Marketing Science and Technology Journal
2026, 3(2), 13-19

AI-Driven Credit Risk Prediction using PSO-Optimized SVM

Sam Nemati
Islamic Azad University, Science and Research Branch, Tehran, Iran. rouhollah.javahari@srbiau.ac.ir

Alireza Yari
ICT Research Institute, Tehran, Iran. a_yari@itrc.ac.ir

ABSTRACT:
In the rapidly expanding eCommerce ecosystem, credit risk from Buy Now Pay Later (BNPL) and installment payments poses significant challenges to platform sustainability and its cash flow. This study adapts a hybrid PSO-SVM model—optimized via Particle Swarm Optimization (PSO) for hyperplane tuning and feature selection —to predict buyer defaults using transactional data patterns to banking data. Leveraging real-world dataset processing, the model excels at nonlinear pattern detection in high-volume eCommerce transactions, outperforming baseline Support Vector Machine (SVM) in dynamic retail environments. It supports platforms in real-time risk scoring, fraud prevention, and personalized financing, enhancing customer retention in the era of digital transformation. In this research the proposed PSO-SVM offering scalable improvements in comparison to the baseline.

ARTICLE HISTORY
Received: 01/02/2026
Accepted: 01/07/2026

KEY WORDS:
eCommerce Credit Risk,
BNPL Defaults, PSO-SVM Hybrid,
Feature Selection,
AI Analytics,
Digital Payments,
Fraud Prediction

I. Introduction

Credit risk assessment is a cornerstone of modern financial systems and regulatory compliance under Basel II/III frameworks [1]. Traditional credit scoring models, particularly logistic regression–based Internal Ratings-Based (IRB) approaches, have long been favored due to interpretability and regulatory acceptance. However, such linear models assume homoscedasticity and linear separability—assumptions rarely satisfied in real-world financial data [2], [3].

Emerging economies present additional challenges. In countries such as Iran, macroeconomic instability, sanctions, hyperinflation, exchange-rate volatility, and incomplete credit bureau coverage introduce nonlinear interactions and temporal shifts in borrower behavior. Conventional statistical methods typically report predictive accuracies below 85% in such environments [4]. Even ensemble methods may struggle when datasets are imbalanced, noisy, or temporally non-stationary.

Machine learning (ML) methods, particularly Support Vector Machines (SVM), offer strong theoretical guarantees based on structural risk minimization and maximum-margin optimization [5]. Through kernel functions—especially the Radial Basis Function (RBF)—SVMs effectively capture nonlinear relationships inherent in credit datasets [6]. However, SVM performance is highly sensitive to hyperparameters such as the penalty parameter and kernel bandwidth [7].

Metaheuristic optimization algorithms, particularly Particle Swarm Optimization (PSO), provide efficient global search capabilities for hyperparameter tuning [8], [9]. Compared to grid search or Bayesian optimization, PSO offers faster convergence and reduced computational overhead while avoiding local minima traps [10]. Prior research demonstrates PSO-SVM improvements of 5–15% in financial classification tasks [11], [12].

Despite these advances, literature gaps remain:

- Limited studies focus on Iranian banking data spanning multiple economic regimes.
- Few works incorporate business-oriented evaluation metrics such as expected loss.
- Deployment feasibility under hardware constraints is rarely addressed.
- Sensitivity analysis of PSO parameters is often superficial.

This study addresses these gaps through a comprehensive hybrid PSO-SVM framework evaluated on large-scale real-world data.

II. LITERATURE REVIEW

Basel regulatory frameworks institutionalized Probability of Default (PD) estimation through IRB models [1]. While interpretable, logistic regression assumes linearity and independence among predictors, limiting its effectiveness in complex credit datasets [2], [3].

Tree-based ensemble methods such as Random Forest and gradient boosting (XGBoost, LightGBM) improved predictive performance by modeling nonlinear feature interactions [13]. XGBoost, in particular, demonstrated strong performance on benchmark credit datasets [14]. However, these models demand high computational resources and extensive labeled data [4].

SVMs, introduced by Cortes and Vapnik [5], provide a theoretically grounded approach based on margin maximization. Huang et al. [6] demonstrated the superiority of RBF-SVM for credit scoring tasks. However, hyperparameter tuning remains critical for optimal performance [7].

PSO, introduced by Kennedy and Eberhart [8], is inspired by swarm intelligence and has been successfully applied in financial optimization tasks [9]. Danenas [11] and Tang et al. [12] confirmed PSO-SVM superiority in European and Asian credit markets. More recent studies explore hybrid evolutionary models and deep kernel extensions [15], [16]. However, comprehensive deployment-oriented studies in emerging markets remain scarce.

III. METHODOLOGY

A. Dataset

The dataset used in this study comprises 50,000 anonymized retail banking customers collected between 2018 and 2025. This period spans multiple macroeconomic regimes, including intensified international sanctions, currency shocks, and the COVID-19 economic contraction. Such regime shifts introduce non-stationarity in borrower behavior and repayment capacity, which is known to degrade model stability if not properly addressed [4].

Unlike benchmark datasets such as German Credit or Lending Club, which operate in relatively stable economies [4], the Iranian dataset incorporates high inflation volatility, exchange-rate fluctuations, and liquidity constraints. These macroeconomic indicators were explicitly included as explanatory variables to enhance predictive robustness, following recommendations in emerging market credit modeling research [12], [16].

The class distribution (80% non-default, 20% default) reflects realistic conservative lending portfolios. Since imbalanced datasets bias classifiers toward the majority class [17], careful resampling strategies were implemented.

The chronological split ensures forward-looking evaluation as table 1.

TABLE I: THE CHRONOLOGICAL SPLIT

Data range	Chronological spli
2018–2022	Training
2023	Validation
2024–2025	Testing

This temporal design prevents data leakage and aligns with regulatory validation guidelines under Basel frameworks [1].

B. Advanced Preprocessing Pipeline

Credit datasets often contain noise, missing entries, and heterogeneous feature scales. To ensure methodological rigor, preprocessing was embedded within each cross-validation fold to avoid look-ahead bias [4].

1) Missing Value Treatment

Median imputation was applied to numerical variables due to its robustness against outliers common in financial data [3]. Categorical variables were imputed using mode substitution.

For robustness evaluation, KNN-based imputation was tested as an ablation. While KNN marginally improved minority recall (+0.4%), computational cost increased significantly without notable AUC gains.

2) Imbalance Handling

Given the 1:4 minority ratio, three strategies including

SMOTE was superior because it preserves decision boundary geometry, which is critical for margin-based classifiers like SVM [6], [17]. Undersampling reduced model stability due to information loss.

3) Feature Engineering and Selection

Recursive Feature Elimination (RFE) using SVM was applied to identify the most informative features. RFE iteratively removes the least important features based on margin contribution, improving generalization [6].

The final subset preserved 98% of predictive variance while reducing dimensionality from 20 to 15 features, lowering computational complexity.

Macroeconomic features (inflation rate, FX volatility) exhibited significant interaction effects with behavioral variables, confirming nonlinear dependencies consistent with findings in [12] and [16].

C. Support Vector Machine Formulation (Detailed)

SVM seeks a maximum-margin hyperplane that minimizes structural risk rather than empirical risk [5]. This principle provides strong generalization guarantees under limited sample conditions.
Soft-margin SVM solves:
minimize:

$$(1/2)||w||^2 + C\,\Sigma\,\xi_i$$

subject to:

$$y_i\,(w \cdot \varphi(x_i) + b) \geq 1 - \xi_i$$

RBF kernel:

$$K(x_i, x_j) = exp(-\gamma\,||x_i - x_j||^2)$$

maps input vectors into an infinite-dimensional Hilbert space, allowing nonlinear separability [6]. The dual formulation, solved via Sequential Minimal Optimization (SMO), reduces quadratic programming complexity [5].

D. Particle Swarm Optimization (In-depth)

PSO simulates social sharing of information among particles in a swarm [8]. Each particle updates its trajectory based on personal best and global best positions.
Velocity update:

$$v_i^{t+1} = w_t v_i^t + c_1 r_1 (Pbest_i - x_i^t) + c_2 r_2 (gbeSt - X_i^t)$$

The inertia weight w_t balances exploration and exploitation [9].
Unlike grid search, which discretizes the search space inefficiently, PSO operates in continuous space and adapts dynamically [10].
Early stopping was triggered after 10 stagnant iterations to prvent overfitting.

IV. EXPERIMENTAL RESULTS

A. Evaluation

Given the imbalance, AUC-ROC and PR-AUC were prioritized [4].
To ensure robustness, 10-fold stratified CV with 3 temporal splits is applied in the evolution. The multi model comparison, such as Friedman test and pairwise comparison, such as Wilcoxon signed-rank are considered.

B. Performance Interpretation

The performance of different model is show in table 2 in comparison with the proposed model.

TABLE II: PERFORMANCE COMPARISON

Model	Accuracy	F1	AUC
Logistic Regression	82.1%	78.5%	0.87
Grid-SVM	84.2%	80.8%	0.89
Random Forest	91.7%	89.3%	0.94
XGBoost	93.4%	91.4%	0.96
LightGBM	92.8%	90.8%	0.95
CatBoost	93.1%	91.3%	0.96
MLP	91.4%	89.3%	0.93
PSO-SVM	**96.5%**	**94.9%**	**0.98**

Statistical tests confirm significance (Friedman $\chi^2=78.3$, $p<0.001$).

- SMOTE outperforms ADASYN (+1.2% F1).
- Optimal $\gamma \approx 0.045$.
- RBF kernel superior to polynomial and linear.
- 50 particles optimal trade-off between runtime and AUC.

It is believed that a superior performance (AUC = 0.98) arises from margin maximization reducing overfitting [5], nonlinear RBF mapping capturing macroeconomic interactions [6] and PSO avoiding suboptimal hyperparameters [8].

Neural networks underperformed likely due to limited dataset size relative to parameter complexity, consistent with findings in [4].

V. CONCLUSION AND FUTURE WORK

This study demonstrates that PSO-optimized SVM provides a robust and scalable framework for credit risk modeling in emerging markets characterized by macroeconomic volatility and limited credit histories.

The model achieves: 96.5% of accuracy and 0.98 of AUC. Unlike traditional IRB logistic regression models, the proposed framework captures nonlinear borrower behavior and macroeconomic interactions. Compared to ensemble methods, PSO-SVM offers superior interpretability through margin-based decision functions.

The practical implications of the proposed idea can be considered as follows:

- Enables real-time BNPL risk scoring
- Reduces capital reserve requirements via improved PD calibration

- Enhances fraud detection through nonlinear boundary detection

This research had constrained such and Static batch training, single country dataset and Interpretability less transparent than logistic regression. Therefore, for the future work topics such as, Online PSO adaptation for streaming credit data, Federated learning frameworks for inter-bank collaboration, Multi-objective PSO optimizing accuracy and calibration simultaneously, Deep kernel SVMs integrating neural tangent kernels and Explainable AI integration (SHAP/LIME) can be considered.

References:

[1] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273-297, 1995.

[2] J. Kennedy and R. Eberhart, "Particle swarm optimization," in Proc. IEEE Int. Conf. Neural Netw., 1995, pp. 1942-1948.

[3] H. Tang et al., "Risk assessment of credit field based on PSO-SVM," IEEE Conf. Proc., 2020.

[4] P. Danenas, "PSO-based linear SVM classifier selection for credit risk," Expert Syst. Appl., vol. 39, 2012.

[5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. KDD, 2016, pp. 785-794.

[6] N. V. Chawla et al., "SMOTE: Synthetic minority over-sampling," J. Artif. Intell. Res., vol. 16, pp. 321-357, 2002.

[7] Y. Zhang et al., "Financial credit risk prediction model based on PSO-SVM," ACM Trans., 2024.

[8] P. Danenas and I. Palpanas, "Credit risk evaluation using evolutionary SVM," Proc. Comput. Sci., vol. 18, 2012.

[9] M. Sun et al., "Credit risk simulation using ML ensembles," Secur. Commun. Netw., 2022.

[10] C. Yuanqing et al., "PPO-enhanced credit scoring," PLoS ONE, vol. 20, 2025.

[11] Heidari et al., "PSO-XGBoost for Iranian P2P credit," J. Finance Data Sci., vol. 10, 2024.

[12] Basel Committee, "Basel III: Internal ratings-based approach," BIS, 2017.

[13] N. Gasmi et al., "SMPSO feature selection for credit risk," Proc. SPIE, vol. 13067, 2024.

[14] J. Yu et al., "TSC-SVM for enterprise credit assessment," Math. Probl. Eng., 2025.

[15] X. Li et al., "Deep kernel learning for credit risk," IEEE Access, vol. 12, 2024.

[16] R. Zhang, "SMOTE-SVM comparison study," J. Finance Econ., vol. 9, 2023.

[17] L. Wang et al., "Online PSO for dynamic credit scoring," Neural Comput. Appl., 2024.

[18] H. Xiong et al., "Deep reinforcement learning for SCF credit," Expert Syst. Appl., vol. 245, 2024.