UNDERSTANDING THE HETEROGENEOUS IMPACT OF REMITTANCES ON SAVING BEHAVIOUR IN UZBEKISTAN: A MACHINE LEARNING APPROACH

Masuda Isaeva

Department of Retail Business, Asakabank, Tashkent, Uzbekistan

ABSTRACT

Uzbekistan is among the top remittance-receiving countries in the world. However, there is a lack of empirical research on how remittances influence recipients' financial inclusion and spending patterns. Existing studies primarily rely on survey-based or macroeconomic data, which often overlook key characteristics of remittances and their recipients.As a result, these studies reveal a homogeneous effect of remittances and fail to provide practical insights for financial institutions. This study addresses this research gap by exploring the factors shaping saving behaviour among remittance recipients in Uzbekistan. Using a dataset from 2017 to 2023, machine learning models, and SHAP values, we investigate the heterogeneous contribution of characteristics of recipients and remittances on the adoption of bank deposits. Our results reveal that gender, age, amount, and duration of remittances significantly influence deposit acceptance, while the origin of remittances has no effect. The insights derived from this study can aid policymakers and financial institutions in designing targeted strategies to encourage saving among remittance recipients and enhancing financial inclusion in country.

KEYWORDS

Bank deposits, machine-learning (ML) models, remittances, SHAPLEYvalues

1. INTRODUCTION

Financial inclusion refers to individuals' adoption and use of financial products and services [1]. It is an indicator of the availability, accessibility, and usage of the formal financial system by all participants in the economy. Account ownership and adoption of retail loans and deposits are fundamental measurements of financial inclusion. Financial inclusion is a gateway that enables individuals to use financial services to facilitate economic development [2].

Financial inclusion was neglected in Uzbekistan for a long time. In 2017, the government initiated reforms in the financial sector. The main goal of these reforms was to provide citizens, households, and SMEs with access to financial institutions. Development strategies included expanding financial inclusion by adopting digital payments, bank accounts, retail loans, and deposits[3].

As a result, access to financial services has increased since 2018. Between 2017 and 2023, the volume of retail loans grew tenfold. The volume of remittance inflows through financial institutions doubled by 2022 but slightly declined in 2023. The volume of retail deposits grew around fivefold, despite high interest rates[4](Figure 1).

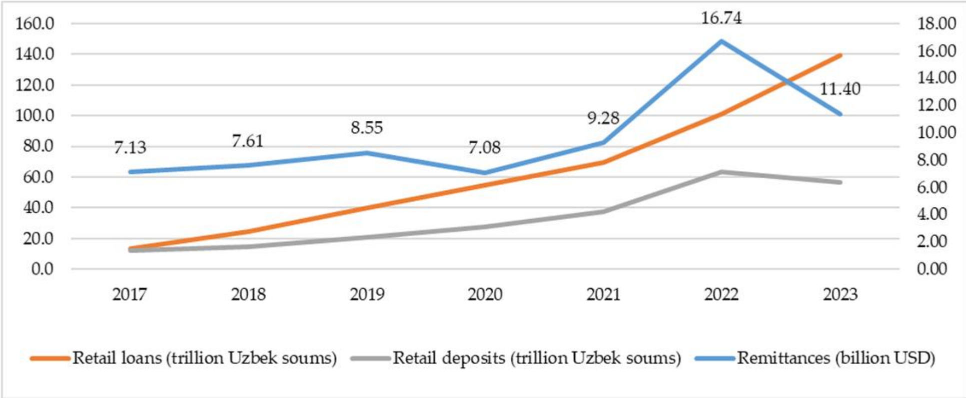


Figure 1. Growth of retail loans and deposits [4].

Migration and remittances are a primary source of income for many people. Remittance recipients often lack prior experience with banks, resulting in low awareness of financial products. However, they are likelier to adopt banking services and products than non-recipients [5].

Remittances are peer-to-peer cross-border money transfers to migrants' home countries, through official channels from the countries where they temporarily work or live[6]. Several factors affect how received remittances are spent. The origin of remittances or the destination country of the migrant is one of the factors determining factors how this money is spending. According to[7], in Kenya, remittances from North America significantly increased the adoption of bank deposits, while remittances from Europe showed no effect. In North and Central Asia, remittances from the US are saved and spent wisely [8]. Among the top ten remittance-receiving countries, remittances from developing countries have enhanced bank savings and account openings. In contrast, remittances from developed countries have facilitated the development of the stock market [9]. In Nigeria, remittances originating from within Africa have increased expenditures on education and food, while those from outside Africa are invested in real estate[6]. In Romania[10]no differences between remittances from EU and non-EU countries were found.

Such contradicting findings indicate that the impact of remittances varies by country, emphasizing the need for country-specific studies to gain accurate insights. Despite Uzbekistan being one of the top remittance-receiving countries in the world, there is limited research on the effect of remittances on saving behaviours. Moreover, existing literature relies on survey or macro-level data, which does not capture the specific features of remittances, the changing impact of variables over time, or the actual financial behaviour of recipients. The "black-box" nature of machine-learning models also limits the interpretability of findings. Consequently, these studies reveal a homogeneous effect of remittances, failing to provide practical insights for financial institutions. This study addresses these research gaps by utilizing transactional data from a commercial bank and employing machine-learning models and SHAP values to reveal the contribution of each value of independent variables.

This study makes several contributions to the literature. First, it is the first study to empirically investigate the effect of remittance characteristics and recipient profiles on saving behaviour. Second, using the latest transactional data, this research goes beyond the limitations of survey and macro-level data. Third, this study develops a highly accurate machine-learning model and explores the contribution of each value of independent variables. This novel approach enables the interpretation of results and reveals the heterogeneous effect of these variables on the outcome. Finally, this study offers practical recommendations for financial institutions looking to encourage remittance recipients to open deposit accounts, as well as for policymakers aiming to leverage remittances to enhance financial inclusion and drive economic growth.

The rest of this paper is structured as follows: Section 2 provides a review of relevant literature. Section 3 introduces the dataset and outlines the methodology. Section 4 evaluates the performance of machine learning models and examines the influence of key variables. Section 5 presents a discussion, while Section 6 summarizes the findings and concludes the study.

2. LITERATURE REVIEW

2.1. Impact of Origin of Remittances on Financial Inclusion

Many studies have found controversial findings about whether and in what way remittances benefit financial inclusion and economic development at both macro and micro levels. It is generally agreed that remittances reduce poverty and stimulate financial inclusion. Here, we move beyond the general discussion on the overall impact of remittances to focus on the specific characteristics of remittances and receivers, as the magnitude of the effect varies among these characteristics.

Many remittance corridors exist between developed and developing countries, each with unique characteristics regarding volume, sending cost, transfer method, and influence on the spending behaviour of recipients (Table 1). Studies analyze differences in remittances received from these corridors. According to [11], remittances from middle-income countries have increased the usage of ATMs and debit cards in Nigeria. In Kenya, remittances from North America have had a positive and statistically significant effect on deposit volumes, while remittances received from Europe have shown no effect [7]. [9, 8] explain these differences by the skill levels of migrants. The authors suggest that remittances from relatively low-skilled migrants significantly impact financial inclusion more than those from their high-skilled counterparts.

Table 1. Summary of literature on the impact of remittance corridors

Relationship	**Result**	**Sending country**	**Receiving country**	**Author**
Remittances → Bank Deposits	Remittances from North America positively affect the adoption of bank deposits, while the impact of remittances from Europe is insignificant.	North America, EU, Rest of the world	Kenya	[7]
Remittances → Financial Inclusion	Remittances from middle-income countries increase the usage of ATMs and debit cards.	Middle-income countries, High-income countries	Nigeria	[11]
Remittances → Entrepreneurship	Remittances from the Gulf Cooperation Council (GCC) and developed countries positively impact entrepreneurship compared to remittances from India.	Gulf States, Developed countries, India	Nepal	[12]

Remittances → Poverty and Inequality	Remittances from India decrease income inequality and have the largest impact on poverty reduction.	India, developed countries	Nepal	[13]
Remittances → GDP	There is no difference in the effect of remittances from EU countries and non-EU countries.	EU, Rest of the world	Romania	[10]
Remittances → spending behaviour	Remittances from African countries increase expenditures on education and healthcare, while remittances from outside Africa increase investments in real estate.	African counters, Rest of the world	Nigeria	[6]
Remittances → financial development	Remittances from GCC countries have a more significant impact on financial inclusion than those from G7 countries.	GCC countries, G7 countries	China and India	[9]
Remittances → spending behaviour	Remittances from Western Europe, the United States, and Canada increase healthcare and education expenditures more than those from GCC countries.	Western Europe, the US, Canada, GCC countries	Kerala (India)	[14]

However, conflicting conclusions also exist. In Nepal, remittances from India, which are much lower on average, have shown the most significant impact on poverty reduction compared to remittances from developed countries[13]. In Romania, [10] found no difference between the effects of remittances received from EU countries and those from non-EU countries.

Besides money, migrants transfer technologies and business ideas to their home countries. By reducing credit constraints and increasing the availability of capital, remittances create favourable conditions for entrepreneurship[15]. In Nepal, return migrants from GCC and developed countries have a higher probability of starting their businesses than return migrants from India [12]. Examining migration from Moldova [16] found that migrants in EU countries tend to remit longer than migrants in CIS countries.

2.2. Determinants of Saving Behaviour

Savings behaviour refers to the financial choices individuals or households make when allocating their income between consumption and savings. This decision is influenced by various factors, including the availability of formal savings instruments and trust in financial institutions, which in turn affect investment rates and overall economic growth[24].

The most commonly used control variables in research on saving behaviour are age and gender. The significant effect of the age variable in many studies suggests that older individuals are more likely to hold deposit accounts in financial institutions [17]. As people age, their intention to save more for post-retirement life increases. Receiving remittances also positively contributes to the formation of saving behaviour among young people [18, 19]explains a non-linear relationship between age and savings through financial behaviour, noting that while middle-aged individuals have higher financial literacy, younger and older individuals exhibit more active financial behaviour. However, this pattern does not hold universally. Empirical evidence from studies by [20, 21, 13] reveals no relationship between age and the savings behaviour of remittance recipients.

Men and women demonstrate significantly different saving behaviours. First, women tend to prioritize daily consumption, such as food and other household needs, over savings[22]. Second, women often do not make financial decisions, while men are typically responsible for the household's financial affairs. Consequently, men are more active in adopting financial products, including bank deposits [23, 18]. Third, women are more likely to save for short-term needs, while men save for medium- and long-term goals [24]. Finally, women generally have a lower risk tolerance, which negatively affects their saving behaviour[25]. However, [26, 27, 19] have found that gender has an insignificant effect on saving behaviour.

Restrictions on international travel and customs controls during the COVID-19 pandemic severely limited the use of informal remittances, such as carrying cash. As a result, remittance senders had to rely on mobile (digital) money transfers. This shift to mobile remittances improved users' digital and financial literacy and facilitated the adoption of other financial services by previously unbanked individuals [28, 26, 29]. The increased use of mobile has led to a rise in mobile deposit accounts[30]. While mobile money usage does not necessarily encourage people to save more, it promotes the transition from informal saving to formal saving. Thus, mobile remittances serve as a gateway to financial inclusion, amplifying their overall impact [31].

2.3. Utilizing ML models in banking

In the modern financial sector, leveraging data science enables institutions to gain deep insights into customers' behaviour before and after product adoption, as well as their 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 machine-learning models to analyze existing customer data, aiming to boost cross-selling, revenue, and service levels [32].

Researchers have used historical data to predict future purchases. For instance, [33]analyzed data from an insurance company in Turkey to explore whether insurance purchase behaviour could be predicted based on past customer behaviour. The author suggests using the DT algorithm, as it shows the best performance. [34] examined machine-learning techniques to predict the likelihood of additional health insurance adoption in South Africa. The author employed random forest, K-nearest neighbors, XGBoost, and LR on data from existing customers. This study found that RF performed best, achieving the highest accuracy and F1 score.

Traditionally, retail banks have used demographic data and statistical models to predict a customer's likelihood of purchasing financial products. [35]recommends addressing the limitations of this approach by incorporating multi-year transactional data. The study's findings highlight that transaction data provides valuable insights into customer behaviours and significantly enhances the predictive accuracy of cross-selling models. [36]introduced a hybrid model utilizing recurrent neural networks to process both sequential and non-sequential features, capturing behavioural and non-behavioural customer-specific information. When tested on online shopping data, the model demonstrated that combining relevant features improves predictive performance, enhancing evaluation metrics by 12%.

In another example, [37] applied the LR model to assess mobile phone usage for financial transactions, demonstrating its statistical significance. [38] used the LR model to predict bank performance, aiming to anticipate potential bank failure and develop an early warning system. Similarly, [39] applied various models, including Extra Trees Classifier, RF, Categorical Boost Classifier (CatBoost), 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, [40] applied SVM, K-nearest neighbors algorithm, and RF. Comparative results showed that SVM was more robust than K-nearest neighbors and RF.

In credit scoring, [41] developed regression models to estimate borrowers' default probabilities. The author conclude that machine-learning models outperformed FICO credit scoring. Among the tested models, XGBoost showed higher accuracy but was less effective for cumulative expected credit loss. Although accuracy is an important measure for model evaluation, it may not suffice for evaluating the performance of the credit-scoring model. [42]applied stochastic gradient descent, K-nearest neighbors, and RF models to predict bank term deposit subscriptions. In this study, RF exhibited the best performance among the evaluated models. [32] highlighted Convolutional Neural Networks' superior performance over traditional models like RF and LR. The author emphasized the potential of this model in transforming digital marketing strategies in banking. [43] introduced a class membership-based model. It is a transparent approach adapted to heterogeneous data that leverages nominal variables in its decision-making function.

After reviewing the literature, we decided to choose XGBoost, CatBoost, and RF models for our analyses, as they align with our research questions, context, and data. Model performance is evaluated using accuracy, precision, recall, F1 scores, and receiver operating characteristics (ROC and AUC). In the last step, we present SHAP values to identify the contribution of each value of each factor.

3. MODEL, DATA, AND METHODOLOGY

3.1. Model

The basic regression model in this study is:

$$Sav_i = \alpha_i + \beta_i R_i + \gamma_i X_i + \varepsilon_i \qquad (1)$$

where Sav_i is a dummy variable indicating whether an individual has a bank deposit or not. $R_i$denotes remittances, $X_i$represents the recipient's characteristics, α, β, and γ are the estimated regression coefficients, and ε_i signifies the error term.

The characteristics of remittances play a crucial role in remittance analysis [6]. To comprehensively account for these factors, we incorporate five variables: the origin of remittances, type (online or cash), frequency of receipts, total amount received, and duration of remittance inflows. Table 2 presents definitions and sources for the variables.

Table 2. Definitions and sources of variables

Variable	Definition	Source
Saving	An individual having a deposit account in a financial institution	[44]
Amount of remittances	The total amount of remittances received	[11]
Number of receipts	Number of remittance transactions	[45]
Duration of remittance receipt	The time between the first and latest remittance receipts	[46]
Origin of remittances	Remittance-sending country	[47]
Type of transactions	Remittances received via online transaction or cash	[48]
Age	Age of recipient	[27]
Gender	Gender of recipient	[27]

3.2. Data

Our study follows a typical data analytics approach. We analyze the business context, inspect the original data, prepare it for modeling, apply models, evaluate their performance, and present SHAP values per disciplinary standards.

A challenge in studying the relationship between remittances and bank deposits is the lack of datasets containing transactional data on both. In this study, data is provided by a commercial bank in Uzbekistan. The remittance data represents transactions of recipients who received remittances via KoronaPay. The bank deposits data reflects the presence of fixed-term or savings deposits. The final dataset consists of aggregated remittance and bank deposit data.

The total remittance amount represents the cumulative sum of all received remittances, measured in millions of Uzbek soums. The receipt duration denotes the number of months between the first and last remittance transactions. Dummy variables are used to categorize remittances from 11 countries (e.g., the UK, Germany, and Russia), covering 98% of total remittances to Uzbekistan.Remittances from other countries are combined into a single variable. Additionally, internal remittances from Uzbekistan are grouped into a separate dummy variable, which is excluded as per the requirements for dummy variable implementation.

Country dummy variables represent the origin of each remittance. If a recipient received remittances from multiple countries, the country with the largest remittance amount is assigned. Transaction type is coded as “1” if money is received through a mobile application and “0” for cash receipt. If a customer receives both types, the method with the highest amount is assigned. Age is treated as a continuous variable, while gender is represented as a dummy variable, assigned a value of “1” for females and “0” for males. The dependent variable, “saving,” takes the value of “1” if the recipient holds a savings or fixed-term deposit and “0” if they do not. Table 3 describes the dataset.

Table 3. Description of data

continues variables	**count**	**min**	**max**
number_of_receipts	153,444	1,0	185,0
amount_of_remittances	153,444	0,01	2 712,6
duration_of_remittance_receipt	153,444	0,0	69,0
age	153,444	16	87
dummy variables	**1**	**0**	
gender (female/male)	76,766	76 678	
type_of_transactions (online/offline)	88,001	65 443	
UK	432	153 012	
Israil	1,110	152 334	
Kazakhstan	16,876	136 568	
South_Korea	1,589	151 855	
Khyrgystan	3,003	150 441	
Russia	120,739	32 705	
USA	164	153 280	
Turkey	4,303	149 141	
Finland	1,022	152 422	
Czech_Republic	339	153 110	
other_countries	2,957	150 487	
saving	2,529	150 915	

Our dataset includes continuous and dummy variables, with an imbalanced outcome variable that is often “0”. To address this issue, we sample the dataset using the SMOTE-ENN approach. After applying this method, the size of our dataset is increased, with 141,801 instances for “1” in and 131,234 “0” values.

3.3. Methodologies

Tree boosting is one of the most effective machine learning methods. In this study, we utilize three classification models: categorical boosting (CatBoost), extreme gradient boosting (XGBoost), and random forest (RF). CatBoost is a gradient-boosting algorithm based on balanced decision trees with a symmetrical structure. It has two main advantages over other supervised machine-learning models. First, it saves time during the data preprocessing stage as it handles all variable types, including numeric, categorical, and text data. Second, it efficiently uses CPU resources to reduce prediction time and prevent overfitting [5].

XGBoost utilizes multiple cores to create trees and organizes data to minimize search time. This feature allows it to have a shorter training time and better performance than other gradient-boosting algorithms [49]. In addition, gradient-boosting algorithms often outperformother classifiers in accuracy[40].

Random Forest (RF) is an extension of the bagging technique, generating multiple decision trees during training. [50]. RF functions by combining multiple tree predictors, where each tree depends on randomly selected values with identical distributions for all trees. Its performance is influenced by the predictive strength of individual trees and their correlation[51].

To evaluate model performance, we use confusion matrix, accuracy, recall, precision, F1- Score, and ROC-AUC recommended by [39, 52, 41, 53]. Accuracy represents the percentage of correct predictions, calculated as: Accuracy = $(TP + TN) / (P + TN + FP + FN)$, where TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives) represent classification outcomes. Recall, or true positive rate, is the ratio of correctly predicted positives out of all actual positives: Recall = TP / (TP +FN). Precision is the proportion of correctly predicted positives out of all predicted positives: Precision = TP / (TP +FP). F1-score combines precision and recall: $F1 - Score = 2{*}P{*}R / (P + R)$ [39].

The confusion matrix visually distinguishes classified and misclassified outcomes, with one axis representing predicted values and other actual values[40]. The receiver operating characteristic (ROC) curve plots the false positive rate on the X-axis and the true positive rate on the Y-axis [39]. The area under the curve (AUC) enables model comparison. Higher AUC values indicate greater accuracy [53].

We apply Shapley additive explanations (SHAP values) to provide interpretability for machine-learning models. SHAP values are model-independent, relying only on model outputs and observed values. Unlike methods such as average impurity reduction or feature permutation, which provide feature importance without insight into individual values, SHAP values reveal the effects of different variable values on outcomes, transforming “black box” models into “white box” models[54]. Thus, SHAP values transform “black box” machine-learning models into “white box” models.

4. RESULTS

CatBoost: first, we predict the outcome variable on original and sampled datasets using the CatBoost classifier (Figure 2). The CatBoost model demonstrates better performance in terms of the confusion matrix and ROC curve compared to the original dataset. For the original dataset, the CatBoost model correctly predicts 5 positive outcomes and 37,882 negative outcomes. The number of false positives and false negatives is 9 and 672, respectively (Figure 2 (a)). The area under the curve (AUC) is 0.50 (Figure 2 (b)).

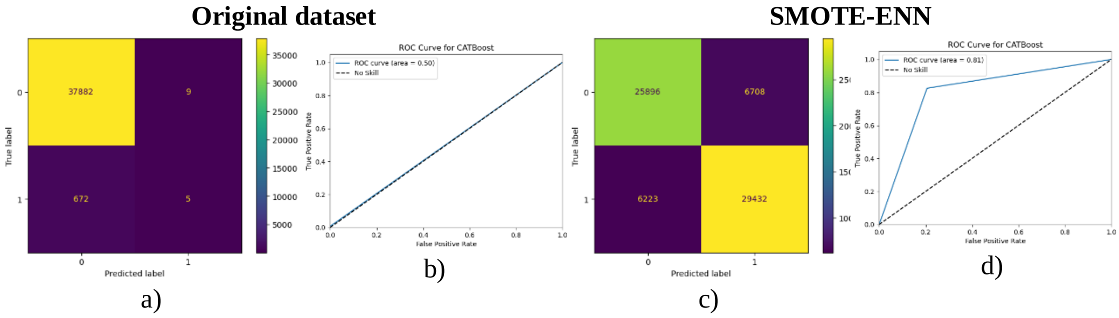


Figure 2. Confusion matrix and ROC curve of CatBoost model on original and sampled (SMOTE-ENN) datasets

On the sampled dataset, the CatBoost model correctly predicts 25,896 positive outcomes and 29,432 negative outcomes. The number of false positives and false negatives are 6,708 and 6,223, respectively (Figure 2 (c)). AUC is equal to 0.81 (Figure 2 (d)).
The first column of Table 4 displays the accuracy, precision, recall, and F1-scores for the CatBoost model when applied to the original dataset. The results—an accuracy of 0.982, precision of 0.429, recall of 0.004, and F1-score of 0.009—suggest that the CatBoost model performs poorly on the original dataset.

To assess potential overfitting, we evaluate the model using five sub-datasets. The dataset is divided into five folds, with each fold serving as the test set in one iteration while the model is trained on the remaining data. This cross-validation technique ensures that each fold is used as a test set once, allowing us to compute performance scores across the entire dataset. The cross-validation results for the original dataset remain consistent across folds, ranging from 0.982 to 0.984, indicating that overfitting is not an issue.

Table 4. Accuracy, precision, recall, and F1-scores of CatBoost model on original and SMOTE-ENN datasets.

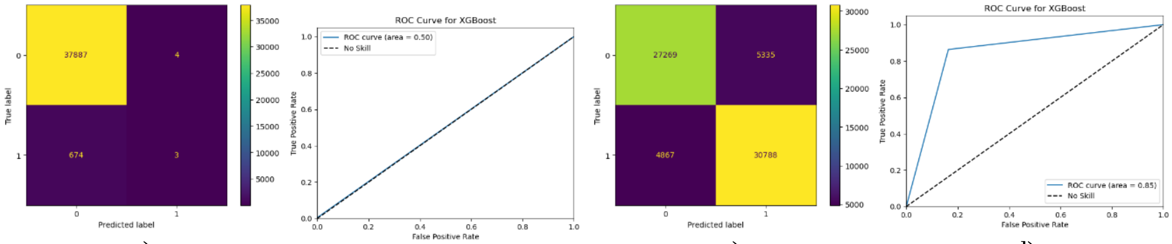
	Original dataset	**SMOTE-ENN**
accuracy_score:	0.982	0.811
precision_score:	0.429	0.814
recall_score:	0.004	0.825
F1-Score	0.009	0.820
Cross-validation:	0.982 0.983 0.983 0.983 0.984	0.778 0.781 0.817 0.805 0.803
scores.mean:	0.964	0.797

The second column of Table 4 presents the accuracy, precision, recall, and F1-scores on the SMOTE-ENN dataset. An accuracy score of 0.811, precision score of 0.814, recall score of 0.825, and F1-score of 0.820 indicate that the CatBoost model performs much better on the SMOTE-ENN dataset than on the original dataset. However, the cross-validation scores for the

SMOTE-ENN dataset are less stable across folds, varying from 0.778 to 0.817. This small variation suggests a minor overfitting issue.

XGBoost: The number of false positives and false negatives is 4 and 674 respectively (Figure 3 (a)). The area under curve AUC= 0.50 (Figure 3 (b)).

Original dataset **SMOTE-ENN**



a) b) c) d)

Figure 3.Confusion matrix and ROC curve of XGBoost model on original and sampled (SMOTE-ENN) datasets

On the SMOTE-ENN dataset, the XGBoost model correctly predicts 27,269 positive outcomes and 30,788 negative outcomes. The number of false positives and false negatives is 5,335 and 4,867, respectively (Figure 3 (c)). The area under the curve AUC= 0.85 (Figure 3 (d)).

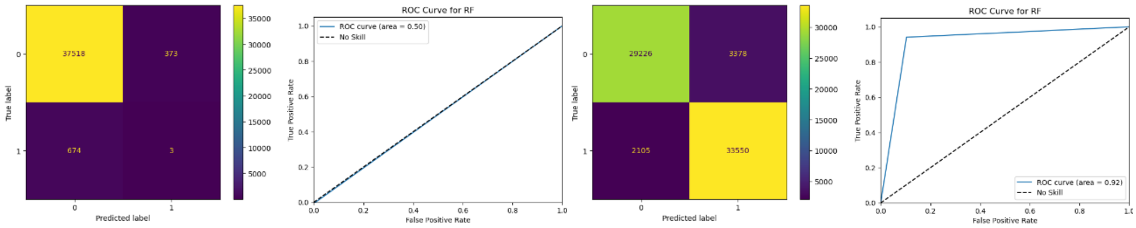
The first column of Table 5 presents the accuracy, precision, recall, and F1-scores for the XGBoost model on the original dataset. An accuracy score of 0.98, precision score of 0.0.43, recall score of 0.004, and F1-score of 0.009 indicate that the XGBoost model performs poorly on the original dataset. The cross-validation scores for the original dataset are stable across folds, ranging from 0.982 to 0.984. This very small variation suggests no overfitting.

Table 5. Accuracy, precision, recall, and F1-scores of XGBoost model on original and SMOTE-ENN datasets.

	Original dataset	SMOTE-ENN
accuracy_score:	0.982	0.851
precision_score:	0.429	0.852
recall_score:	0.004	0.864
F1-Score	0.009	0.858
Cross-validation:	0.982 0.983 0.984 0.984 0.984	0.812 0.820 0.852 0.835 0.833
scores.mean:	0.964	0.830

The second column of Table 5 presents the accuracy, precision, recall, and F1-scores of the XGBoost model on the SMOTE-ENN dataset. An accuracy score of 0.851, precision score of 0.852, recall score of 0.864, and F1-score of 0.858 indicate that the XGBoost model performs very well on the SMOTE-ENN dataset. The cross-validation scores for the SMOTE-ENN dataset are stable across folds, ranging from 0.812 to 0.852. This slight variation suggests a small overfitting issue.

Original dataset **SMOTE-ENN**



a) b) c) d)

Figure 4. Confusion matrix and ROC curve of RF model on original and sampled (SMOTE-ENN) datasets

RF: On the original dataset, the RF model correctly predicts 3 positive outcomes and 37,518 negative outcomes. The number of false positives and false negatives are 373 and 674, respectively (Figure 4 (a)). The area under curve AUC= 0.50 (Figure 4 (b)). On the SMOTE-ENN dataset, the RF model correctly predicts 29,226 positive outcomes and 33,550 negative outcomes. The number of false positives and false negatives are 3,378 and 2,105, respectively (Figure 4 (c)). The area under curve AUC= 0.92 (Figure 4 (d)).

The first column of Table 6 presents the accuracy, precision, recall, and F1-scores of the RF model on the original dataset. An accuracy score of 0.97, precision score of 0.008, recall score of 0.004, and F1-score of 0.006 indicate that the RF model performs poorly on the original dataset. However, the cross-validation scores for the original dataset are stable across folds, ranging from 0.969 to 0.974. This slight variation indicates that there is no overfitting issue.

Table 6. Accuracy, precision, recall, and F1-scores of RF model on original and SMOTE-ENN datasets.

	Original dataset	**SMOTE-ENN**
accuracy_score:	0.973	0.920
precision_score:	0.008	0.910
recall_score:	0.004	0.941
F1-Score	0.006	0.924
Cross-validation:	0.974 0.972 0.974 0.969 0.972	0.914 0.915 0.924 0.909 0.916
scores.mean:	0.972	0.915

The second column of Table 6 presents the accuracy, precision, recall, and F1-scores of the RF model on the SMOTE-ENN dataset. An accuracy score of 0.920, precision score of 0.910, recall score of 0.941, and F1-score of 0.924 indicate that the RF model performs very well on the SMOTE-ENN dataset. The cross-validation scores for the SMOTE-ENN dataset are stable across folds and range from 0.909 to 0.924. This variation suggests that there is no overfitting issue.

We do not use any hyper-parameter optimization, as our all models demonstrate high performance. These results show that all machine-learning algorithms perform well on the SMOTE-ENN dataset, achieving high accuracy, recall, precision, and F1-scores. Among the algorithms considered, the RF model shows superior overall performance, with an accuracy score of 0.992, a precision score of 0.997, a recall score of 0.987, and F1-score of 0.992. The consistent cross-validation scores across folds indicate no overfitting. Therefore, the combination of the RF model and SMOTE-ENN sampling method is recommended as the optimal choice for building deposits cross-selling prediction models on imbalanced data.

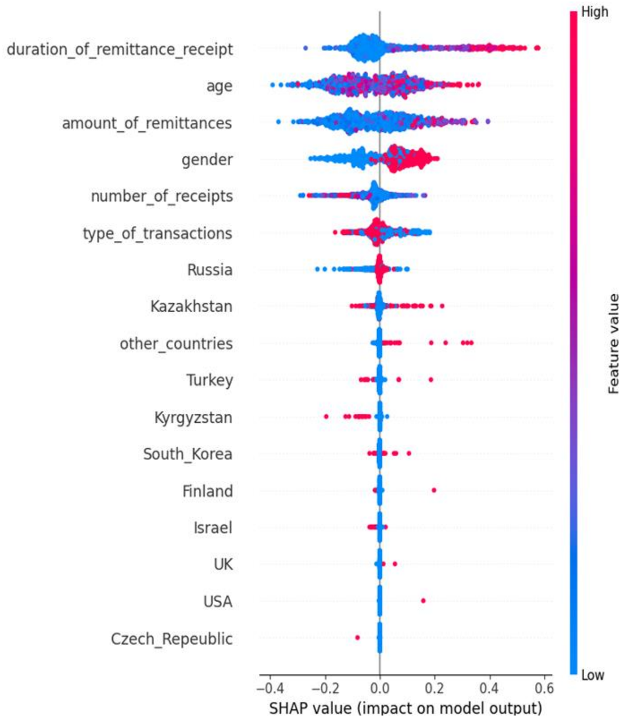


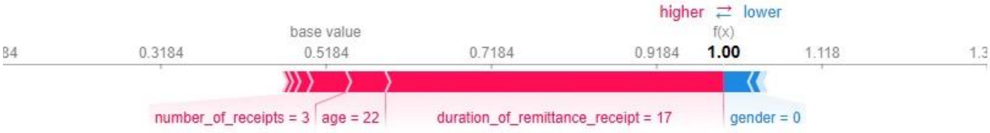
Figure 5. Summary plot of SHAP values for the RF model

SHAP values: in the next step, we calculate SHAP values for the RF model to understand the contribution of each value of the independent variables to a particular outcome. A summary plot visualizes the relationship between an independent variable and an outcome variable(Figure 5).

Figure 5 shows that high values of remittance amount, remittance receipt duration, gender, and age have a strong positive impact on the model's output. In contrast, low values for the number of receipts and transaction types are associated with a positive impact on the model's outcome. This means that online remittances are less likely to be saved compared to offline remittances. Low and high values for all countries, except Kyrgyzstan, have minimal impact. High values of Kyrgyzstan, however, have a strong negative impact.

Next, we analysed the contribution of features to specific "0" and "1" outcomes. SHAP force plots provide insights into how features contribute to the model's prediction. Force plots are ideal for explaining how a model calculates a particular outcome (Figure 6). In the force plot, features with larger impacts on the predicted outcome are located closer to the dividing line between red and blue. The width of each cell indicates the size of the impact. In Figure 6 (a), the age is "28" and the gender is "0", indicating the client is a 28-year-old man. The amount of remittances received is 3 million soums. A remittance receipt duration of "0", and the number of receipts of "1", indicate that this is a new client. A transaction type of "1" means that remittances are received online. Red represents features that increase the outcome, while blue represents features that lower it.

a)



b)

Figure 6. Force plot of SHAP values on “0” (a) and “1” (b) outcomes

Figure 6(b) shows the force plot for an instance classified as adopting a bank deposit (savings = 1). The predicted value for this instance is f(x) = 1. The client is a 22-year-old man. The blue colour of the gender variable indicates a negative contribution of this value to the predicted outcome A remittance receipt duration of “17” indicates that the client has been using the bank for over a year. A number of receipts equal to “3” indicates this is the client’s third receipt The red colour of these features indicates that they increase the predicted outcome.

In the final stage, we utilize a stacked force plot to illustrate the nonlinear relationship between the outcome variable and key independent variables, including age, remittance amount, and remittance duration. These variables are chosen due to their varied impact, as observed in Figure 5. The stacked force plot of SHAP values is especially effective in visualizing how different values of specific independent variables influence the outcome (Figure 7). The analysis reveals that age has a strong positive effect on the outcome variable, particularly for individuals aged 31 to 36 and those over 63.

The effect of age on the outcome variable is high and positive for clients aged 31 to 36 and over 63. This can be explained by the tendency of clients aged 30-36 to save money for their children’s education or future expenses. In Uzbekistan, elderly parents of migrants frequently receive remittances from their children working abroad. This money is often saved by parents, who are typically over 60 years old, to support their children in purchasing real estate or vehicles upon their return. Serving as trusted custodians, these parents effectively function as informal savings managers, safeguarding the financial resources of their migrant children.

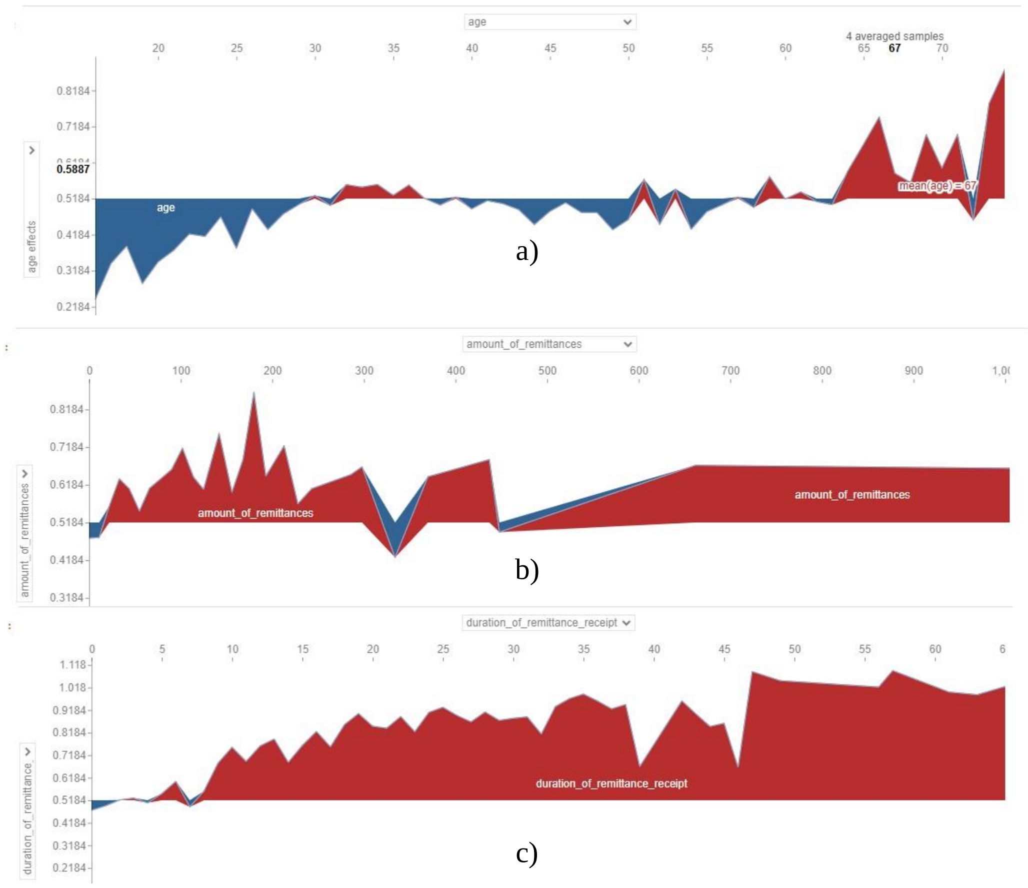


Figure 7. Stacked force plot (a) age, (b) amount of remittances, (c) duration of remittance receipts

The impact of the remittance amount is high and positive for clients receiving between 30-300 million soums and those receiving more than 450 million soums. These amounts suggest that clients receiving up to 30 million soums spend it on basic needs, whereas clients receiving over 30 million soums start to save. However, remittances between 300-450 million soums have the lowest impact. The effect of remittance duration is positive starting from the fourth month and increases consistently. This trend can be explained by the fact that migrants initially need to cover essential expenses during the first 3-4 months before they start saving.

5. DISCUSSION

Our findings indicate that the origin of remittances does not significantly influence the saving behaviour of remittance recipients in Uzbekistan. This result aligns with the results of [62, 63] and contrasts with other studies. These discrepancies may be attributed to the scope of our analyses. We use eleven countries to represent the origin of remittances. Previous research often focused on comparisons between two countries or regional groups such as the EU, GCC, or domestic and international remittances.

In Uzbekistan, female recipients are more likely to adopt bank deposits than their male counterparts. This finding contradicts previous research of[23, 18] and others. A potential explanation for this difference is gender distribution within our dataset. Our dataset includes an almost equal distribution of male and female clients, unlike previous studies, which were often male-dominated. Additionally, we find that online remittances negatively impact saving behaviours, contradicting the conclusions of [28, 29, 26]and other researchers. This discrepancy may be attributed to differences in data collection periods. The cited studies relied on data gathered before or during the COVID-19 pandemic, whereas our dataset focuses on the post-pandemic period, during which the previously observed positive influence of online transactions may have diminished.

In Uzbekistan, the saving behaviours of remittance recipients are influenced by factors such as the age and gender of the recipient and the amount and duration of remittances. The duration of remittance receipts emerges as the strongest predictor of adopting a bank deposit. However, the impact of these factors varies as their values increase, suggesting heterogeneity in their effects.

6. Conclusion

The impact of remittances on financial inclusion remains a topic of debate in policy discussions and country-level studies, with findings often producing mixed results. It is generally accepted that remittances stimulate financial inclusion. In this study, we move beyond the overall impact of remittances and focus on the specific characteristics of remittances and recipients, as the magnitude of the effect varies depending on these variables.

The primary purpose of this study is to analyse the factors shaping the saving behaviour of remittance recipients. To achieve this, we use data from clients who received remittances between 2018-2023 as well as data from clients who hold fixed-term or saving deposits in a commercial bank in Uzbekistan. Our dependent variable is categorical (whether or not the individual has a deposit account), and we assume a nonlinear relationship between the independent and dependent variables. We apply the CatBoos, XGBoost, RF models, and SMOTE-ENN data sampling methods. We use SHAP values to interpret the outputs of these models and reveal the heterogeneous impact of different values of independent variables.

Our findings indicate that gender, age, remittance amount, and remittance duration significantly influence the likelihood of deposit adoption. Female recipients are more likely to hold bank deposits than their male counterparts, while remittance duration emerges as the strongest predictor of saving behaviour. Additionally, cash remittances are more frequently saved than online transfers. Age has a consistently positive effect on savings after 60, whereas the country of remittance origin does not show a significant impact.

Based on these insights, we propose several recommendations for financial institutions:

- Financial institutions should work to increase remittance inflows to attract more retail deposits. Key steps include reducing transaction fees and developing new, affordable, secure online money transfer services.
- Although remittances are a low-profit product, they increase the adoption of bank deposits. Banks should target remittance recipients to promote deposit products.
- Potential clients should be targeted based on age, gender, or remittance amount. Targeting customers based on the origin of remittances should be avoided.
- Financial institutions should build a recommender system based on machine-learning models similar to ours and deploy it to attract more deposit holders.
- Financial institutions should use remittances as an entry point to attract previously unbanked individuals and households into the financial system. A public campaign aiming at enhancing the digital and financial literacy of remittance recipients should be organized.

- Financial institutions should collect reliable and comprehensive data on remittances. The availability of data is central to making accurate decisions. The Central Bank of Uzbekistan should organize the collaboration between commercial banks to enhance the data collection on remittance inflows, senders and recipients' demographics, transaction details, country of origin, and transaction type. This data is essential for drawing the portfolio of senders and recipients, for understanding the importance of different corridors, and for developing strategies to harness the full potential of remittances.

For data scientists, we recommend employing a combination of the Random Forest model and the SMOTE-ENN sampling technique when working with imbalanced datasets. Using SHAP values can provide interpretability to ML models.

This is the first empirical study to examine the impact of remittance and recipient characteristics on saving behaviour. This study contributes to the literature in several ways. First, it uses the most recent transactional data, covering the period from 2017 to 2023, whereas previous studies used survey data. Second, we reveal the heterogeneous impact of different values of remittance features on recipients' saving behaviour. In contrast, previous studies have either investigated the homogeneous effect of remittances or did not interpret the outcomes of machine-learning models. Third, we developed a machine-learning model with high performance. Financial institutions can use such models to build recommender systems and to promote deposit products. In other words, the results of this study will be of interest to financial institutions seeking to convert low-profit remittances into long-term deposits and investments and to policymakers seeking to harness the full potential of remittances for economic development.

This study has some limitations. The first limitation is data-related: we obtained data from a single bank and did not compare the differences between banks. However, this limitation provides direction for future research. Another avenue for future research is exploring government policies for channelling diaspora remittances into diaspora investment. A comparative analysis of such diaspora engagement policies would help to draw lessons from the experiences of other countries that have successfully implemented similar policies.

REFERENCES

[1] Varma, Shakil. (2020). *Enhancing Financial Capability and Inclusion in Uzbekistan: A Demand-Side Assessment*. Washington, D.C.: International Finance Corporation.

[2] Demirgüç-Kunt, A. (2022). *The Global Findex Database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19.* World Bank Publications.

[3] Kubayevich, K. F. (2024). "Current Analysis and Current Issues of Ensuring the Financial Stability of the Banking System in Uzbekistan."*European Journal of Business Startups and Open Society*, pp. 169–176.

[4] Central Bank of Uzbekistan (CBU). (2023). "Statistic bulletin of Central Bank." Tashkent.

[5] Prokhorenkova, Liudmila. (2018). "CatBoost: Unbiased boosting with categorical features."*Advances in Neural Information Processing Systems.*

[6] Fonta, William. (2015). "Dynamics of remittance utilization by Nigerian households."*Progress in Development Studies*, pp. 343–357.

[7] Arthur, Emmanuel Kwesi. (2020). "Diaspora Remittances and Financial Inclusion in Kenya."*European Journal of Business and Management Research.*

[8] Prokhorova, A. (2017). "Remittances in North and Central Asian Countries: Enhancing Development Potential."*Asia-Pacific Population Journal.*

[9] Forhad, Md Abdur Rahman. (2023). "Effect of Remittance-Sending Countries' Type on Financial Development in Recipient Countries: Can the Pandemic Make a Difference?"*Journal of Risk and Financial Management.*

[10] Mehedintu, Anca. (2019). "Remittances, migration, and gross domestic product from Romania's perspective."*Sustainability.*

[11] Ajefu, Joseph Boniface. (2019). "Migrant remittances and financial inclusion among households in Nigeria."*Oxford Development Studies*, pp. 319–335.

[12] Devkota, Jeevanath. (2016). "Do return migrants use remittances for entrepreneurship in Nepal?"*Journal of Economics and Development Studies*, pp. 90–100.

[13] Acharya, Chakra. P. (2012). "The impact of remittance on poverty and inequality: A micro-simulation study for Nepal."*National Graduate Institute for Policy Studies*, pp. 1–32.

[14] Sunny, Justin. (2020). "Remittances, investment, and new emigration trends in Kerala."*Review of Development and Change*, pp. 5–29.

[15] Olivie, Iliana. (2022). *The Role of Remittances in Promoting Sustainable Development.* Directorate-General For External Policies.

[16] Pinger, Pia. (2010). "Come back or stay? Spend here or there? Return and remittances: The case of Moldova."*International Migration*, pp. 142–173.

[17] Lubambu, K. M. (2014). *The Impacts of Remittances on Developing Countries.* Directorate-General for External Policies of the Union, Policy Department.

[18] Bashir, Saqib. (2022). "Financial Inclusion Profile: Socio-Economic Determinants, Barriers, and the Reasons for Saving and Borrowing in Pakistan."*International Journal of Asian Social Science*, pp. 449–462.

[19] Morgan, Peter. (2020). "Financial literacy, financial inclusion, and savings behavior in Laos."*Journal of Asian Economics.*

[20] Quartey, Peter. (2019). "Inter-linkages between remittance and savings in Ghana."*International Journal of Social Economics*, pp. 152–166.

[21] Gatsi, J. (2020). "Effects of International and Internal Remittances on Financial Inclusion in Ghana."*Financial Markets, Institutions and Risks*, pp. 2521–1250.

[22] Jumena, Bayu Bastian. (2022). "Saving Behaviour: Factors That Affect Saving Decisions (Systematic Literature Review Approach)."*Journal Economic Resource.*

[23] Kadoya, Yoshihiko. (2020). "Financial literacy in Japan: New evidence using financial knowledge, behaviour, and attitude."*Sustainability.*

[24] Bashir, Taqadus. (2013). "Gender Differences in Saving Behavior and its Determinants (Evidence from Punjab, Pakistan)." Available at: http://ssrn.com/abstract=2253031.

[25] Fisher, Patti. (2010). "Gender differences in personal saving behaviours."*Journal of Financial Counseling and Planning.*

[26] Osei-Gyebi, Samuel. (2023). "The effect of remittance inflow on savings in Nigeria: The role of financial inclusion."*Cogent Social Sciences.*

[27] Virak, Khiev & Yuriy, Bilan. (2022). "The role of formal and informal remittances as the determinants of formal and informal financial services."*Equilibrium: Quarterly Journal of Economics and Economic Policy*, pp. 727–746.

[28] Chu, Albert B. (2018). "Mobile technology and financial inclusion." In *Handbook of Blockchain, Digital Finance, and Inclusion*, pp. 131–144.

[29] Dinarte, Lelis. (2021). "Neither by Land nor by Sea: The Rise of Electronic Remittances during COVID-19."*The World Bank.*

[30] Ismailov, Askar. (2019). "The effect of mobile money usage on borrowing, saving, and receiving remittances: Evidence from Tanzania."*University of Tsukuba.*

[31] Gupta, Lokesh. (2021). "Gharar-free remittance: Powered by blockchain." In *Fintech, Digital Currency, and the Future of Islamic Finance: Strategic, Regulatory and Adoption Issues in the Gulf Cooperation Council*, pp. 107–125.

[32] Modak, Chinmoy. (2024). "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.

[33] Özdemir, Yunus Emre, and Selim, Bayraklı.(2022). "A case study on building a cross-selling model through machine learning in the insurance industry."*Avrupa Bilim ve Teknoloji Dergisi*, pp. 364–372.

[34] Mavundla, Khulekani. (2024). "Predicting Cross-Selling Health Insurance Products Using Machine-Learning Techniques."*Journal of Computer Information Systems*, pp. 1–18.

[35] Boustani, Noureddin. (2024). "Improving the predictive accuracy of the cross-selling of consumer loans using deep learning networks."*Annals of Operations Research*, pp. 613–630.

[36] Kalkan, Ibrahim Erdem,& Cenk, Şahin. (2023). "Evaluating cross-selling opportunities with recurrent neural networks on retail marketing."*Neural Computing and Applications*, pp. 6247–6263.

[37] Mbukwa, Justine N., & Francis, Lwesya. (2023). "The adoption of mobile financial services in Tanzania: The application of logit model."*Journal of Statistics and Management Systems*, pp. 897–914.

[38] Obeid, Rami. (2022). "Early warning of bank failure in the Arab region: A logit regression approach."*Asian Journal of Economics and Empirical Research*, pp. 91–99.

[39] Anand, Mayank. (2022). "Prediction of loan behavior with machine learning models for secure banking."*Journal of Computer Science and Engineering (JCSE)*, pp. 1–13.

[40] Islam, Umar. (2022). "Detection of distributed denial of service (DDoS) attacks in IoT-based monitoring system of banking sector using machine learning models."*Sustainability*, p. 8374.

[41] Munkhdalai, Lkhagvador. (2019). "An empirical comparison of machine-learning methods on bank client credit assessments."*Sustainability*, pp. 1–23.

[42] Zaki, Ahmed Mohamed. (2024). "Predictive analytics and machine learning in direct marketing for anticipating bank term deposit subscriptions."*American Journal of Business and Operations Research*, pp. 79–88.

[43] Tékouabou, Stéphane Cédric Koumétio. (2022). "A machine learning framework towards bank telemarketing prediction."*Journal of Risk and Financial Management.*

[44] Kokorović, Jukan. (2020). "Remittances as an opportunity to increase savings and financial inclusion of youth in South East Europe."*Economic Research-Ekonomska Istraživanja*, pp. 2606–2619.

[45] Aycinena, Diego. (2010). "The impact of remittance fees on remittance flows: Evidence from a field experiment among Salvadoran migrants."*University of Michigan Report.*

[46] Banerjee, Soumyadeep. (2017). "Do Financial Remittances Build Household-Level Adaptive Capacity? A Case Study of Flood-Affected Households in India."*KNOMAD.*

[47] Azizi, SeyedSoroosh. (2020). "Impacts of remittances on financial development."*Journal of Economic Studies*, pp. 467–477.

[48] Guermond, Vincent. (2022). "Whose money? Digital remittances, mobile money, and fintech in Ghana."*Journal of Cultural Economy*, pp. 436–451.

[49] Ramraj, Santhanam. (2016). "Experimenting XGBoost algorithm for prediction and classification of different datasets."*International Journal of Control Theory and Applications*, pp. 651–662.

[50] Dogra, Varun. (2022). "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.

AUTHOR

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 AI model development, statistical modeling, and the implementation of machine learning solutions, specifically in retail loan cross-selling.