective, since accurate identification of delinquent clients can help reduce default losses. Overall, the findings show that RL can identify compact and informative feature subsets that improve both model efficiency and decision quality. Comparative experiments with wrapper, filter, and embedded methods further show that wrapper methods remain computationally expensive, filter methods are faster but less effective, and embedded methods provide moderate performance. In contrast, the proposed RL-based approach offers a better balance between predictive performance, subset compactness, and computational cost, making it a suitable alternative for credit scoring and delinquency prediction tasks.

4.1. Evaluation results

Dataset 1 consists of 50,000 records after oversampling (40,000 training and 10,000 testing). The target variable (Credit Bucket) is classified into eight categories: < 5k, 5K–10K, 10K–20K,20K-30K, 30K–50K, 50K–70K, 70K–100K, and >100K. Table 6 summarizes the results for using RF, KNN, and LR to test dataset 1.

Table 6. Performance matrix for dataset 1 using RL

Algorithm	Accuracy	Precision	Recall	F1-score	AUC-ROC	Selected features
RF	62.84%	0.63	0.63	0.63	0.66	Age, Income, Marital status, Gender
KNN	60.25%	0.61	0.60	0.60	0.64	Age, Income, Marital status, Net monthly expense
LR	59.90%	0.60	0.59	0.59	0.62	Gender, Club membership

The dataset 2 contains 50,000 records (40,000 training and 10,000 testing). The target variable (Delinquency behaviour) is classified into two categories: good or bad. The same features for dataset 1 will be used in dataset 2 with additional feature which is the label predicted from dataset 1 (Credit bucket). Table 7 summarizes the results for RF, KNN and LR.

For dataset 1 as Table 8, the results showed that traditional wrapper methods achieved moderate predictive performance. Overall, although embedded and wrapper-based methods achieved competitive performance, they consistently underperformed relative to the proposed RL-based approach, which attained higher accuracy and better generalization with comparable computational efficiency.

The dataset 2 extended the first dataset by incorporating the predicted eligible credit bucket as an additional input feature. Among conventional techniques, backward elimination achieved the best wrapper performance, as mentioned on Table 9.

Table 7. Performance matrix for dataset 2 using RL

Algorithm	Accuracy	Precision	Recall	F1-score	AUC-ROC	Selected feature
RF	70.75%	0.74	0.70	0.72	0.75	Marital status, Net monthly expense, Profession, Governorate, eligible limit bucket.
KNN	63.87%	0.64	0.63	0.63	0.67	Age, Gender, Profession, Club membership, eligible limit bucket.
LR	66.43%	0.66	0.66	0.66	0.69	Age, Profession, Industry, eligible limit bucket.

Table 8. Baseline feature selection results for dataset 1

Algorithm	Accuracy	Precision	Recall	F1-score	AUC-ROC	Selected features
Forward selection	50.90	0.51	0.50	0.50	0.54	Age, Income, Monthly expenses, club membership
Backward elimination	58.64	0.59	0.58	0.58	0.61	Age, Income, Marital status, Club membership, Profession, Dependents, Government
RFE	60.25	0.61	0.60	0.60	0.63	Age, Income, Profession, Gender, Dependents
Mutual information (filter)	55.40%	0.55	0.54	0.54	0.57	Age, Income, Profession, Marital status
Tree-based importance (embedded)	59.75%	0.60	0.60	0.60	0.62	Age, Income, Marital status, Profession

Table 9. Baseline feature selection results for dataset 2

Algorithm	Accuracy	Precision	Recall	F1-score	AUC-ROC	Selected features
Forward selection	61.72	0.62	0.61	0.61	0.66	Age, Gender, Marital status, eligible credit bucket
Backward elimination	66.10	0.66	0.65	0.65	0.70	Age, Gender, Profession, eligible credit bucket, Dependents
RFE	64.85	0.64	0.64	0.64	0.68	Age Income, Profession, Gender, Dependents
Mutual information (filter)	63.05	0.63	0.62	0.62	0.67	Age, Profession, eligible credit bucket
Tree-based importance (embedded)	65.22	0.65	0.65	0.65	0.69	Age, Profession, Marital status, eligible credit bucket

RL–based feature selection incurs additional computational overhead; it provides superior predictive performance compared with conventional approaches. As summarized in Table 10, wrapper methods (Forward, Backward, and RFE) performed the highest computational cost due to repeated model retraining, resulting in long execution times (≈100–120 min for dataset 1). Filter methods are computationally efficient (≈8–10 min) with minimal memory usage but produce lower-quality feature subsets. Embedded methods offer a trade-off between efficiency and performance, achieving moderate accuracy (≈57–59%) with execution times of ≈30–35 min. In contrast, the Q-learning approach attains the highest predictive performance (≈63%

accuracy, F1 ≈ 0.63) with moderate computational cost (≈60 min), making it well suited for applications where accuracy is prioritized and computational resources are available.

Table 10. Computational and cost efficiency

Method category	Time	Memory	Feature subset size	Predictive quality
Wrapper (forward selection)	High (≈ 120 min)	Medium	4-7	Moderate (Acc.≈50-59%)
Wrapper (backward elimination)	High (≈ 110 min)	Medium	6-8	Moderate (Acc. ≈ 58%)
Wrapper (RFE)	High (≈ 100 min)	High	5	Moderate (Acc.≈60%)
Filter (mutual information)	Low (≈ 10 min)	Low	3-5	Low–Moderate (Acc. ≈ 55%)
Embedded (tree-based feature selection)	Medium (≈ 35 min)	Medium	4-5	Moderate (Acc. ≈ 59%)
RL (Q-learning)	Medium High (≈ 60 min)	Medium	4	High (Acc.≈ 63%)

5. FUTURE WORK

While this research has demonstrated the effectiveness of reinforcement learnin specifically using Q-learning in optimizing feature selection for credit scoring and delinquency prediction, there are several topics require further investigation to enhance the agent's performance, scalability, and generalization across diverse financial datasets. A key direction for future work lies in hyperparameter tuning of the Q-learning algorithm paramters. The present study utilized fixed values for critical parameters, including the learning rate alpha (α), discount factor (γ), exploration probability (ϵ), and number of episodes. However, these parameters strongly influence the agent's learning stability and its balance between exploration and exploitation. Future research will focus on identifying optimal hyperparameter configurations that maximize predictive performance and training efficiency [25]. This objective can be pursued through the application of various optimization strategies:

- Grid search and random search: to evaluate a wide range of candidate values and identify combinations yielding the highest classification accuracy and F1-score.
- Bayesian optimization: to model the reward surface of Q-learning as a probabilistic function, allowing adaptive refinement of parameter selection with fewer evaluations.
- Genetic algorithms and evolutionary strategies: hyperparameters are optimized through evolutionary search using mutation and crossover operators.

Beyond hyperparameter optimization, future extensions will explore deep reinforcement learning (DRL) techniques such as deep Q-networks (DQN), double DQN, and dueling DQN. These methods integrate neural networks to approximate Q-values, improving scalability and generalization in high-dimensional feature spaces, especially when applied to complex financial datasets with hundreds of potential features [26]. DRL approaches can further enable nonlinear feature interaction modelling, capturing subtle dependencies that traditional Q-learning may miss.

Addationaly, we can extend extend the proposed framework through federated learning for privacy-preserving credit scoring, real-time deployment in fintech decision systems, and case studies on unbanked populations to further validate practical impact.

6. CONCLUSION

This study presented an adaptive RL-based feature selection framework for credit scoring and delinquency prediction. By formulating feature selection as a MDP and using a Q-learning agent, the proposed method was able to explore feature subsets dynamically and identify compact sets of informative variables without relying on static ranking rules or exhaustive search. The experimental results on the evaluated datasets show that the framework achieved competitive, and in several cases improved, performance compared with conventional filter, wrapper, and embedded feature selection methods. In particular, the selected feature subsets, when combined with standard classifiers such as RF, improved predictive accuracy while maintaining reasonable computational efficiency. The use of predicted credit bucket information in the delinquency modelling stage also strengthened the overall risk assessment process.

From a practical perspective, the framework offers value for financial institutions by supporting earlier and more consistent credit decisions. In credit scoring, it can assist in estimating client eligibility using a reduced but informative feature set, while in delinquency prediction it improves the identification of higher-risk clients, which is important for monitoring and loss prevention. These capabilities can help banks and fintech institutions reduce manual effort, improve decision consistency, and build more scalable risk assessment systems, especially in environments where data are heterogeneous or traditional credit history is limited.

Overall, the findings suggest that RL can serve as an effective and flexible approach for feature selection in credit risk modelling. At the same time, the study acknowledges several limitations, including the scalability constraints of tabular Q-learning, sensitivity to parameter settings, and the need for broader

validation on additional real-world datasets. Future work may therefore extend this framework using more scalable deep RL models and broader deployment scenarios in practical financial systems.

FUNDING INFORMATION

This research was conducted without external financial support and was fully self-funded.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mostafa Mohamed SeifElnasr	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓
Yasser Omar				✓	✓	✓				✓	✓	✓	✓	
Saleh Mesbah			✓	✓	✓					✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The author declares that there is no conflict of interest regarding the publication of this paper.

DATA AVAILABILITY

The datasets used in this study include a combination of proprietary and publicly available data. The proprietary datasets were obtained from a previous employer and have been fully anonymized; therefore, they cannot be publicly shared due to confidentiality and data protection agreements.

REFERENCES

[1] D. Tripathi, D. R. Edla, V. Kuppili, A. Bablani, and R. Dharavath, "Credit scoring model based on weighted voting and cluster based feature selection," *Procedia Computer Science*, vol. 132, pp. 22–31, 2018, doi: 10.1016/j.procs.2018.05.055.

[2] S. Tran and P. Verhoeven, "Kelly criterion for optimal credit allocation," *Journal of Risk and Financial Management*, vol. 14, no. 9, p. 434, Sep. 2021, doi: 10.3390/jrfm14090434.

[3] Y. Chen, W. Ding, Q. Hua, J. Yang, T. Zhou, and F. Fu, "A survey on feature selection techniques for biomedical data: methods and applications," *Neurocomputing*, vol. 670, p. 132534, Mar. 2026, doi: 10.1016/j.neucom.2025.132534.

[4] W. Wang, C. Lesner, A. Ran, M. Rukonic, J. Xue, and E. Shiu, "Using small business banking data for explainable credit risk scoring," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 08, pp. 13396-13401, Apr. 2020, doi: 10.1609/aaai.v34i08.7055.

[5] T. Handhika, A. Sabri, and M. Murni, "Reinforcement learning on the credit risk-based pricing," in *2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST)*, IEEE, Dec. 2021, pp. 233-236. doi: 10.1109/ICCMST54943.2021.00056.

[6] Z. Ma, W. Hou, and D. Zhang, "A credit risk assessment model of borrowers in P2P lending based on BP neural network," *PLoS ONE*, vol. 16, no. 8 August, p. e0255216, Aug. 2021, doi: 10.1371/journal.pone.0255216.

[7] S. Akkoç, "An empirical comparison of conventional techniques, neural networks and the three stage hybrid adaptive neuro fuzzy inference system (ANFIS) model for credit scoring analysis: the case of Turkish credit card data," *European Journal of Operational Research*, vol. 222, no. 1, pp. 168–178, Oct. 2012, doi: 10.1016/j.ejor.2012.04.009.

[8] "Egypt non performing loans ratio," CEIC data provider. Accessed: Apr. 15, 2026. [Online]. Available: https://www.ceicdata.com/datapage/en/indicator/egypt/non-performing-loans-ratio

[9] V. Kuppili, D. Tripathi, and D. R. Edla, "Credit score classification using spiking extreme learning machine," *Computational Intelligence*, vol. 36, no. 2, pp. 402–426, May 2020, doi: 10.1111/coin.12242.

[10] D. Durand, "Risk elements in consumer instalment financing," in *National bureau of economic research*, NBER, 1941. [Online]. Available: https://www.nber.org/books-and-chapters/risk-elements-consumer-instalment-financing

[11] X.-L. Li and Y. Zhong, "An overview of personal credit scoring: techniques and future work," *International Journal of Intelligence Science*, vol. 02, no. 04, pp. 181–189, 2012, doi: 10.4236/ijis.2012.224024.

[12] M. Herasymovych, K. Märka, and O. Lukason, "Using reinforcement learning to optimize the acceptance threshold of a credit scoring model," *Applied Soft Computing Journal*, vol. 84, p. 105697, Nov. 2019, doi: 10.1016/j.asoc.2019.105697.

[13] P. Ziemba, A. Radomska-Zalas, and J. Becker, "Client evaluation decision models in the credit scoring tasks," *Procedia Computer Science*, vol. 176, pp. 3301–3309, 2020, doi: 10.1016/j.procs.2020.09.068.

[14] A. Kumar, D. Shanthi, and P. Bhattacharya, "Credit score prediction system using deep learning and K-means algorithms," *Journal of Physics: Conference Series*, vol. 1998, no. 1, p. 012027, Aug. 2021, doi: 10.1088/1742-6596/1998/1/012027.
[15] S. Rasoul, S. Adewole, and A. Akakpo, "Feature selection using reinforcement learning." Jan. 23, 2021. [Online]. Available: http://arxiv.org/abs/2101.09460
[16] J. S. Pedro, D. Proserpio, and N. Oliver, "Mobiscore: towards universal credit scoring from mobile phone data," in *International conference on user modeling, adaptation, and personalization*, vol. 9146, 2015, pp. 195-207, doi: 10.1007/978-3-319-20267-9_16.
[17] H. Chen and Y. Xiang, "The study of credit scoring model based on group lasso," *Procedia Computer Science*, vol. 122, pp. 677-684, 2017, doi: 10.1016/j.procs.2017.11.423.
[18] M. Ghasemi, A. H. Moosavi, and D. Ebrahimi, "A comprehensive survey of reinforcement learning: from algorithms to practical challenges." Feb. 01, 2025. [Online]. Available: http://arxiv.org/abs/2411.18892
[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
[20] H. Han, W. Y. Wang, and B. H. Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning," in *Lecture Notes in Computer Science*, 2005, pp. 878–887. doi: 10.1007/11538059_91.
[21] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.
[22] S. Zhang, J. Lu, and X. Dong, "Deep reinforcement learning for feature selection: a review and research agenda," *IEEE Access*, vol. 10, pp. 20250–20266, 2022.
[23] U. Agrawal, V. Rohatgi, and R. Katarya, "Normalized mutual information-based equilibrium optimizer with chaotic maps for wrapper-filter feature selection," *Expert Systems with Applications*, vol. 207, p. 118107, Nov. 2022, doi: 10.1016/j.eswa.2022.118107.
[24] M. A. Jahin, M. F. Mridha, N. Dey, and M. J. Hossen, "Human-in-the-loop feature selection using interpretable kolmogorov-arnold network-based double deep Q-network." *IEEE Open Journal of the Computer Society*, Jan. 08, 2026, doi: 10.1109/OJCS.2026.3652986.
[25] J. P. Noriega, L. A. Rivera, and J. A. Herrera, "Machine learning for credit risk prediction: a systematic literature review," *Data*, vol. 8, no. 11, p. 169, Nov. 2023, doi: 10.3390/data8110169.
[26] F. P. Such, V. Madhavan, E. Conti, J. Lehman, K. O. Stanley, and J. Clune, "Deep neuroevolution: genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning." Apr. 20, 2018. [Online]. Available: http://arxiv.org/abs/1712.06567

BIOGRAPHIES OF AUTHORS

Mostafa Mohamed SeifElnasr is the head of enterprise integration at Lunate Capital Limited. He has more than 16 years of experience in information technology and over 12 years of specialized experience in data science and big data technologies. His professional focus includes data integration, data engineering, advanced analytics, and intelligent decision-support systems within the financial services sector. His research interests include machine learning, reinforcement learning, financial data modelling, credit risk analytics, and large-scale data integration architectures, with a particular emphasis on applying data-driven techniques to credit scoring and risk management in banking and asset management environments. He can be contacted at email: mmseif87@yahoo.com.

Yasser Omar received the Ph.D. degree in biomedical engineering from Cairo University, Cairo, Egypt. He has been a Professor of AI with the Department of Computer Science, Faculty of Computing and Information Technology, Arab Academy for Science Technology and Maritime Transport (AASTMT). His current research interests include bioinformatics, medical imaging, data visualization, machine learning, and computing algorithms. He can be contacted at email: dr_yaser_omar@yahoo.com.

Prof. Dr. Saleh Mesbah is an associate professor of Information Systems and the Head of the Remote Sensing and GIS Unit at the College of Engineering, Arab Academy for Science, Technology and Maritime Transport (AASTMT). His academic and research expertise spans remote sensing, geographic information systems (GIS), human–computer interaction, and information systems. He has authored and co-authored numerous peer-reviewed publications in these areas, with research contributions focusing on spatial data analysis, geospatial technologies, and the integration of information systems with advanced computing techniques. His work has been applied across interdisciplinary domains, combining theoretical research with practical engineering and technological applications. He can be contacted at email: saleh.mesbah@aast.edu.