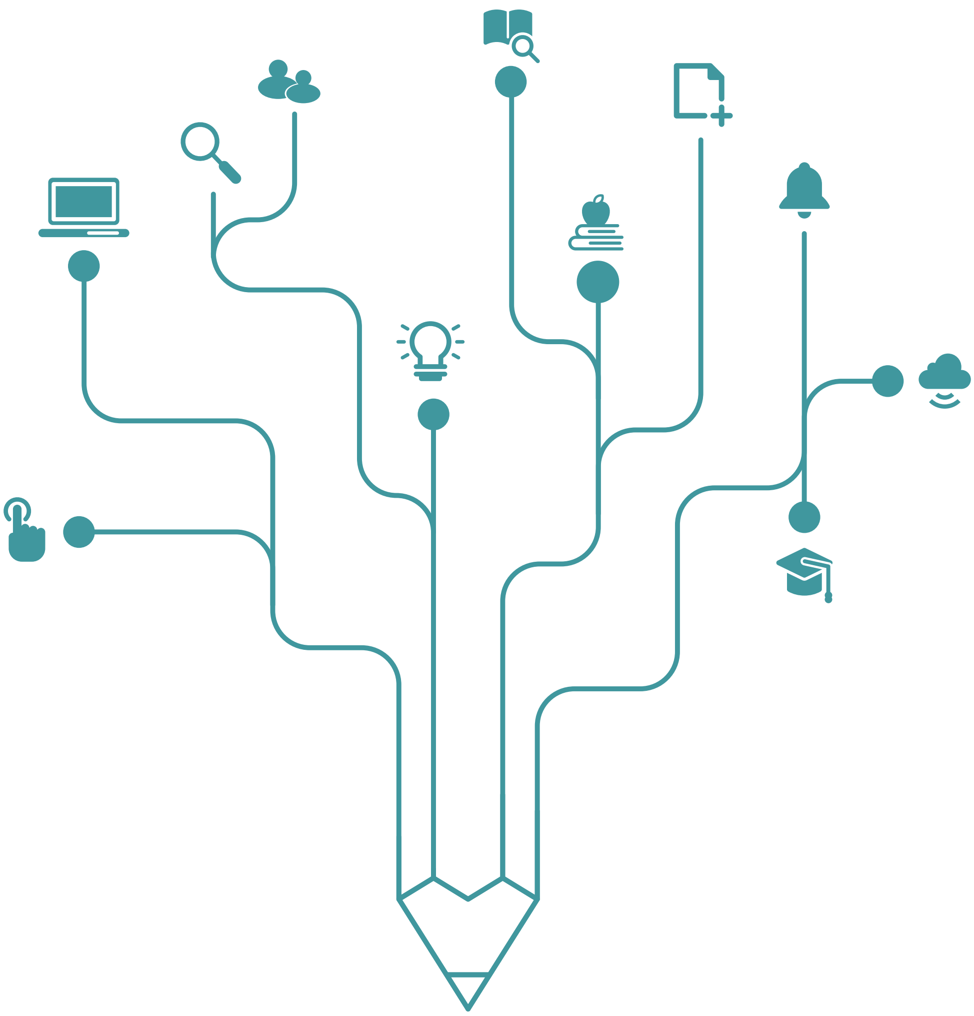
KDIS Working Paper 23-09

An explainable machine learning model for consumer credit scoring in Mexico

David Ugarte Chacon (KDI School of Public Policy and Management)
Seohyun Lee (KDI School of Public Policy and Management)
Jaehyuk Pakr (KDIS School of Public Policy and Management)



KDI SCHOOL
KDI School of Public Policy and Management

An explainable machine learning model for consumer credit scoring in Mexico

David Ugarte Chacon *[1], Seohyun Lee † [1], Jaehyuk Park ‡ [1]

[1]KDI School of Public Policy and Management, Republic of Korea

Abstract

This paper proposes an explainable machine learning model for consumer credit scoring in Mexico, an emerging economy. We develop an extreme gradient boosting (XGBoost) model using non-traditional data from the Financial Inclusion National Survey. To address the black box problem, we explore the feature importance by estimating the Shapley values that measure the average marginal contributions across all possible subsets of features. The key drivers of consumer credit defaults include the adverse economic effects due to the COVID-19 pandemic and financial attitudes and behaviors. By exploring the distributions of the Shapley values by age and income, we find the evidence of non-linearity of the feature explanations.

JEL Classification: C58, G21
Keywords: Machine learning, explainability, xAI, consumer credit scoring

*First author, elhdos2004@gmail.com.
†Corresponding author, slee6@kdis.ac.kr.
‡jpark@kdis.ac.kr.

1 Introduction

In recent years, artificial intelligence (AI) and machine learning (ML) methods have been rapidly developed and adopted in various types of credit scoring models. Instead of relying on traditional financial transactions and payment data, lenders attend to new sources of data to gauge creditworthiness of potential borrowers. The studies using alternative data sources include digital footprints obtained from mobile phone usage data (Berg et al., 2020) and proprietary data from e-commerce platforms or fintech companies (Frost et al., 2019; Gambacorta et al., 2019; Jagtiani and Lemieux, 2019). By leveraging big data with the use of AI/ML algorithms, lenders can improve speed and accuracy of their credit scoring predictions (Lessmann et al., 2015).

Yet the applications of AI/ML in credit risk management could raise potential problems. For instance, unexpected interconnectedness among financial markets and institutions could increase as it uses various data sources which seem traditionally unrelated (FSB, 2017). A widespread use of highly nonlinear black box models may lead to unintended consequences. Without a framework providing explanations about the black box, banks may be unable to justify their lending decisions if challenged. Also, AI/ML credit scoring models based on sociodemographic characteristics, such as gender, race, and ethnicity, may lead to heated public debates on the fairness of AI/ML models, causing a trust deficit (Coyle, 2020). Bartlett et al. (2022) provide the empirical evidence that, even without face-to-face interactions, profit-maximizing fintech lenders can induce discrimination by setting higher interest rates for the minority groups in the US mortgage lending market.

Against this backdrop, explainable AI/ML models for credit scoring predictions have been explored in several applications, including the US mortgage loans (Fuster et al., 2022), the UK mortgage loans (Bracke et al., 2019), and the SME loans (Bussmann et al., 2021, Guegan and Hassani, 2018).

The existing literature, however, mainly focuses on the explainable AI/ML models of advanced economies with ample micro-level financial data. A notable exception is Tantri (2021), which applies the ML algorithms in lending practices using loan applications data of India. For emerging markets and developing economies (EMDEs), the structured financial data is scarce and a sound credit reporting system has not been fully established. Loan approval is highly dependent on soft information that is not easy to be verified by third parties.[1] Under this circumstance, assessing borrowers' creditworthiness becomes more difficult, which, in turn, creates frictions in the credit market and limits access to credit.

In this paper, we attempt to fill the gap in the literature by developing an explainable ML model for predicting the default probability of consumer credit in Mexico. In particular, we employ an extreme gradient boosting model and use the Financial Inclusion National Survey of Mexico to exploit non-traditional socioeconomic and behavioral data to predict consumer credit defaults. By doing so, we demonstrate the potential usefulness of soft information that has not been systemically incorporated in the typical loan approval process.

We highlight that Mexico's financial inclusion survey data is suitable for constructing an explainable ML credit scoring model for EMDEs. Similar to other EMDEs, Mexico's financial system is quite small and the banking sector plays an important role in providing credit to the economy. Interest rate spreads tend to be wide due to high market concentration in the banking sector, especially for consumer loans, mortgages, and credit cards, and information asymmetries between lenders and borrowers (OECD, 2022). A weak correlation between borrowing rates and the default probability of consumer loans reflects Mexico's less-developed information infrastructures and credit risk management system in the banking sector. As of 2022, only 7 out of 35 commercial banks reporting to the National Commission on Banking and Securities, CNBV, conduct the default risk assessments for consumer loans and their predictions of the default probability of personal loans do not have any systemic relationships with interest rates (CNBV, 040-33R-R4, 2022 3rd bimester).

Our contributions are threefold. First, by employing an extreme gradient boosting model, we utilize a large number of non-financial variables to predict default probability while dealing with imbalanced dataset where the number of observations of the class of prediction, defaults, is extremely smaller than the other class. Second, we further apply the SHapley Additive exPlanations

[1]For the recent studies on the use of hard and soft information, see Agarwal and Hauswald, 2010 and Fisman et al., 2017, among others.

(SHAP) to analyze the key features of default predictions. The Shapley value (Lundberg and Lee, 2017) is a useful model-agnostic measure to explain prediction outcomes with the marginal contributions of features across all possible individuals and incidents. Third, by comparing our machine learning model to logistic regression models, we confirm the robustness of the XGBoost model and enhance explainability. The study also suggests that AI/ML credit scoring models for EMDEs may prove useful by deepening financial inclusion for the most vulnerable.

The rest of the paper is organized as follows. Section 2 explains data and introduces the empirical methodologies for the prediction of default probability of consumer credit. Section 3 discusses the results and Section 4 concludes.

2 Methodology

2. 1 Data

We use data from the Financial Inclusion National Survey (ENIF, Encuesta Nacional de Inclusion Financiera) collected by National Institute of Statistics and Geography (INEGI) of Mexico (INEGI, 2021). The survey has been conducted every three years, starting from 2012. We use the most recent wave of the survey conducted during the period from June 28 to August 13, 2021, on individuals above 18 years old. The survey is representative of the 90.3 million adult population in Mexico at the time, with 13,352 observations. The sample was stratified by six geographical regions, and by rural and urban locations. The survey contains comprehensive information about the socioeconomic characteristics of households (gender, location, education, income, and employment). It also has questions about the houses they live in, savings and credit from both formal and informal sources, payments, insurance, retirement savings, and the use of financial intermediaries. One of the key features included in the survey is the information about financial attitudes and behaviors of the respondents (e.g. habits of recording financial transactions, spending/saving behaviors, attitudes towards money) and financial literacy. On top of the regular questions of the survey series, the 2021 survey further asks about financial vulnerability and resilience during and after the COVID-19 pandemic. The examples include, due to COVID-19, (i) financially affected;

(ii) had reduction in earnings; (iii) had health or funeral expanses; (iv) relied on other funding sources to cope with economic emergencies.

The data covers broad ranges of consumer loans–credit cards, payroll, personal, auto, retail credit cards, cooperatives, mortgage and fintech loans. Due to the design of the survey questionnaire asking only whether or not a person has loans, the amounts of outstanding loans are not available, nor their respective interest rates.[2] The total number of households with at least one type of consumer credit is 4,552, 34% of total number of survey respondents. The default of consumer credit is defined as being delinquent for at least one type of consumer loan at the time of the survey. The proportion of households with defaulted consumer loans within total number of households with at least one type of consumer loans in our dataset is 31%. [3]

For our analysis, we have adopted a meticulous data management strategy by partitioning our dataset into three subsets: the training set, the validation set, and the test set.[4]

2. 2 Logistic regression model

We employ the logistic regression as a benchmark model for the evaluation of the performance of ML credit scoring model.

$$ln\left(\frac{p_i}{1-p_i}\right) = \mathbf{x_i'}\beta \tag{1}$$

where p_i is the probability of default for individual i, $\mathbf{x}'$ is a vector of the values of explanatory variables. Using a large number of explanatory variables (or features in machine learning terminology) in our dataset may cause statistical issues, such as noise to the coefficient calculation and multicollinearity. Hence, we construct two logistic models only with the top 20 variables that have the highest explanatory power, based on their information gains and Shapley values, respectively[5].

[2] See Appendix A.1. Figures A1-A2 for the aggregate data at the national level on default rate and the balance of consumer credit.

[3] See Appendix A.1. Figures A3-A4 for the number of households with consumer credit and defaults by loan type.

[4] See Appendix A.1. for more details.

[5] See Appendix A.2. for the details of estimation process and the estimation results of the logistic regression model.

2. 3 Extreme Gradient Boosting model

We employ the Extreme Gradient Boosting (XGBoost) classifier, a technique that has demonstrated significant versatility, ranging from patient mortality predictions to the assessment of individual credit scores (Yan et al. (2020); Liu et al. (2021); Li et al. (2020)). XGBoost, a potent machine learning algorithm, is rooted in gradient-boosting decision trees, exhibiting excellent scalability in various circumstances, and a rapid learning trajectory (Chen and Guestrin (2016)). This classifier excels in its capacity to effectively build boosted trees that can run in parallel — either regression or classification trees — optimizing the objective function value throughout the trees.

One standout advantage of the XGBoost classifier lies in its ability to render an intricate decision-making process more interpretable. While unraveling the intricacies of black-box modeling strategies remains challenging, XGBoost's recursive tree-based decision system enables the identification of the importance of each individual feature by tracking its cumulative use in each decision step. The resulting metric quantifies the relative importance of each feature, an invaluable asset when estimating the features that significantly enhance the model's outcomes, particularly when tied to salient valuation parameters.

2. 4 Explaining model predictions using Shapley

We apply the SHapley Additive exPlanations (SHAP or Shapley) (Lundberg and Lee (2017)) to explain the determinants of our machine learning predictions of default probability. SHAP is a method grounded in game theory used to explain the output of machine learning models. To explain the prediction of an instance by computing the contribution of each feature to the prediction, SHAP assigns each feature an importance value for a particular prediction.

The SHAP value for each feature in our XGBoost algrorithm is calculated through the following process. First, from our trained model on the dataset, the base value, or average prediction of the model on the training dataset, is identified, which serves as the starting point for all subsequent SHAP calculations.

Next, each individual prediction made by the model is 'explained' by attributing portions of the prediction to each feature used in the model. This is done by considering all possible subsets of features and evaluating how much the model's prediction changes when a particular feature is added or removed from a subset. It is important to note that these combinations take into account the order in which features are added, reflecting the reality of complex models where the impact of a feature often depends on the presence of other features. This process is repeated across all features and all instances to obtain a SHAP value for each feature, for each instance. The result SHAP values provide insights into how each feature contributes to the predictions, either by increasing or decreasing the predicted value, relative to the base value. The greater the magnitude of a feature's SHAP value, the greater the impact of that feature on the model's output.

3 Results

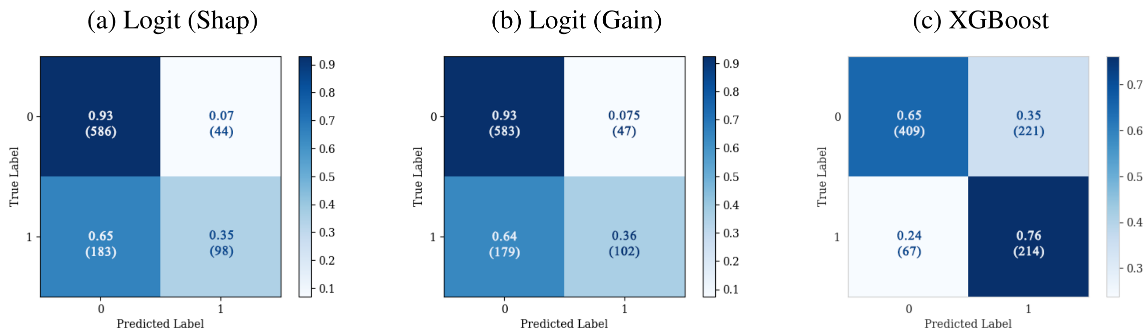
3. 1 Predictions

To evaluate the performance of the models, we first illustrate the true and predicted default classification for the logistic regressions and the XGBoost model in Figure 1. The two logistic regression models have relatively small false positives (FP) – consumer loans predicted as default that actually did not – of 7%. However, they have greater false negatives (FN) – consumer credit predicted not to default but actually does – of 65%. On the contrary, the XGBoost model performs better in predicting defaults with a relatively small proportion of wrong predictions in defaults with a lower FN rate (24%), albeit the performance in predicting non-default worsens with the FP rate of 35%. The confusion matrix confirms that, the XGBoost predictions tend to outperform linear logistic regression models in case of our imbalanced dataset in which the number of the observations of the class of prediction (defaults) is extremely fewer than the other class (non-defaults).

3. 2 Performance measures

Figure 2 shows the Receiver Operating Characteristics (ROC) curve, plotting the TP rate against the FP rate at each threshold. A random classifier would be on the 45-degree line while a better

FIGURE 1 Confusion Matrix

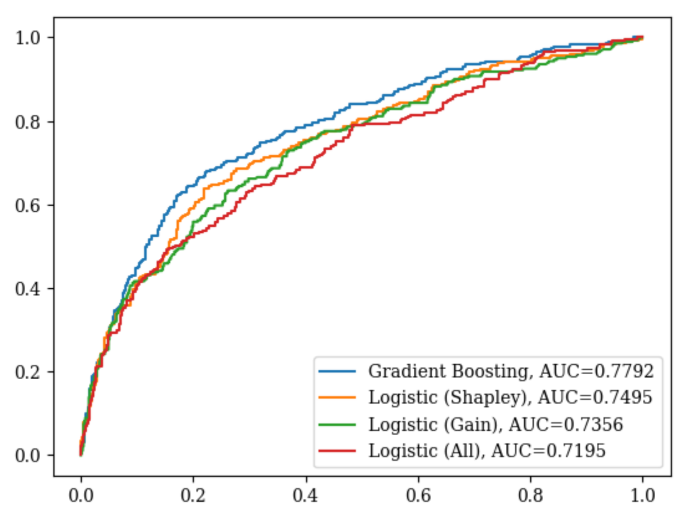


Notes: The confusion matrix summarizes the predicted classification of credit defaults against true classification of defaults in the sample. If the label is 1 (0), it means default (not default). Darker color of each of the four possible outputs indicates higher probability.

classifier would be above the random classifier line, having higher TP rates at any given level of the FP rate. The ROC curve for the XGBoost model is above the other three logistic regression models, indicating that the machine learning algorithm appears to better predict the defaults using our survey dataset.

A summary indicator of the ROC curve is the Area Under the Curve (AUC) measuring how far is the ROC curve from the random classifier. Consistent with the ROC curves, the AUC for the XGBoost model (78%) is larger compared to those of the logistic regression models, ranging from 72% to 75%.

FIGURE 2 The ROC curves



Notes: The y-axis represents the True Positive rate and the x-axis represents the False Positive rate.

TABLE 1 Model performance measures

Model	AUC	Brier Score	Log Loss	Precision	Recall	F1	Partial Gini
XGBoost	0.7791	0.2047	0.5984	0.4919	0.7615	0.5977	0.7328
Logistic (Shapley)	0.7495	0.1747	0.5294	0.6901	0.3487	0.4633	0.7074
Logistic (Gain)	0.7356	0.1775	0.5358	0.6845	0.3629	0.4744	0.6929
Logistic (All)	0.7195	0.1852	0.5646	0.6310	0.4199	0.5042	0.6801

Notes: AUC is the area under the ROC curve. Brier score is the mean squared error of the differences between probability forecasts and actual outcomes. Log loss measures how close the classifier output is to the correct output ($= \sum -y log(p) - (1-y) log(1-p)$, where y is actual output, p is prediction). Precision is computed as the ratio of $TP/(TP+FP)$. Recall is computed as the ratio of $TP/(TP+FN)$. F1 indicates the harmonic mean of precision and recall scores. Partial Gini is the AUC score for the ROC curve only up to a prediction score of 0.4 (Lessmann et al. (2015)).

Table 1 reports other key performance measures. Based on broad metrics like Brier score and log loss, the scores are higher for the XGBoost model, suggesting that the XGBoost predictions are less accurate compared to the logistic regression predictions. However, these metrics do not directly take account for the imbalanced number of events among the classes in a dataset. In our case of default predictions where one class (non-defaults) outnumbers the other (defaults), precision and recall would provide more useful information. For default prediction, FNs (predicted as non-defaults but actually default) are more costly than FPs (predicted as defaults but actually not) and thus recall is more relevant than precision. Given potential trade-offs between precision and recall, it would be desirable to achieve a high recall while sacrificing less in precision. Relative to the logistic regressions, our XGBoost model has higher recall and lower precision. F1 score and partial Gini also indicate better performance of the XGBoost model.

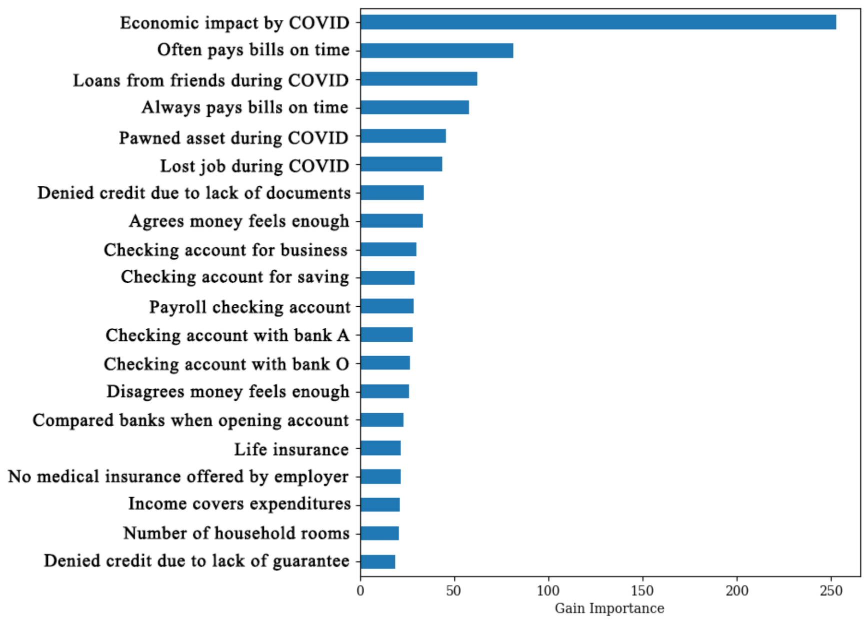
3. 3 Explanations of predictions

We first identify the relative feature importance of the XGBoost predictions based on information gain [6]. Figure 3 shows top 20 most important features explaining the defaults of consumer credit. The probability of default can be largely explained by whether one had economic impacts due to the COVID-19 pandemic. Among top 20 features, the COVID-19 related factors are four in total, including borrowed from friends, pawned assets, lost job during the pandemic. Other variables

[6] Information gain measures how much information a feature adds to the model. For a given split in a decision tree, information gain is the reduction in entropy or Gini impurity that results from the split. The larger the decrease in impurity (higher information gain), the more important the feature is considered to be. This is calculated for each feature over all trees in the model, and summed to provide a measure of the feature's overall importance.

with high feature importance include financial attitudes and behaviors, having checking accounts with specific banks, having insurance, and reasons for credit card denial.

FIGURE 3 Gain importance: the XGBoost model



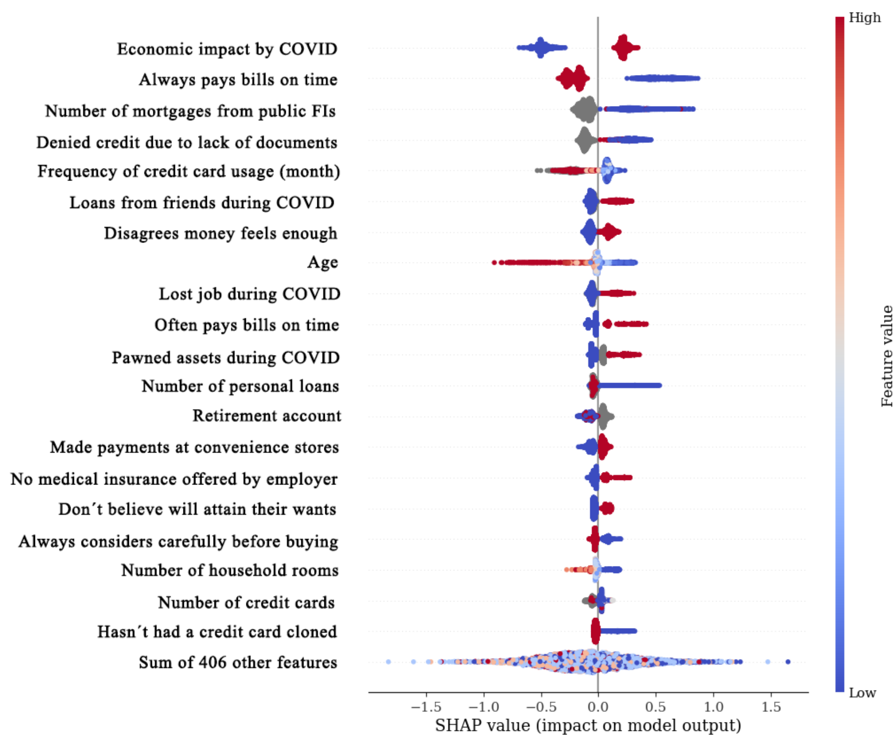
Notes:

Next, we move to the Shapley values of the XGboost model (Figure 4). Each point of a SHAP summary plot represents a shapley value for a feature and an instance (a data point). Similar to the information gain results, the four COVID-19 pandemic related variables are included in the list of the top 20 most important features of the XGBoost model based on the Shapley value. For a person who responded that the COVID-19 adversely affected, the probability of default in any types of consumer loans increases. The financial behavior is one of the important features in default predictions of the XGBoost model. Those who responded that they *always* pay bills on time have lower default probability.

Figures 5-6 report the Shapley values from the logistic regression model with 20 features selected based on the XGBoost Shapley values and information gain. One of the most important feature is the behavioral aspect, *always* pays bills on time, and it reduces the default probability. If a person was denied credit card due to lack of documentation (among other reasons, including not able to prove income, no credit history, lack of guarantee), he is less likely to default on loans.

The COVID-19 pandemic related features are also important for the logistic regression and the directions of their impacts are consistent with the XGBoost model results.

FIGURE 4 SHAP summary plot: the XGBoost model

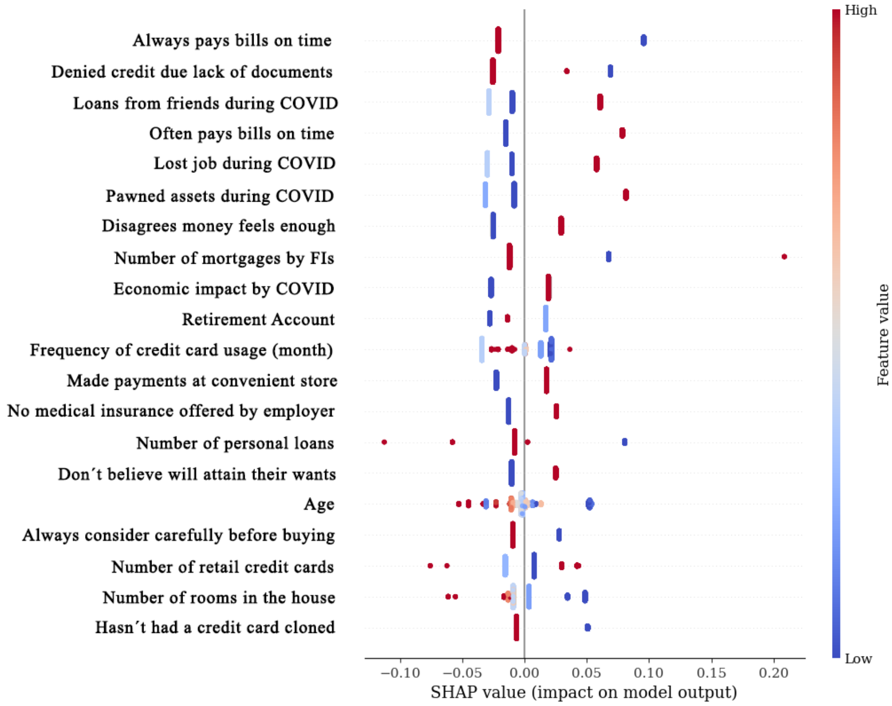


Notes: Red (blue) indicates higher (lower) value of the feature. Grey indicates missing values.

Figure 7 shows the distribution of the Shapley values by features. Panel (a) reconfirms that, on average, the default probability decreases with age. For those who are in their retirement age (older than 65 years old), the negative correlation between the Shapley value and age becomes larger and the Shapley values are much more dispersed. Panel (b) reports the SHAP dependence plot by income level. Overall, consumer loans borrowed by high income households are more likely to default. There is an exception – households with income between 10,000 and 15,000 in USD are more likely to have lower default rates, compared to those who earn slightly less. However, the variance of the Shapley values seems quite large for higher income levels above USD20,000.

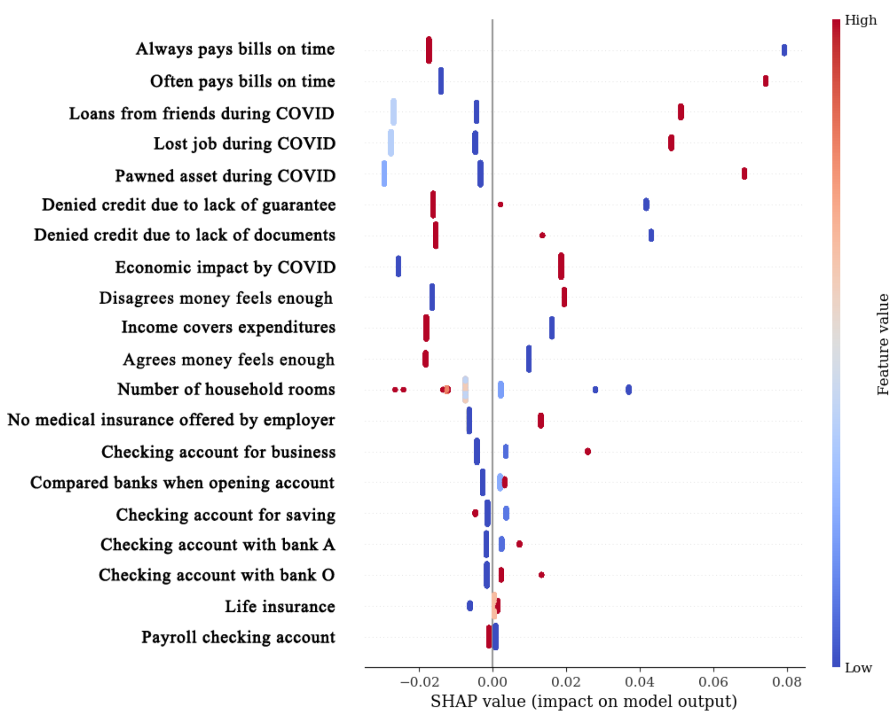
Lastly, waterfall plots can illustrate the feature importance of an individual default prediction as compared to the average prediction. Individual A (Figure 7) has responded that he was *not* economically impacted by the COVID-19 pandemic, which reduces the probability of default.

FIGURE 5 SAHP summary plot: the logistic regression model (Shap)



Notes: The features included in the model is the top 20 features based on the Shapley values of the XGBoosting model. Red (blue) indicates higher (lower) value of the feature. Grey indicates missing values.

FIGURE 6 SAHP summary plot: the logistic regression model (Gain)

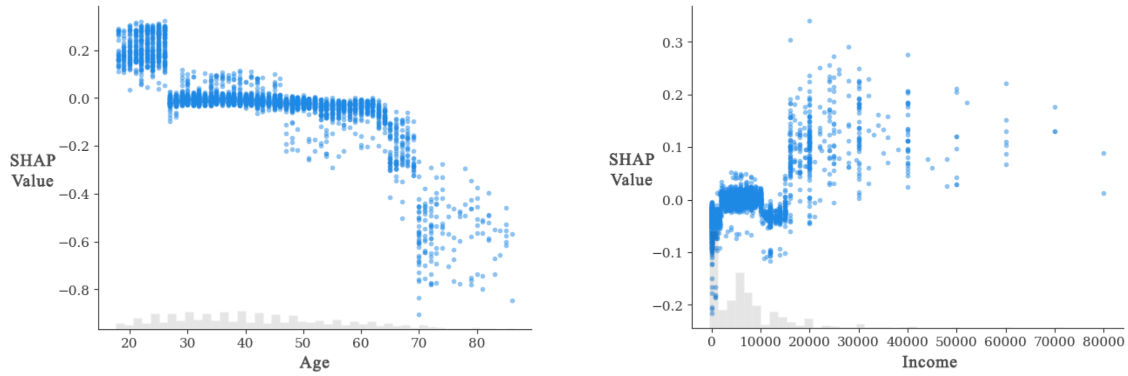


Notes: The features included in the model is the top 20 features based on the information gain of the XGBoosting model. Red (blue) indicates higher (lower) value of the feature. Grey indicates missing values.

FIGURE 7 The distribution of SHAP value

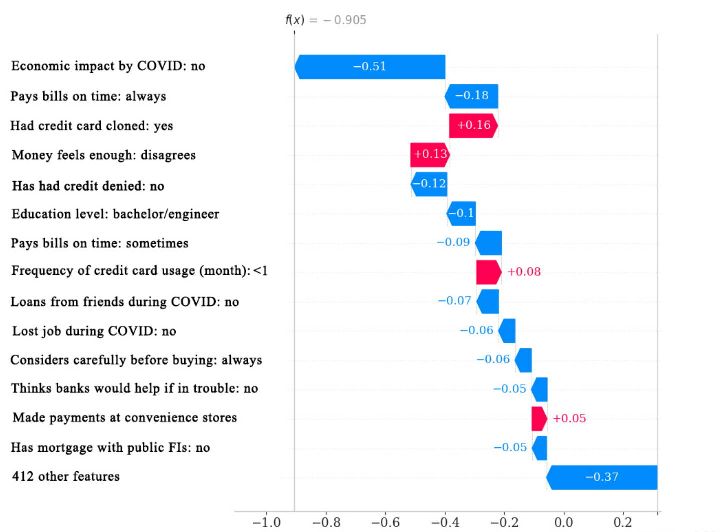
(a) By age

(b) By income



Individual A's financial attitudes and behavioral aspects also affect the default prediction. For instance, he doesn't feel money is enough and this response increases the likelihood of default in the XGBoost prediction.

FIGURE 8 Feature influences for Individual A



Notes: $E(f(x))$ indicates the baseline (average) default probability. Red (blue) denotes increases (decresase) in the default probability.

4 Conclusions

Our study aims to provide an empirical evidence of benefits of explainable machine learning in predicting consumer credit defaults using non traditional survey data of Mexico. We employ the XGBoost model to predict consumer credit defaults. In our default prediction application where the data is imbalanced, the XGBoost model outperforms the logistic regression model. For unboxing the black box, we apply the SHAP and for our 2022 survey data, the adverse economic conditions due to the COVID-19 pandemic are important features in predicting defaults.

Our research has important policy implications for the financial market applications of AI/ML. The regulators could consider the requirement of explainability for the AI/ML model that are inherently complex and difficult to interpret. For banks legally obliged to explain the credit decisions, the Shapley can provide useful information about the important features of the decision at the individual level.

References

[1] Agarwal, S. and Hauswald, R. (2010). Distance and private information in lending. *The Review of Financial Studies*, 23(7):2757–2788.

[2] Bartlett, R., Morse, A., Stanton, R., and Wallace, N. (2022). Consumer-lending discrimination in the fintech era. *Journal of Financial Economics*, 143(1):30–56.

[3] Berg, T., Burg, V., Gombović, A., and Puri, M. (2020). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies*, 33(7):2845–2897.

[4] Bracke, P., Datta, A., Jung, C., and Sen, S. (2019). Machine learning explainability in finance: an application to default risk analysis. Working paper.

[5] Brauninger, J. (1980). A quasi-newton method with cholesky factorization. *Computing*, 25:155–162.

[6] Bussmann, N., Giudici, P., Marinelli, D., and Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics*, 57:203–216.

[7] Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

[8] Coyle, D. (2020). The tensions between explainable ai and good public policy. *Brookings TechStream, September*, 15.

[9] Fisman, R., Paravisini, D., and Vig, V. (2017). Cultural proximity and loan outcomes. *American Economic Review*, 107(2):457–492.

[10] Frost, J., Gambacorta, L., Huang, Y., Shin, H. S., and Zbinden, P. (2019). Bigtech and the changing structure of financial intermediation. *Economic Policy*, 34(100):761–799.

[11] FSB (2017). Artificial intelligence and machine learning in financial services: market developments and financial stability implication. Technical report, Financial Stability Board.

[12] Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., and Walther, A. (2022). Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*, 77(1):5–47.

[13] Gambacorta, L., Huang, Y., Qiu, H., and Wang, J. (2019). How do machine learning and non-traditional data affect credit scoring? new evidence from a chinese fintech firm. Available at SSRN: `https://ssrn.com/abstract=3506945`.

[14] Guegan, D. and Hassani, B. (2018). Regulatory learning: How to supervise machine learning models? an application to credit scoring. *The Journal of Finance and Data Science*, 4(3):157–171.

[15] INEGI (2021). Encuesta nacional de inclusion financiera. Technical report, Instituto Nacional de Geografía e Historia.

[16] Jagtiani, J. and Lemieux, C. (2019). The roles of alternative data and machine learning in fintech lending: evidence from the lendingclub consumer platform. *Financial Management*, 48(4):1009–1029.

[17] Lessmann, S., Baesens, B., Seow, H.-V., and Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1):124–136.

[18] Li, H., Cao, Y., Li, S., Zhao, J., and Sun, Y. (2020). Xgboost model and its application to personal credit evaluation. *IEEE Intelligent Systems*, 35(3):52–61.

[19] Liu, J., Wu, J., Liu, S., Li, M., Hu, K., and Li, K. (2021). Predicting mortality of patients with acute kidney injury in the icu using xgboost model. *Plos one*, 16(2):e0246306.

[20] Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

[21] OECD (2022). Oecd economic surveys: Mexico 2022. *OECD Publishing, Paris*.

[22] Tantri, P. (2021). Fintech for the poor: Financial intermediation without discrimination. *Review of Finance*, 25(2):561–593.

[23] Yan, L., Zhang, H.-T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., Sun, C., Tang, X., Jing, L., Zhang, M., et al. (2020). An interpretable mortality prediction model for covid-19 patients. *Nature machine intelligence*, 2(5):283–288.

Appendix

A.1. Data preparation

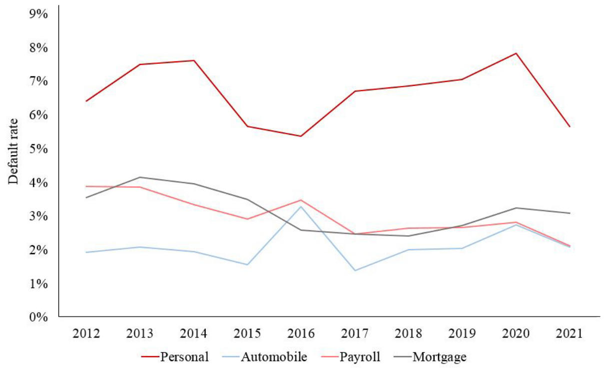
The initial stage of our model building process is grounded on what we term the 'training set'. By using a method of random sampling, we designate approximately 75% of our total dataset as the training set. This portion forms the bedrock for the early phases of our model development, where preliminary parameters are estimated, and features are engineered. However, to ensure that our model does not become overly adjusted or overfitted to this training data, we introduce a further level of refinement and evaluation with the validation set.

The validation set, which comprises 25% of the original training set, plays a critical role in the model optimization process. This set provides an opportunity to assess and refine the model's parameters and evaluate the performance using data unseen during the initial model creation stage. This method of iterative evaluation and refinement helps us establish a robust model without overfitting the data. Once this iterative process of refinement and validation is completed, we finalize our model parameters. Importantly, we make no further changes or optimizations to the model parameters or feature engineering process. This approach helps us maintain the integrity of our model, ensuring that it remains unbiased towards our specific sample data.

The last, yet significant, stage involves the test set. This dataset comprises the remaining 25% of our total data, which has been kept separate and untouched during the entire model development process. It is used to evaluate our finalized model's ability to predict outcomes on unseen data, thus offering an objective measure of its performance and generalizability.

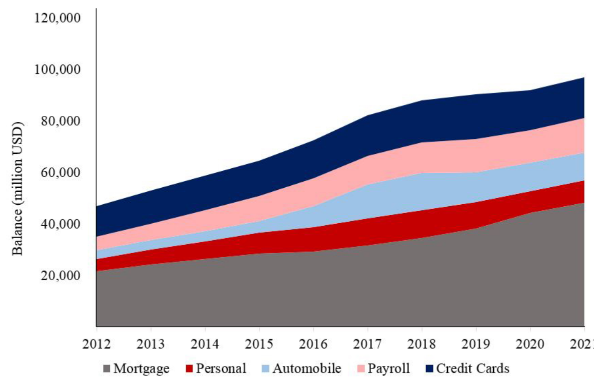
In our data preprocessing, we used a method called one-hot encoding to handle categorical features in our dataset. Categorical features refer to variables that have a finite number of distinct categories but lack any inherent order, such as different answers of a questionnaire, which many machine learning algorithms are unable to work with in their raw form. To resolve this, we apply one-hot encoding to transform each category of these variables into a new binary feature (0 or 1), allowing for a more compatible representation of these categorical data within our machine learning models. Each category is thus represented as a unique binary vector in the transformed dataset, ensuring that our models can effectively incorporate this categorical information. After data pre-processing, we obtain 426 features in total for the XGBoost model predictions.

A1. Default rate by credit type



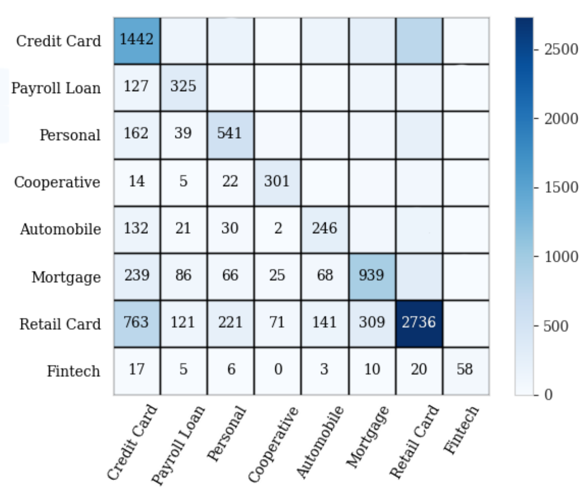
Source: National Commission on Banking and Securities (CNBV), forms 04013aR2B, 04030aR3, 04031aR3, and 04033aR3, which are built with the information reported by the regulated commercial FIs to the Commission.

A2. The balance of consumer credit by type



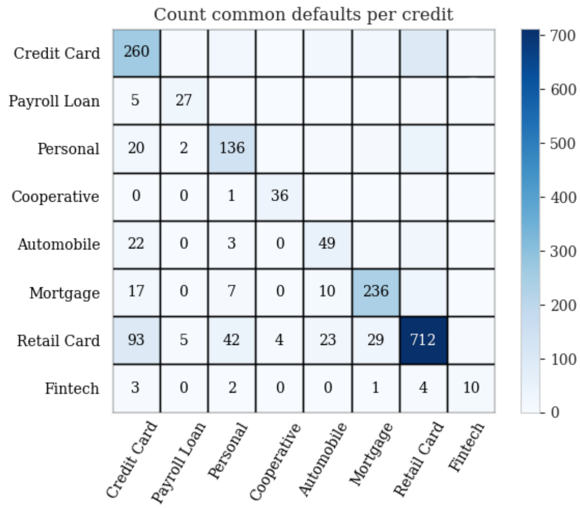
Source: National Commission on Banking and Securities (CNBV), forms 04013aR2B, 04033aR1, 04030aR1, 04031aR3, and 04012dR2.

A3. The number households with consumer credit by type



Notes: Coopertives include credit unions and microfinance entities.

A4. The number of consumer loan defaults by type



Notes: Coopertives include credit unions and microfinance entities.

A.2. Logistic regression results

The logistic regression model undergoes the standard pre-processing normalization of missing values by mean filling. Also, the Newton-Cholesky optimization algorithm (5) is employed to iteratively update the coefficients using both the first- and second-order derivatives of the log likelihood function. L2 regularization is employed to reduce the probability of overfitting by adding a squared coefficient term to the logistic loss function, which overly penalizes large coefficient weights, resulting in a simpler and better generalizing model.

A1. Summary Statistics of features

Column	Count	Mean	Std	Min	25%	50%	75%	Max
Economic impact by COVID	4552	0.583919	0.492961	0	0.0	1.0	1.0	1
Often pays bills on time	4552	0.218585	0.413332	0	0.0	0.0	0.0	1
Loans from friends during COVID	2658	0.381866	0.485935	0.0	NaN	NaN	NaN	1.0
Always pays bills on time	4552	0.773726	0.418465	0	1.0	1.0	1.0	1
Pawned asset during COVID	2658	0.232882	0.422747	0.0	NaN	NaN	NaN	1.0
Lost job during COVID	2658	0.374342	0.484044	0.0	NaN	NaN	NaN	1.0
Denied credit card due to lack of documents	1206	0.071310	0.257449	0.0	NaN	NaN	NaN	1.0
Agrees money feels enough	4552	0.352812	0.477897	0	0.0	0.0	1.0	1
Checking account for business	3305	0.076248	0.265435	0.0	NaN	NaN	NaN	1.0
Checking account for saving	3305	0.106505	0.308530	0.0	NaN	NaN	NaN	1.0
Payroll checking account	4552	0.445299	0.497053	0	0.0	0.0	1.0	1
Checking account with Bank A	3185	0.078493	0.268988	0.0	NaN	NaN	NaN	1.0
Checking account with Bank O	3185	0.031711	0.175258	0.0	NaN	NaN	NaN	1.0
Disagrees money feels enough	4552	0.396529	0.489230	0	0.0	0.0	1.0	1
Compared banks when opening account	3185	0.232339	0.422390	0.0	NaN	NaN	NaN	1.0
Life insurance	1936	0.664773	0.472192	0.0	NaN	NaN	NaN	1.0
No medical insurance offered by employer	4552	0.323374	0.467816	0	0.0	0.0	1.0	1
Income covers expenditures	4552	0.549209	0.497627	0	0.0	1.0	1.0	1
Number of household rooms	4552	3.998023	1.570505	1	3.0	4.0	5.0	13
Denied credit card due to lack of guarantee	1206	0.045605	0.208714	0.0	NaN	NaN	NaN	1.0

Notes: The included features are the top 20 highest values based on information gain of the XGBoost model. The full list of features and summary statistics are available upon request.

A2. Coefficients of the logistic regressions (I)

	coefficient	std. err.	Z	p-value	[0.025]	[0.975]
Economic impact by COVID	0.9769	0.0882	11.0776	0.0000	0.8041	1.1498
Number of mortgages from public FIs	0.6301	0.4587	1.3735	0.1696	-0.2690	1.5292
Pawned assets during COVID	0.4694	0.1155	4.0624	0.0000	0.2429	0.6958
Disagrees money feels enough	0.3397	0.0872	3.8961	0.0001	0.1688	0.5105
Made payments at convenience stores	0.3319	0.0820	4.0475	0.0001	0.1712	0.4926
Loans from friends during COVID	0.3193	0.1006	3.1738	0.0015	0.1221	0.5165
Lost job during COVID	0.3171	0.1000	3.1700	0.0015	0.1210	0.5132
Don´t believe will attain their wants	0.2036	0.0925	2.2004	0.0278	0.0222	0.3849
Number of credit cards	0.1463	0.0805	1.8177	0.0691	-0.0115	0.3040
Retirement account	0.1386	0.1601	0.8658	0.3866	-0.1752	0.4524
No medical insurance offered by employer	0.1168	0.0851	1.3716	0.1702	-0.0501	0.2836
Age	-0.0116	0.0030	-3.8306	0.0001	-0.0175	-0.0057
Frequency of credit card usage (month)	-0.0171	0.0134	-1.2778	0.2013	-0.0434	0.0091
Number of household rooms	-0.0593	0.0262	-2.2665	0.0234	-0.1106	-0.0080
Always considers carefully before buying	-0.2163	0.0883	-2.4497	0.0143	-0.3894	-0.0432
Denied credit due to lack of documents	-0.2850	0.3163	-0.9012	0.3675	-0.9050	0.3349
Often pays bills on time	-0.4323	0.4009	-1.0782	0.2810	-1.2181	0.3536
Number of personal loans	-0.4479	0.2618	-1.7109	0.0871	-0.9611	0.0652
Hasn´t had a credit card cloned	-0.5493	0.1212	-4.5295	0.0000	-0.7870	-0.3116
Always pays bills on time	-1.4741	0.3968	-3.7148	0.0002	-2.2518	-0.6963

Notes: The variables included in the logistic regression are the top 20 highest Shapley values in the baseline XGB model.

A3. Coefficients of the logistic regressions (II)

	coefficient	std. err.	Z	p-value	[0.025]	[0.975]
Economic impact by COVID	0.9807	0.0887	11.0528	0.0000	0.8068	1.1546
Pawned asset during COVID	0.4580	0.1146	3.9962	0.0001	0.2334	0.6826
Loans from friends during COVID	0.3352	0.1007	3.3280	0.0009	0.1378	0.5326
Lost job during COVID	0.3350	0.0997	3.3590	0.0008	0.1395	0.5305
Disagrees money feels enough	0.2378	0.0994	2.3926	0.0167	0.0430	0.4326
Checking account for business	0.2147	0.1726	1.2444	0.2134	-0.1235	0.5530
Checking account with Bank O	0.2118	0.2624	0.8070	0.4197	-0.3026	0.7261
Compared banks when opening account	0.1689	0.1134	1.4902	0.1362	-0.0533	0.3911
No medical insurance offered by employer	0.1211	0.0926	1.3080	0.1909	-0.0604	0.3026
Life insurance	0.0717	0.1295	0.5536	0.5799	-0.1821	0.3255
Payroll checking account	0.0541	0.0895	0.6038	0.5460	-0.1214	0.2296
Checking account with Bank A	0.0344	0.1717	0.2004	0.8412	-0.3022	0.3710
Checking account for saving	-0.0243	0.1590	-0.1528	0.8785	-0.3360	0.2874
Number of household rooms	-0.0706	0.0253	-2.7951	0.0052	-0.1202	-0.0211
Income covers expenditures	-0.1513	0.0889	-1.7016	0.0888	-0.3256	0.0230
Agrees money feels enough	-0.1646	0.1077	-1.5284	0.1264	-0.3756	0.0465
Denied credit card due to lack of guarantee	-0.2047	0.3494	-0.5859	0.5579	-0.8896	0.4802
Denied credit card due to lack of documents	-0.2508	0.3160	-0.7936	0.4275	-0.8703	0.3687
Often pays bills on time	-0.4303	0.1961	-2.1949	0.0282	-0.8146	-0.0460
Always pays bills on time	-1.5059	0.1887	-7.9811	0.0000	-1.8758	-1.1361

Notes: The variables included in the logistic regression are the top 20 highest information gain values in the baseline XGB model.