

## **AN IMPROVED APPROACH TO CREDIT RISK ASSESSMENT BASED ON THE RANDOM FOREST ALGORITHM IN CREDIT SCORING SYSTEMS**

**Egamberdiyev N.A**

**Doctor of Philosophy (PhD) in Technical Sciences**

**Mahmudova D.A**

### **Abstract**

This article analyzes the effectiveness of classical scoring approaches and modern machine learning models used for assessing customer creditworthiness in banking systems. During the study, the performance of Logistic Regression, Decision Tree (J48), Naive Bayes, k-Nearest Neighbor, and Random Forest algorithms in credit scoring tasks was experimentally evaluated using WEKA and KNIME software environments. The experiments were conducted based on the German Credit Data and Default of Credit Card Clients datasets.

The research results demonstrated that the Random Forest algorithm provides higher stability and classification accuracy compared to other classical models when working with complex and imbalanced credit datasets. In addition, the study investigates the trade-off between interpretability and predictive accuracy in credit scoring systems and discusses the practical application of artificial intelligence models in banking environments.

The proposed approach improves the accuracy of credit risk assessment, reduces the proportion of non-performing loans, and supports the automation of decision-making processes in modern banking systems.

**Keywords:** Credit scoring, credit risk assessment, Random Forest, machine learning, artificial intelligence, WEKA, KNIME, classification, banking systems, ensemble learning.

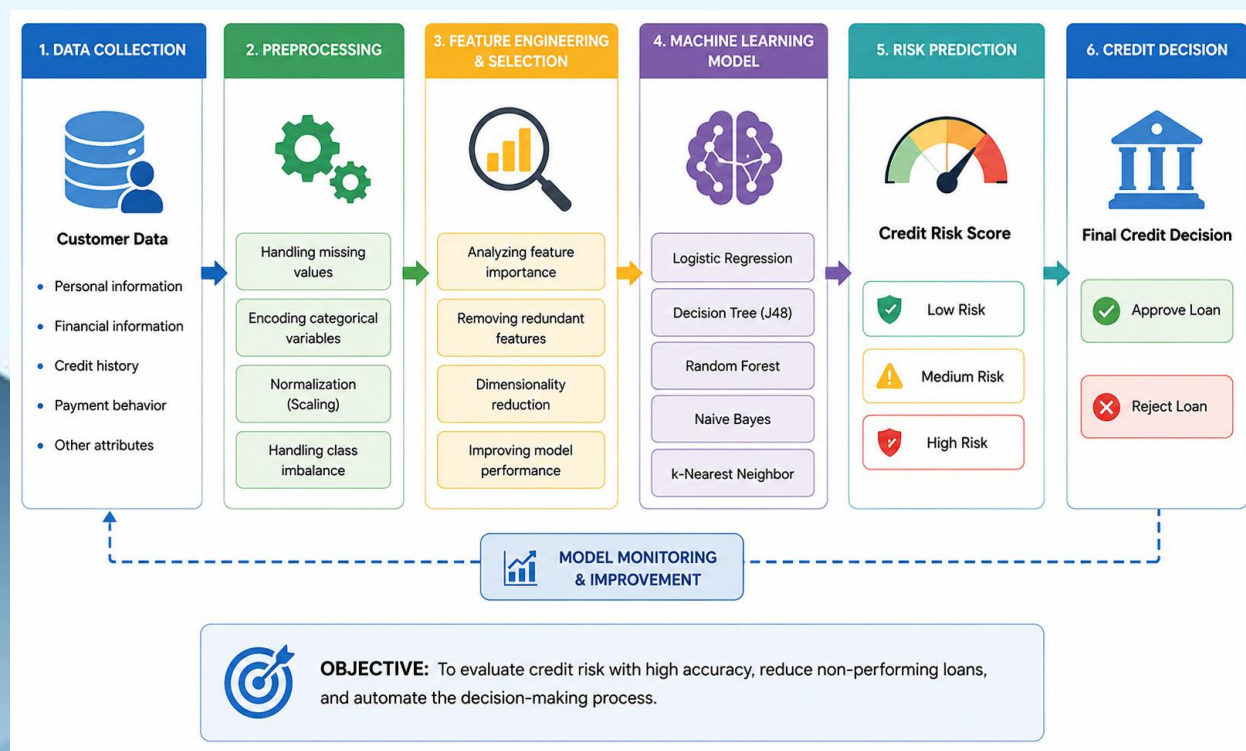
### **Introduction**

Effective credit risk management is one of the key factors ensuring the stability of modern banking systems. The quality of a bank's credit portfolio directly depends on the accuracy of customer creditworthiness assessment, while incorrect risk evaluation may lead to an increase in non-performing loans. The rapid digitalization of banking services and the growing number of credit applications have significantly increased the demand for intelligent and automated credit scoring systems.

Traditional credit scoring approaches are mainly based on statistical models and expert judgment. Although these methods are widely applied in banking practice, they are often limited to linear relationships among variables. However, real-world banking data are typically complex, high-dimensional, and imbalanced. In particular, the relatively small proportion of default cases in credit datasets negatively affects the performance of classical classification models. Therefore, the application of machine learning algorithms in credit risk assessment has become an important scientific and practical research area.

One of the major advantages of artificial intelligence models is their ability to identify hidden patterns within large-scale datasets, analyze complex relationships among attributes, and predict credit risk with higher accuracy. In recent years, algorithms such as Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and Neural Networks have been widely adopted in modern credit scoring systems.

The primary objective of this study is to compare the effectiveness of classical and modern classification algorithms in credit scoring systems and to investigate the advantages of the Random Forest algorithm for credit risk assessment.



**Figure 1.** General Architecture of the Credit Scoring System

Figure 1 illustrates the general workflow architecture of the credit scoring system. Initially, customer-related financial and social information is collected. In the preprocessing stage, the data are transformed and prepared for machine learning models through attribute transformation and data cleaning procedures. Subsequently, the machine learning model predicts the level of credit risk, and the final credit decision is generated based on the prediction results.

### **The Role of Classification Models in Credit Scoring Systems**

From the perspective of machine learning, creditworthiness assessment is considered a classification problem. During the classification process, customers are assigned to predefined classes according to their financial and behavioral attributes. In credit scoring systems, these classes are generally represented as “reliable customers” and “high-risk customers.”

For many years, Logistic Regression has been one of the primary scoring instruments used in banking practice. The main advantage of this model lies in its interpretability, allowing financial institutions to explain the reasons behind credit approval or rejection decisions. However, Logistic Regression has limitations in representing complex and nonlinear relationships among variables.

Logistic Regression is one of the most widely applied classical classification algorithms in credit scoring systems. The model estimates the probability of customer default using a probabilistic approach and mathematically represents the influence of input attributes on prediction outcomes. The model can be expressed as follows:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

where:

- $x_i$  - represents the input attributes of the model;
- $P(y=1|x)$  denotes the probability that a customer belongs to the positive class;
- $\beta_i$  represents the weight coefficients associated with the attributes.

Decision Tree algorithms are also practically convenient for banking systems because they represent decision-making rules in a visual and interpretable form.

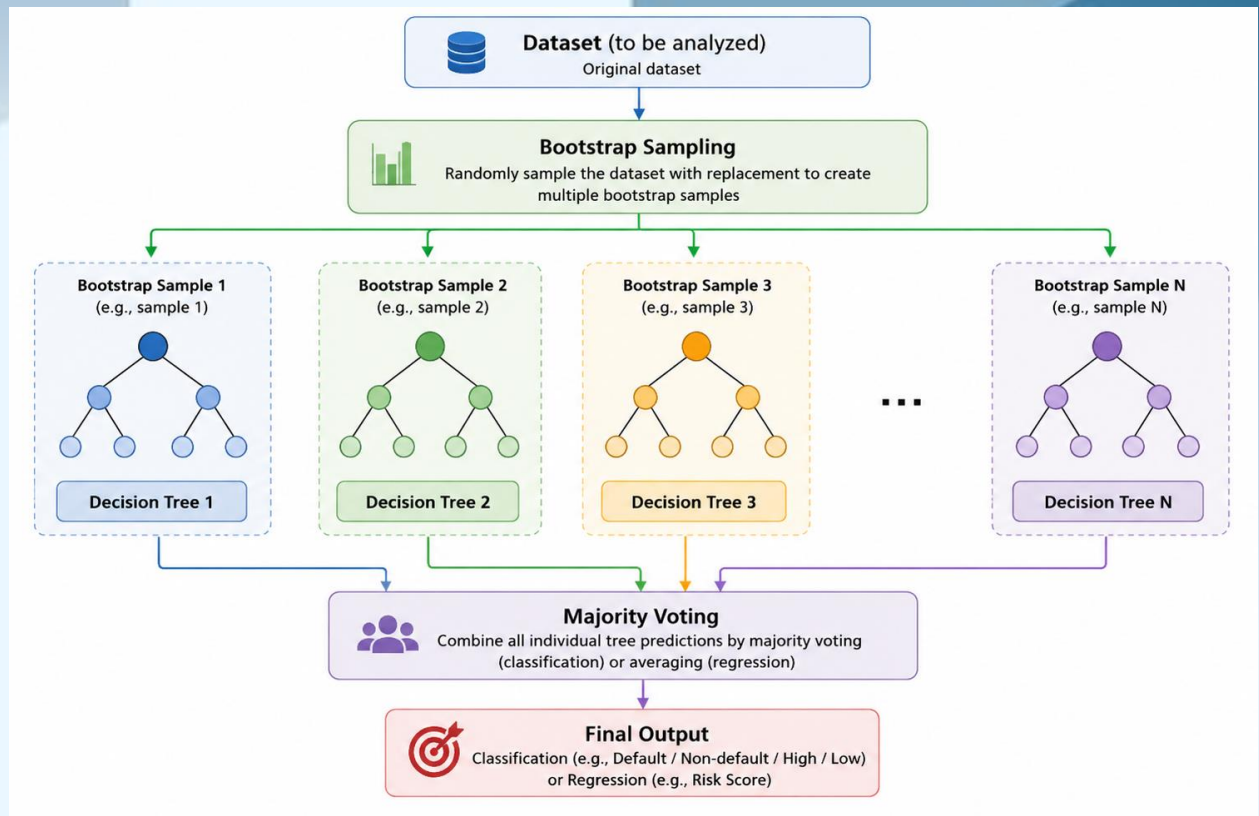
Nevertheless, a single decision tree model may become excessively fitted to training data, resulting in overfitting problems.

Random Forest is an ensemble learning algorithm that generates the final classification result by combining multiple decision trees. The model operates based on bootstrap sampling and majority voting mechanisms. Each decision tree is trained using a randomly selected subset of the dataset, while the final prediction is determined according to the collective voting of all trees.

One of the major advantages of the ensemble approach is its ability to improve model generalization capability. While a single decision tree may become excessively fitted to training data, the combination of multiple trees in the Random Forest algorithm significantly reduces this problem. As a result, the model:

- is less prone to overfitting;
- demonstrates stable performance on noisy data;
- identifies complex relationships among attributes;
- maintains high classification accuracy on imbalanced datasets.

These characteristics are particularly important in credit scoring systems because real-world banking datasets are often complex, nonlinear, and imbalanced in nature.



**Figure 2.** Architecture of the Random Forest Algorithm

Figure 2 illustrates the operational mechanism of the Random Forest algorithm. Initially, multiple random subsets are generated from the original dataset using bootstrap sampling. Separate decision trees are then trained independently on each subset. The final classification result is produced through the majority voting mechanism based on the outputs of all decision trees. This ensemble-based approach improves model stability and enhances generalization capability.

### Research Methodology

In this study, two benchmark datasets widely used in international credit risk assessment research were selected:

- German Credit Data;
- Default of Credit Card Clients Dataset.

The German Credit dataset contains 1000 instances with 20 attributes, while the Credit Card dataset consists of 30,000 observations and 24 attributes related to customer financial behavior and repayment history.

The following classification algorithms were applied in the experiments:

- Logistic Regression;
- Decision Tree (J48);

- Random Forest;
- Naive Bayes;
- k-Nearest Neighbor (kNN).

The experiments were conducted using WEKA and KNIME Analytics Platform environments. In the WEKA environment, a 10-fold cross-validation strategy was applied to ensure reliable model evaluation. In the KNIME environment, the datasets were divided into training and testing subsets using a 70/30 ratio.

During the data preprocessing stage:

- categorical attributes were encoded;
- normalization procedures were performed;
- missing values were checked;
- class imbalance was analyzed.

These preprocessing procedures were essential for improving the generalization capability of machine learning models and reducing the probability of incorrect classification.

### Experimental Results and Discussion

In credit scoring tasks, relying solely on accuracy is insufficient for evaluating model performance. Particularly in the presence of imbalanced credit datasets, accuracy may not fully reflect the actual effectiveness of a classification model. Therefore, modern machine learning research widely employs additional evaluation metrics such as Precision, Recall, and F1-score.

Precision indicates the proportion of objects classified as “positive” that are truly positive. Recall represents the proportion of actual positive instances correctly identified by the model. The F1-score combines these two metrics into a single evaluation indicator and is especially important when analyzing model performance on imbalanced datasets.

**Table 1.**

Confusion Matrix Results of the CRW-RF+ Model

|  | Predicted Reject | Predicted Approve |
|--|------------------|-------------------|
|  |                  |                   |

|                |        |         |
|----------------|--------|---------|
| Actual Reject  | TN = 4 | FP = 0  |
| Actual Approve | FN = 1 | TP = 12 |

The confusion matrix results demonstrate that the proposed model successfully distinguished risky customers with high accuracy. In particular, the false positive value of 0 indicates that no risky customers were incorrectly approved by the model. In credit scoring systems, false positive cases are considered among the most critical classification errors because incorrectly approved risky customers may lead to significant financial losses for banks.

These results confirm that the proposed CRW-RF+ model maintains stable performance even under imbalanced dataset conditions. Furthermore, the model achieved an effective balance between Precision and Recall metrics.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

These evaluation metrics provide a deeper analysis of misclassification risk in credit scoring systems. In particular, maintaining a high Recall value is crucial for identifying problematic customers. Incorrectly classifying risky customers as reliable may result in substantial financial losses for banking institutions.

One of the main characteristics of credit scoring datasets is class imbalance. In real banking environments, the number of customers who successfully repay loans is usually significantly higher than the number of problematic borrowers. This imbalance directly affects the training process of machine learning models and may cause classical algorithms to incorrectly classify minority class instances.

During the research, the Random Forest algorithm demonstrated higher stability compared with other classical models under imbalanced dataset conditions. In particular, the ensemble-based combination of decision trees improved the detection performance of minority class instances. As a result, the model maintained an optimal balance between Precision and Recall metrics.

These findings indicate that relying only on overall accuracy is insufficient in credit scoring systems. In some cases, a model may achieve high accuracy while still incorrectly approving risky customers. Therefore, evaluation metrics such as Precision, Recall, F1-score, and ROC-AUC play a crucial role in credit risk assessment tasks.

**Table 2.**

Comparison of Standard Random Forest and CRW-RF+ Models

| <b>Metric</b>     | <b>Standard RF</b> | <b>CRW-RF+</b> |
|-------------------|--------------------|----------------|
| Accuracy          | 88.24%             | 94.12%         |
| Precision         | 0.9231             | 1.0000         |
| Recall            | 0.9231             | 0.9231         |
| F1-score          | 0.9231             | 0.9600         |
| Specificity       | 0.7500             | 1.0000         |
| Balanced Accuracy | 0.8365             | 0.9615         |
| ROC-AUC           | 0.9615             | 0.9615         |
| False Positive    | 1                  | 0              |

According to the experimental results, the proposed CRW-RF+ model outperformed the classical Random Forest algorithm across all major evaluation metrics. In particular, the Precision value of 1.0000 indicates that all customers approved by the model were truly reliable borrowers.

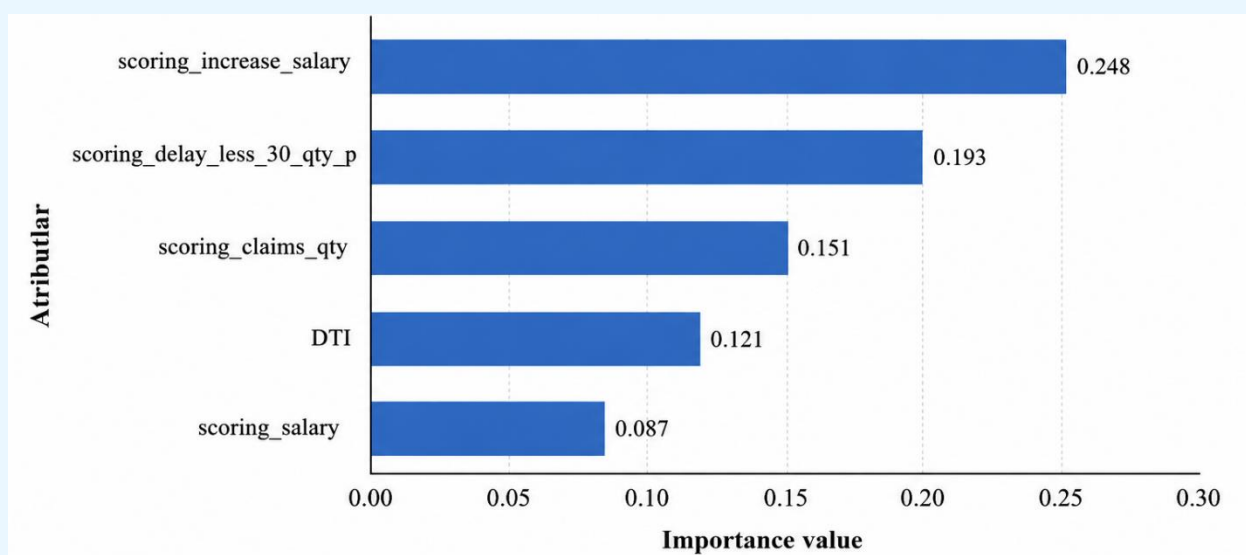
The increase in Balanced Accuracy from 0.8365 to 0.9615 confirms that the proposed model maintained stable performance even under imbalanced dataset conditions. Furthermore, the false positive value of 0 represents a highly important practical result for banking systems, since incorrectly approved risky customers may lead to significant financial losses.

These findings demonstrate that the proposed approach is more effective than the classical Random Forest model for credit risk assessment tasks.

One of the major advantages of the proposed model is its ability to determine the influence level of attributes on credit risk assessment. Through the feature importance mechanism, the Random Forest algorithm identifies the attributes that contribute most significantly to the final prediction outcome. This approach enables deeper analysis of risk indicators in credit scoring systems.

During the study, the most influential attributes identified by the model were analyzed. The results showed that income stability, the number of delays in credit history, debt-to-income ratio (DTI), and the overall financial condition of customers had the greatest impact on credit risk prediction.

These findings are consistent with real-world banking risk indicators. In particular, overdue cases in credit history and income stability were identified as critical factors in assessing customer creditworthiness.



**Figure 3.** Feature Importance Analysis in the CRW-RF+ Model

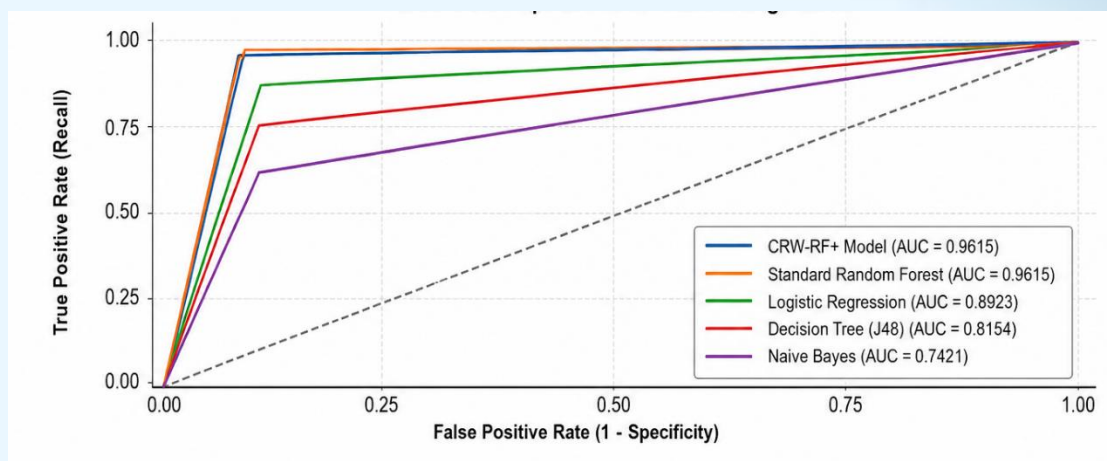
Figure 3 presents the most influential attributes identified by the CRW-RF+ model. According to the results, the attributes scoring\_increase\_salary, scoring\_delay\_less\_30\_qty\_p, scoring\_claims\_qty, DTI, and scoring\_salary had the greatest impact on credit risk prediction.

These results confirm that income stability, debt burden, and credit history-related attributes play a significant role in credit scoring systems.

To perform a deeper evaluation of model performance, the ROC-AUC metric was also analyzed. The ROC (Receiver Operating Characteristic) curve illustrates the

model's behavior under different threshold values and demonstrates the relationship between the true positive rate and false positive rate.

The AUC (Area Under Curve) value evaluates the classification capability of the model. The closer the AUC value is to 1, the higher the classification accuracy of the model. During the experiments, the CRW-RF+ model achieved a high ROC-AUC value, indicating its strong ability to reliably distinguish credit risk classes.



**Figure 4.** ROC-AUC Comparison of Credit Scoring Models

The experimental results showed that, in addition to classification accuracy, interpretability is also an important factor in credit scoring systems. Logistic Regression remains widely used in practical banking applications because of its high interpretability. However, the model cannot fully capture complex and nonlinear relationships among variables.

Although the Random Forest algorithm demonstrated higher accuracy and stability, its interpretability is more complex compared with Logistic Regression. Therefore, maintaining a balance between predictive accuracy and explainability has become an important scientific and practical challenge in modern credit scoring systems.

The proposed CRW-RF+ model not only improved the accuracy of credit risk assessment but also provided partial interpretability of model decisions through the feature importance mechanism.

**Table 3.**

*Comparison of Algorithm Accuracy in the WEKA Environment*

| <b>Algorithm</b>    | <b>German Credit</b> | <b>Credit Card</b> |
|---------------------|----------------------|--------------------|
| Logistic Regression | 75.2%                | 81.0%              |
| J48                 | 73.4%                | 81.96%             |
| Random Forest       | 79.0%                | 81.84%             |
| Naive Bayes         | 76.5%                | 65.25%             |
| kNN                 | 77.0%                | 72.95%             |

According to the experimental results, the Random Forest algorithm achieved the highest classification accuracy on the German Credit dataset. This result can be explained by the ensemble nature of the model, where multiple decision trees compensate for individual classification errors and improve overall predictive stability.

Although the J48 algorithm demonstrated slightly higher performance on the Credit Card dataset, Random Forest showed greater overall robustness. In particular, the model performed more reliably than other classical algorithms under imbalanced dataset conditions. The results obtained in the KNIME environment are presented in Table 4.

**Table 4.**

*Comparison of Algorithm Accuracy in the KNIME Environment*

| <b>Algorithm</b>    | <b>German Credit</b> | <b>Credit Card</b> |
|---------------------|----------------------|--------------------|
| Logistic Regression | 75.3%                | 82.5%              |
| Decision Tree       | 73.4%                | 74.3%              |
| Random Forest       | 79.0%                | 81.9%              |
| Naive Bayes         | 76.5%                | 80.6%              |

|     |       |       |
|-----|-------|-------|
| kNN | 77.0% | 72.9% |
|-----|-------|-------|

The experimental results demonstrated that the Random Forest algorithm produced nearly identical results across different software environments. This indicates the strong generalization capability and stability of the model.

Logistic Regression remains important for practical banking systems because of its interpretability. However, the model showed lower predictive performance compared with ensemble-based approaches due to its limited ability to capture complex attribute relationships.

In some datasets, the Naive Bayes algorithm demonstrated a significant decrease in classification accuracy. This can be explained by the existence of correlations among financial attributes, while Naive Bayes assumes conditional independence between variables.

The experiments also confirmed the presence of class imbalance in credit datasets. Such imbalance increases the risk of incorrect classification in credit scoring systems. Nevertheless, the Random Forest algorithm maintained stable performance under imbalanced conditions, indicating its strong potential for real-world banking applications.

### **Scientific Contributions of the Study**

During this research, the effectiveness of classical and modern machine learning algorithms in credit scoring systems was analyzed, and the CRW-RF+ model was proposed for credit risk assessment.

The main scientific contributions of the study can be summarized as follows:

- the impact of class imbalance on model performance in credit scoring datasets was investigated;
- the effectiveness of the Random Forest algorithm under imbalanced dataset conditions was analyzed;
- the CRW-RF+ model was developed based on a credit risk weighted approach;
- the most influential attributes affecting credit risk prediction were identified using the feature importance mechanism;

- the model was comprehensively evaluated using Precision, Recall, F1-score, and ROC-AUC metrics;
- the CRW-RF+ model demonstrated higher stability compared with the classical Random Forest algorithm.

These findings indicate that the application of artificial intelligence models in credit scoring systems can improve the accuracy of credit risk assessment and enhance the automation of banking decision-making processes.

### **Practical Significance of the Proposed Approach**

The results obtained during the study demonstrated that the proposed CRW-RF+ model has strong practical potential for credit scoring systems. The stable performance of the model under imbalanced dataset conditions significantly improved the accuracy of credit risk identification. In particular, the reduction of false positive cases represents an important practical achievement for banking systems, since incorrectly approved risky customers may lead to substantial financial losses.

The proposed approach supports the automation of credit scoring processes, reduces subjective decisions caused by human factors, and enables the rapid processing of large-scale financial data. In addition, the feature importance mechanism allows the identification of the key attributes influencing model decisions, thereby improving the interpretability of credit risk analysis.

The research results indicate that integrating the proposed model into real banking systems may optimize credit approval processes and reduce the proportion of non-performing loans.

### **Conclusions**

This study comprehensively analyzed the effectiveness of classical and modern machine learning algorithms in credit scoring systems. The performance of Logistic Regression, Decision Tree, Naive Bayes, k-Nearest Neighbor, and Random Forest models in credit risk assessment was experimentally evaluated using WEKA and KNIME software environments.

The research results confirmed that the Random Forest algorithm demonstrates high accuracy and stability when working with complex and imbalanced credit datasets. In particular, the ensemble learning approach improved the generalization

capability of the model and increased the effectiveness of identifying risky customers. The proposed CRW-RF+ model achieved superior results compared with the classical Random Forest algorithm in terms of Precision, F1-score, Specificity, and Balanced Accuracy metrics.

Furthermore, the feature importance mechanism enabled the identification of the attributes that have the greatest impact on credit risk assessment. The ROC-AUC analysis, confusion matrix evaluation, and imbalance dataset investigation demonstrated that the proposed model has strong potential for practical implementation in real banking systems.

The findings of this study confirm that the application of artificial intelligence models in credit scoring systems can reduce banking risks, automate decision-making processes, and improve the overall quality of credit portfolios. Future research directions may include deeper optimization of the proposed model using real banking datasets, as well as the integration of ensemble learning and deep learning approaches.

### References

1. Han J., Kamber M. *Data Mining: Concepts and Techniques*. - Morgan Kaufmann, 2012.
2. Witten I.H., Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. - Elsevier, 2016.
3. Breiman L. *Random Forests* // Machine Learning. - 2001.
4. Bishop C. *Pattern Recognition and Machine Learning*. - Springer, 2006.
5. Thomas L.C. *Consumer Credit Models: Pricing, Profit and Portfolios*. - Oxford University Press, 2009.
6. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning*. - Springer, 2009.
7. Larose D. *Discovering Knowledge in Data*. - Wiley, 2014.
8. German Credit Data Dataset. *UCI Machine Learning Repository*.
9. Default of Credit Card Clients Dataset. *UCI Machine Learning Repository*.
10. Scientific and methodological materials of Tashkent University of Information Technologies named after Muhammad al-Khwarizmi.

- 11.Lessmann S., Baesens B., Seow H.-V., Thomas L.C. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research // European Journal of Operational Research. - 2021.
- 12.Maldonado S., López J. Dealing with imbalanced class problems in credit risk prediction using machine learning techniques // Expert Systems with Applications. - 2022.
- 13.Ribeiro M.T., Singh S., Guestrin C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier // Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. - 2021.
- 14.Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. - 2021.
- 15.Bhattacharyya S., Jha S., Tharakunnel K., Westland J.C. Data mining for credit card fraud: A comparative study // Decision Support Systems. - 2023.