

Cross-Selling Car Loans to Remittance Recipients in Uzbekistan: A Machine Learning Approach Using SMOTE-ENN

Masuda Isaeva

*Department of retail business, Asakabank, 67 – Nukus str., Tashkent, Uzbekistan

masudaisaeva@gmail.com

Abstract—The primary objective of this study is to develop a robust model that assists financial institutions in identify potential customers with a higher likelihood of adopting car loans. We use a unique dataset of remittance transactions and vehicle financing data provided by a commercial bank in Uzbekistan. To balance data, we apply data sampling techniques. We then compare the performance of Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF) models in these two datasets. Our analysis reveals that all models perform better on the Synthetic Minority Over-sampling Technique Edited Nearest Neighbours (SMOTE-ENN) dataset. DT outperforms the other models. Based on these insights, we recommend using DT and SMOTE-ENN techniques on imbalanced datasets. The results of this study offer practical implications for data scientists and financial institutions in remittance-receiving countries aiming to leverage remittance flows, boost cross-selling, and increase revenue.

Keywords— cross-selling, machine-learning (ML), remittances, vehicle loans, Uzbekistan

I. INTRODUCTION

In Uzbekistan, labor migration plays a significant role in economic development due to low wages and limited employment opportunities within the country [1]. The volume of remittances is ten times larger than foreign direct investments to the country (Figure 1).

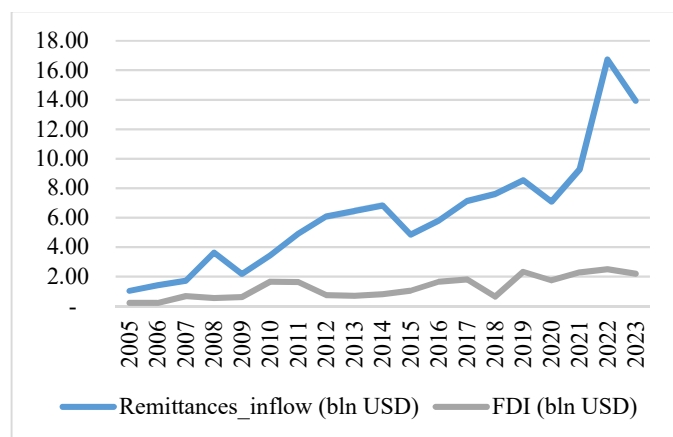


Figure 1. Inflow of remittance (in billion USD) and foreign direct investments to Uzbekistan [3].

The growing volume of remittances is a key driver of poverty reduction in the country. Remittances represent 7% of household income in the lower 40% bracket and up to 12% for those in the upper 60%. Comparatively, the overall share of external remittances from labor migrants comprises 15% of household income [2].

Remittance recipients spend more on wedding ceremonies, gifts, and vehicles than non-recipients do. Automobiles are purchased to signal social status, and display wealth and income rather than fulfilling practical needs [4]. Several factors contribute to such unproductive spending of remittance in Uzbekistan: the lack of remittance-backed products, underdeveloped financial sector development and trust, and low financial literacy among recipients [5].

Financial institutions are interested on redirecting this “unproductive” spending toward “productive” uses, such as investments and savings. The adoption of financial products and services could facilitate capitalizing remittance and generating additional resources for financial institutions.

Considering that remittance recipients are inclined toward vehicle purchases and have steady, albeit unofficial, income sources, this research aims to develop a predictive cross-selling model using ML algorithms to identify which recipients are more likely to adopt car loans. This study addresses two research questions: 1) How do LR, DT, and RF perform in predicting the cross-selling of car loans car loans to remittance recipients in Uzbekistan? 2) How does the application SMOTE-NC and SMOTE-ENN influence the predictive performance of machine learning models?

There are some gaps remain in the existing literature. First, there is limited research on the impact of remittances on car loans, particularly in Uzbekistan. Second, most studies rely on survey data, which often fail to capture real-world dynamics and are prone to biases, limiting the accuracy of findings. Such studies do not offer practical insights for financial institutions.

This study addresses these gaps by analyzing the relationship between remittances and vehicle financing using the latest transactional data. The data on remittance receipts and vehicle loans is provided by a commercial bank in Uzbekistan. This data includes transactions from January 2023 to October 2024. We apply LR, DT, and RF models to examine the nonlinear relationship between independent and dependent variables.

This study has several contributions. First, it enhances the existing literature on remittances and financial services by examining the case of Uzbekistan. Second, we use long-term dataset which consist transactional data of remittance recipients and their adoption of car loans. Third, we develop a high-performance classification model using a limited set of independent variables. Lastly, the study contributes to the methodological framework for handling imbalanced datasets. The findings of this study will be interesting to financial institutions seeking to increase financial inclusion among remittance recipients and policymakers interested in expanding financial inclusion.

The rest of the paper is structured as follows: Section 2 reviews related literature. Section 3 presents data and methodology. Section 4 contains the results. Section 5 offers a conclusion and directions for future research.

This document is a template. An electronic copy can be downloaded from the conference website. For questions on paper guidelines, please contact the conference publications committee as indicated on the conference website. Information about final paper submission is available from the conference website.

II. LITERATURE REVIEW

Cross-selling is defined as the sale of additional, often related items to a customer who has previously made a purchase. One challenge in cross-selling is identifying whether an existing customer is likely to make a subsequent purchase and determining which product they are most likely to buy [6]. With the evolving landscape of financial services and products, banks are increasingly recognizing the importance of enhancing customer engagement and expanding their product offerings beyond traditional services. This study focuses on two key areas of the literature: the adoption of financial products and services by remittance recipients, the application of supervised classifiers in banking, and challenges in the implementation of ML models. Below, we present a summary of the existing literature in these areas.

A. Application of supervised classifiers in banking

Today, by harnessing the power of data science, financial institutions can gain valuable insights into customer's pre- and post-adoption behavior, preferences, and risk profiles. The use of customer datasets in this context represents a paradigm shift in how financial institutions enhance their marketing strategies. Increasingly, financial institutions are focusing on ML models to analyze existing customer data, aiming to boost satisfaction and identify customers who are more likely to purchase additional products [7].

Scholars have used historical data to predict future purchases. For instance, [8] analyzed data from an insurance company in Turkey to explore whether insurance purchase behavior could be predicted based on past customer behavior. The author suggests using DT algorithms as they show the best performance. [9] examined ML techniques to predict the likelihood of additional health insurance adoption in South Africa. The author employed random forest, K-nearest neighbors, XGBoost classifier, and LR to the data of existing

customers. This study found that RF performed best, achieving the highest accuracy and F1 score.

In another example, [10] applied the LR model to assess mobile phone usage for financial transactions. The results of this study demonstrate the statistical significance of the LR model. [11] used the LR model to predict bank performance, aiming to anticipate the potential bank failure and develop an early warning system. Similarly, [12] applied various models, including Extra Trees Classifier, RF, Categorical Boost Classifier, Light Gradient Boosting, and Extreme Gradient Boosting, to evaluate loan credit disapproval on vulnerable applicants. The results of this study showed that the Extra Trees Classifier and RF provided the highest accuracy. To detect distributed denial-of-service attacks on financial institutions, [13] applied SVM, K-nearest Neighbor algorithm, and RF. Comparative results show that SVM is more robust as compared to the K-nearest Neighbor algorithm and RF.

In credit scoring, [14] developed regression models to estimate borrowers' default probabilities. The author conclude that ML models outperformed FICO credit scoring. Among tested models, XGBoost showed higher accuracy but was less effective for cumulative expected credit loss. Although accuracy is the most important measure for model evaluation, it may not suffice for evaluating the ability of the credit-scoring model. [15] applied Stochastic Gradient Descent, K-nearest neighbor, and RF models to predict bank term deposit subscriptions. In this study, the best performance among the evaluated models was exhibited by RF. [7] highlighted Convolutional Neural Networks' superior performance over traditional models like RF and LR. The author underscores the potential of this model in transforming digital marketing strategies in banking. [16] introduced a class membership-based model. It is a transparent approach adapted to heterogeneous data that leverages nominal variables in its decision-making function.

B. Challenges in the implementation of ML models

Class imbalance is a common obstacle to obtaining accurate predictions. Highly imbalanced data biases any machine learning model towards the negative class. There are many approaches to rectify the class imbalance issue, such as random oversampling, random under-sampling, and the Synthetic Minority Over-Sampling Technique for Nominal and Continuous (SMOTE-NC). The results of the research of [17] show that the random under-sampling method used to balance the class distribution has the best performance on RF and Gradient Boosted Tree.

[18, 19] recommended balancing datasets using the Synthetic Minority Oversampling Technique (SMOTE-ENN). [21] compared the performance of four predictive models - artificial neural network, gradient boosting machines, and RF on different sampling methods. The author used random under-sampling, synthetic minority over-sampling technique, density-based synthetic minority over-sampling technique, and SMOTE-ENN for balancing data. The findings of this study indicate that SMOTE-based sampling techniques yielded promising results.

[21] recommended applying the Balanced RF1 and Random Under-Sampling Boost2 classifiers to address the class

imbalance. These algorithms randomly under-sample the majority class to balance the proportion between positive and negative observations. [22] evaluated various ML models with imbalanced data using sampling and ensemble techniques. The study demonstrates that the performance of the RF when data is balanced using the SMOTE over-sampling technique, was superior in terms of precision, recall, F-1, and accuracy.

After reviewing the literature on this topic, we decided to choose LR, DT, and RF models for our analyses, as they are well suited to our research questions and context. Model performance is evaluated using accuracy, precision, recall, and AUC curves. In the last step, we present SHAP values to identify the contribution of each factor.

III. DATA AND METHODOLOGY

Our study follows a typical data analytics approach. We analyze the business context, inspect the original data, prepare it for modeling, apply models, evaluate the performance of them and present SHAP values according to the standards of the discipline.

A. Methodology.

Logistic regression (LR) model is a widely used algorithm for binary classification in statistical methods and ML contexts. The model is particularly popular in banking. It is favored for its easy interpretation and implementation [11]. LR transforms probability into odds and then applies a logarithmic function to these odds. This ensures that estimated probabilities range between zero and one. This process effectively removes both the lower and upper bounds of the probability [10]. Scholars have compared their proposed methods to LR. The findings of these studies demonstrate the power of LR as a benchmark in credit scoring [14, 23].

Decision tree (DT) algorithm incrementally builds a tree structure by splitting data into increasingly smaller subsets. It uses Chi-square, Gini-Index, and Information gain methods. DT with improved Chi-square feature selection performs especially well in terms of recall for multiclass text classification. [22]. The model follows if-then rules. When an input condition is met, it returns a decision, predicting the target value. If not, it carries out further analysis [24].

Random forest (RF) is an ensemble learning method. It operates by creating multiple decision trees during training and providing a class mode in classification or mean in regression of individual trees. [22]. In this way, RF works more than a single tree from the classification results perspective. However, while RF generally provides strong results, gradient-boosted trees often surpass RF in accuracy [13].

Synthetic Minority Over-sampling Technique for Nominal and Continuous variables (SMOTE-NC) is an oversampling technique [17]. SMOTE-ENN technique first applies SMOTE to generate synthetic samples for the minority class and then uses Edited Nearest Neighbors to clean the dataset by removing noisy samples [19].

To evaluate model performance, we use Confusion Matrix, Accuracy, Recall, Precision, F1- Score, ROC-AUC. Accuracy measures the percentage of correctly predicted values. A

formula to calculate accuracy: $\text{Accuracy} = (TP + TN) / (P + TN + FP + FN)$ Where, TP stands for True Positives, TN=True Negatives, FP=False Negatives FP=False Positives. Recall - true positive rate, is the ratio of correctly predicted positives out of all actual positives: $\text{Recall} = TP / (TP + FN)$. Precision represents the proportion of correctly predicted positives out of all predicted positives: $\text{Precision} = TP / (TP + FP)$. F1-score is combining precision and recall: $F1 - \text{Score} = 2 * P * R / (P + R)$ [12].

The confusion matrix: visually distinguishes between classified and misclassified outcomes, with one axis representing predicted values and other actual values [13]. The receiver operating characteristic (ROC) curve visualizes classifier performance. It represents two- dimensional plots: false positive rates on the X-axis and true positive rates on the Y-axis [12]. The area under the curve (AUC) enables model comparison. Models with higher AUC values indicate greater accuracy [23].

B. Description of data.

A commercial bank of Uzbekistan provided data for this study. Table I presents the summary of our dataset.

TABLE I. SUMMARY OF DATA

Continues variables	count	mean	std	min	max
duration	15 149	3	5	0	21
avrg_check	15 149	471	720	0,001	10 000
number_of_reciepts	15 149	4	8	1	174
dummy variables					
	count	0		1	
female_reciever	15 149	13 054		2 095	
female_sender	15 149	13 162		1 987	
UK	15 149	14 792		357	
Germany	15 149	14 927		222	
Kazakhstan	15 149	14 731		418	
South_Korea	15 149	14 589		560	
Kyrgyzstan	15 149	14 954		195	
Latvia	15 149	14 735		414	
Lithuania	15 149	14 510		639	
Poland	15 149	14 355		794	
Russia	15 149	7 244		7 905	
Turkey	15 149	14 969		180	
Uzbekistan	15 149	12 935		2 214	
Finland	15 149	14 629		520	
USD	15 149	8 322		6 827	
Andijan	15 149	14 427		722	
Bukhara	15 149	9 370		5 779	
Fergana	15 149	14 199		950	
Jizzakh	15 149	14 895		254	
Karshi	15 149	14 620		529	
Khorezm	15 149	13 875		1 274	
Namangan	15 149	14 218		931	
Navai	15 149	14 567		582	
Karakalpakstan	15 149	14 486		663	
Samarkand	15 149	14 515		634	
Sirdarya	15 149	14 799		350	
Surkhandarya	15 149	14 899		250	
Tashkent_city	15 149	13 597		1 552	
car_loan	15 149	14 588		561	

Our dataset has three continuous variables and twenty-nine dummy variables. Duration is a continuous variable. It is measured in months and shows the time passed between the first and last receipts of remittances. The minimum value of zero indicates that the client received a single remittance. The maximum duration is equal to 21 months. On average, recipients receive remittances over three months. Avg_check represents the average check amount per recipient in US dollars. The values of avg_check range from 0.001 to 10,000 USD, with a mean of 471 USD. The number of receipts reflects the total quantity of remittances per recipient, varying from one to 174. On average, there are four remittances per recipient.

Female_reciever and female_sender indicate the gender of the recipient and sender respectively. Where a value of 1 signifies a female, and 0 represents a male. Remittances originating from 12 primary countries are represented by dummy variables (e.g., UK, Germany, Russia), which account for 95% of all the remittances to Uzbekistan. Remittances from other countries are grouped into a single variable, which we omit as required for dummy variable usage. In Uzbekistan, remittances can be sent in US dollars and Uzbek soums. USD variable indicates the currency of remittances, where 1 represents USD and 0 represents Uzbek soums. Uzbekistan has thirteen regions. Twelve regions are used to represent the originality of the recipient and one region is omitted.

Our target variable, car_loan is a dummy variable equal to 1 when the recipient has a car loan and 0 otherwise.

Of the receipts, 561 adopted car loans. This means we have an imbalanced dataset. Imbalanced data is a common challenge in classifier learning, as it can bias toward the majority class. Balancing the dataset is essential to improve model performance. We applied the SMOTENC technique to oversample the minority class. This technique allows oversampling data containing both categorical and continuous variables. After oversampling SMOTE-ENN dataset contains 12 630 instances of the positive class (1) and 10 080 instances of the negative class (0). SMOTE-NC dataset 14 588 instances of positive and negative classes.

IV. RESULTS

A. Logit model

Figure 2 presents the result of the LR model on the SMOTE-NC and SMOTE-ENN datasets. It includes the confusion matrix and ROC curve. For the SMOTE-NC dataset, the LR model correctly predicts 3 284 positive outcomes and 2 655 negative outcomes. The number of false positives and false negatives is 993 and 362, respectively (Figure 2 (a)). The area under curve (AUC) is 0.81 (Figure 2 (c)).

On the SMOTE-ENN dataset, the LR model correctly predicts 2 721 positive outcomes and 2 205 negative outcomes. The number of false positives and false negatives is 266 and 473, respectively (Figure 2 (b)). AUC is equal to 0.87 (Figure 2 (d)).

The first column of Table II presents accuracy, precision, recall and F1-scores of the LR model on the SMOTE-NC dataset. An accuracy score of 0.814, precision score of 0.768, recall score of 0.901 and F1-score of 0.829 indicate that the LR model performs well on the SMOTE-NC dataset.

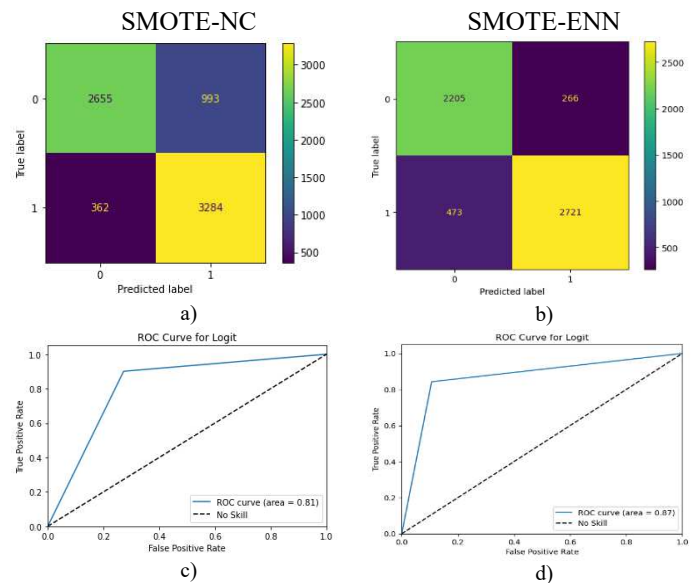


Figure 2. Confusion matrix and ROC curve of LR model

To evaluate potential overfitting, we test the model on five sub-datasets. We divide the entire dataset into five folds. In each iteration, we use one-fold as the test set. We train the model on the remaining data, using each fold once as a test set. This cross-validation method allows us to obtain scores for entire dataset. The cross-validation scores for the SMOTE-NC dataset vary across folds, ranging from 0.74 to 0.84. This variation signals a potential overfitting issue.

TABLE II. ACCURACY, PRECISION, RECALL AND F1-SCORES OF LR

	SMOTE-NC	SMOTE-ENN
accuracy score:	0.814	0.865
precision score:	0.768	0.911
recall score:	0.901	0.843
F1-Score	0.829	0.836
Cross-validation:	0.808 0.818 0.811	0.709 0.880 0.879
scores.mean:	0.803	0.862

The second column of Table II presents the accuracy, precision, recall and F1-scores for the SMOTE-ENN dataset. Accuracy score of 0.865, precision score of 0.911, recall score of 0.843 and F1-score of 0.836 indicate that the LR model performs better on SMOTE-ENN dataset than SMOTE-NC dataset. However, the cross-validation scores for the SMOTE-ENN dataset are not stable across folds, varying from 0.70 to 0.88. This significant variation suggests potential overfitting.

B. Decision tree

Figure 3 presents the results of the DT model on the SMOTE-NC and SMOTE-ENN datasets. On the SMOTE-NC dataset, the DT model correctly predicts 3 439 positive outcomes and 3 301 negative outcomes. The number of false positives and false negatives is 341 and 207 respectively (Figure 3 (a)). The area under curve AUC= 0.92 (Figure 3 (c)). On SMOTE-ENN dataset, DT model correctly predicts 3 071 positive outcomes and 2 335 negative outcomes. The number of false positives and false negatives is 136 and 123, respectively (Figure 3 (b)).

Accuracy score of 0.96, precision score of 0.96, recall score

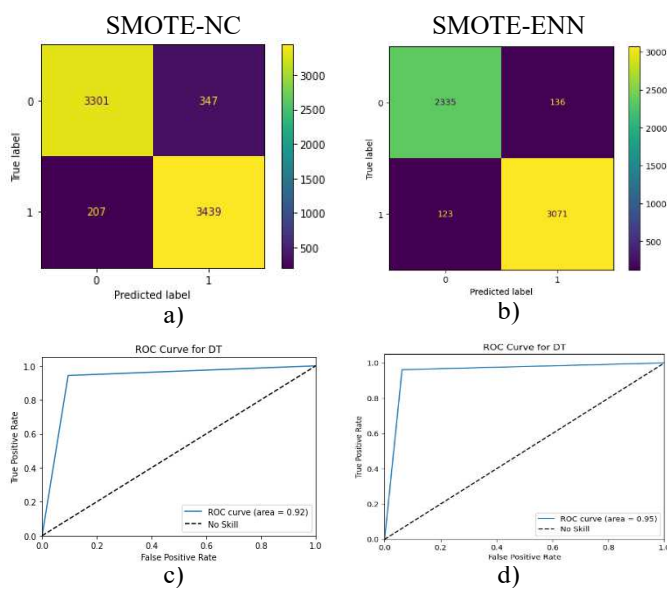


Figure 3. Confusion matrix and ROC curve of DT model.

of 0.96 and F1-score of 0.96 indicate that DT model performs very well on SMOTE-NC dataset. The cross-validation scores for the entire SMOTE-NC dataset are stable across folds and ranges from 0.926 to 0.934. This very small variation suggests that there no overfitting.

TABLE III. ACCURACY, PRECISION, RECALL AND F1-SCORES OF DT

	SMOTE-NC	SMOTE-ENN
accuracy score:	0.955	0.951
precision score:	0.960	0.952
recall score:	0.962	0.962
F1-Score:	0.961	0.957
Cross-validation:	0.927 0.926 0.930	0.927 0.932 0.928
scores.mean:	0.929	0.930

The second column of Table III presents the accuracy, precision, recall and F1-scores of DT model on the SMOTE-ENN dataset. An accuracy score of 0.95, precision score of 0.95, recall score of 0.96 and F1-score of 0.96 indicate that the DT model performs very well on SMOTE-ENN dataset. The cross-validation scores for entire SMOTE-ENN dataset are stable across folds and ranges from 0.927 to 0.934. This small variation indicates that there is no overfitting.

C. Random forest

Figure 5 presents the result of RF model on SMOTE-NC and SMOTE-ENN datasets. On the SMOTE-NC dataset, RF model correctly predicts 3 465 positive outcomes and 3 377 negative outcomes. The number of false positives and false negatives is 271 and 181, respectively (Figure 5 (a)). The area under curve AUC= 0.56 (Figure 5 (c)). On the SMOTE-ENN dataset, RF model correctly predicts 3 080 positive outcomes and 2 393 negative outcomes. The number of false positives and false negatives is 78 and 114, respectively (Figure 5 (b)). The area under curve AUC= 0.96 (Figure 5 (d)).

The first column of Table IV presents the accuracy, precision, recall and F1-scores of RF model in SMOTE-NC data. An accuracy score of 0.94, precision score of 0.93, recall score of

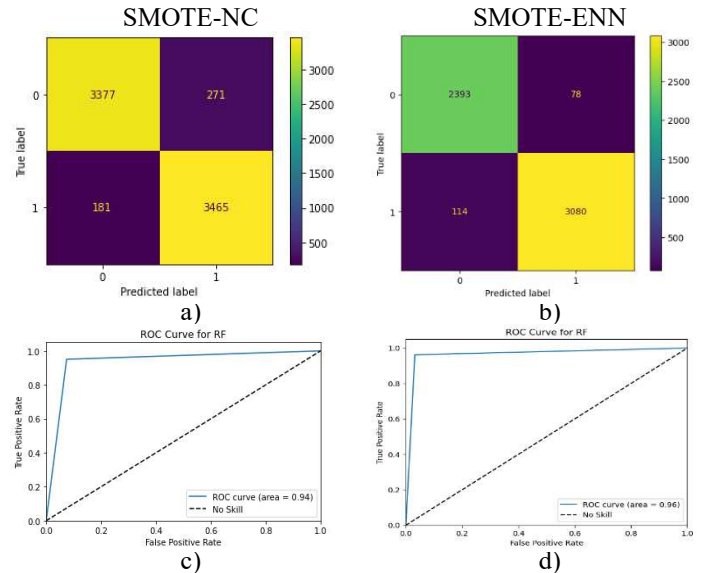


Figure 4. Confusion matrix and ROC curve of RF

0.95 and F1-score of 0.94 indicate that RF model performs very well on SMOTE-NC dataset. However, the cross-validation scores for entire SMOTE-NC dataset are not stable across folds and ranges from 0.93 to 0.95. This slight variation indicates that there a small overfitting issue.

TABLE IV. ACCURACY, PRECISION, RECALL AND F1-SCORES OF RF

	SMOTE-NC	SMOTE-ENN
accuracy score:	0.938	0.965
precision score:	0.927	0.975
recall score:	0.950	0.962
F1-Score:	0.939	0.968
Cross-validation:	0.950 0.940 0.942	0.952 0.939 0.943
scores.mean:	0.943	0.941

The second column of Table V presents the accuracy, precision, recall and F1-scores of RF model on the SMOTE-ENN dataset. An accuracy score of 0.96, precision score of 0.97, recall score of 0.96 and F1-score of 0.97 indicate that RF model performs very well on the SMOTE-ENN dataset. The cross-validation scores for entire SMOTE-ENN dataset are slightly unstable across folds and ranges from 0.924 to 0.952. This variation suggests a potential overfitting issue.

V. CONCLUSION

Remittance inflows are recognized as an integral component of economy of many developing countries. To increase the financial inclusion among remittance recipients, research is need to explore the cross-selling opportunities for these clients. The goal of this study is developing model to identify and assessing remittance recipients who are most eligible for a car loan.

This study analyses data on remittance transactions and car-loans, and applies LR, DT and RF models. To balance the data, we apply SMOTE-NC and SMOTE-ENN techniques. The results demonstrate that all the ML algorithms perform well on the SMOTE-ENN dataset, achieving high accuracy, recall, precision and F1-scores. However, among the algorithms considered, the DT shows superior overall performance.

This study contributes existing literature on remittances and financial services by examining the case of Uzbekistan. We utilize a long-term dataset comprising transactional data on remittance recipients and their adoption of car loans. Models that rely on huge number of variables are often impractical in real-world applications. Collecting such comprehensive data is challenges or in many cases almost impossible. By focusing on a limited number of variables, this study enhances the practical feasibility and broader applicability of its model.

The findings of this study are aligned with findings of [18, 19, 22, 20]. SMOTE-ENN and DT approaches outperform other algorithms. Based on these results we recommend data scientists using combination of SMOTE-ENN and DT for analyzing imbalanced dataset. A limitation of this study is the lack of consideration for external factors such as exchange rate. However, this limitation suggests the direction for future research. Another future direction is analyzing differences in loan adoption behavior of female and male remittance receivers.

Financial institutions can use the model developed in this study to organize cross-selling and capitalize on remittances. Additionally, financial institutions should develop new products and services tailored to match the needs remittance recipient household. Introducing products and services of diaspora banking and Islamic banking will lead to increase in financial inclusion among remittance senders and recipients.

REFERENCES

- [1] I. Ahunov, «Socio-economic consequences of labor migration in Uzbekistan, » Discussion Papers in Economics Economics, 2015.
- [2] K. Bondarenko, «The Impact of the COVID-19 Pandemic: The Case of Remittance Recipient Countries12, » International Organisations Research Journal, pp. 1-18, 2020.
- [3] J. Kakhkharov, «Squandering Remittances Income in Conspicuous Consumption? » в 30 Years since the Fall of the Berlin Wall: Turns and Twists in Economies, Politics, and Societies in the Post-Communist Countries, Singapore, Palgrave Macmillan, 2020, pp. 271-288.
- [4] CBU, «Press releases, » 13 06 2024. Available: https://cbu.uz/ru/press_center/releases/
- [5] I. Kalkan, «Evaluating cross-selling opportunities with recurrent neural networks on retail marketing, » Neural Computing and Applications, pp. 6247-6263, 2023.
- [6] S. Modak, «Machine Learning Model in Digital Marketing Strategies for Customer Behavior: Harnessing CNNs for Enhanced Customer Satisfaction and Strategic Decision-Making, » Journal of Economics, Finance and Accounting Studies, pp. 178-186, 2024.
- [7] S. Özdemir, «A case study on building a cross-selling model through machine learning in the insurance industry, » Avrupa Bilim ve Teknoloji Dergisi, pp. 364-372, 2022.
- [8] A. Mavundla, «Predicting Cross-Selling Health Insurance Products Using Machine-Learning Techniques, » Journal of Computer Information Systems, pp. 1-18, 2024.
- [9] F. Mbukwa, «The adoption of mobile financial services in Tanzania: The application of logit model, » Journal of Statistics and Management Systems, pp. 897-914, 2023.
- [10] R. Obeid, «Early warning of bank failure in the Arab region: A logit regression approach, » Asian Journal of Economics and Empirical Research, pp. 91-99, 2022.
- [11] P. Anand, «Prediction of loan behaviour with machine learning models for secure banking, » Journal of Computer Science and Engineering (JCSE), pp. 1-13, 2022.
- [12] M. Islam, «Detection of distributed denial of service (DDoS) attacks in IOT based monitoring system of banking sector using machine learning models, » Sustainability, p. 8374, 2022.
- [13] H. Munkhdalai, «An empirical comparison of machine-learning methods on bank client credit assessments, » Sustainability, pp. 1-23, 2019.
- [14] K. Zaki, «Predictive analytics and machine learning in direct marketing for anticipating bank term deposit subscriptions, » American Journal of Business and Operations Research, pp. 79-88, 2024.
- [15] M. Tékouabou, «A machine learning framework towards bank telemarketing prediction, » Journal of Risk and Financial Management, 2022.
- [16] A. P. Ratnasari, «Performance of Random Oversampling, Random Undersampling, and SMOTE-NC Methods in Handling Imbalanced Class in Classification Models., » Valley International Journal Digital Library, pp. 494-501, 2024.
- [17] S. Berkman, «Credit Card Fraud Detection with Data Sampling, » n 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), pp. 1-6, 2022.
- [18] M. Isaeva, «Remittances and financial dollarization, empirical evidence from Uzbekistan, » 2023 SeAH Haiam Scholarship Foundation Fellowship Papers, pp. 1-27, 2023.
- [19] V. Muaz, «A comparison of data sampling techniques for credit card fraud detection., » International Journal of Advanced Computer Science and Applications, 2020.
- [20] T. Haag, «Augmented cross-selling through explainable AI--a case from energy retailing, » arXiv preprint arXiv, 2022.
- [21] N. Dogra, «A comparative analysis of machine learning models for banking news extraction by multiclass classification with imbalanced datasets of financial news: challenges and solutions, » Special Issue on Artificial Intelligence in Economics, Finance and Business, pp. 35-52, 2022.
- [22] M. I. Utirovna, «Determinants of loan prepayment and comparison of machine learning approaches, » EEE World Conference on Applied Intelligence and Computing (AIC), pp. 719-725, 2022.
- [23] A. Khatir, «Machine learning models and data-balancing techniques for credit scoring: What is the best combination? » Risk, pp. 1-22, 2022.
- [24] M. Isaeva, «How can Uzbekistan engage its diaspora for national development? » CENTRAL ASIA AND THE CAUCASUS, т. 25, № 1, pp. 36-45, 2024.



Masuda Isaeva (Tashkent, Uzbekistan, 1983) earned her Bachelor's and Master's degrees in Computer Science from Tashkent State Technical University.

She is a Senior Data Scientist at Asakabank in Uzbekistan. With over four years of experience in finance-focused analytics, she is deeply engaged in optimizing financial mechanisms for remittance recipients and developing data-driven strategies for improving banking product adoption rates. Her expertise spans across AI model development, statistical modeling, and the implementation of machine learning solutions, specifically in retail loan cross-selling.