Digital financial inclusion in Latin America: An application of classification models

https://doi.org/10.18041/1900-3803/entramado.1.12332

Sofia Orazi
Instituto de Investigaciones Económicas y Sociales del Sur (IIESS-UNS-CONICET), Departamento de Economía, Universidad Nacional del Sur, Argentina. Escuela de Ciencias Empresariales, Universidad Católica del Norte, Coquimbo, Chile

Lisana B. Martinez
Instituto de Investigaciones Económicas y Sociales del Sur (IIESS-UNS-CONICET) y Departamento de Economía, Universidad Nacional del Sur, Argentina

Hernan Pedro Vigier
Instituto de Investigaciones Económicas y Sociales del Sur (IIESS-UNS-CONICET) y Departamento de Economía, Universidad Nacional del Sur, Argentina

Abstract

Digital financial inclusion is crucial for economic and social development in Latin America, where access to basic financial services is limited and the informality in the economy has ever been reluctant to decrease. This paper uses a novel methodology for classified individuals according to whether they have a mobile account (also known as e-wallets) based on their socio-economic characteristics, the holding of other instruments, and the country of origin among four Latin American nations (Argentina, Brazil, Colombia, and Peru) in 2021. Mobile accounts foster inclusive growth and reduce transaction costs and geographical constraints and were boosted in the post pandemic era. The objective is to identify the most relevant attributes of mobile account ownership, in order to improve digital financial inclusion. For the classification, the results highlight age and income as relevant but also owning a debit card, accessing the internet at home, and having saved in the last year as significant factors regardless the country of origin (meaning that they are very alike) or typically relevant attributes for traditional finance like education or gender.

Keywords

Registration

Inclusión financiera digital en América Latina: una aplicación de modelos de clasificación

Resumen

La inclusión financiera digital es crucial para el desarrollo económico y social en América Latina, donde el acceso a los servicios financieros básicos es limitado y la informalidad en la economía ha mostrado una persistente resistencia a disminuir. Este artículo utiliza una metodología novedosa para clasificar a las personas según tengan o no una cuenta móvil (también conocida como billetera electrónica) en función de sus características socioeconómicas, la tenencia de otros instrumentos y el país de origen entre cuatro naciones latinoamericanas (Argentina, Brasil, Colombia y Perú) en 2021. Las cuentas móviles fomentan el crecimiento inclusivo y reducen los costos de transacción y las restricciones geográficas. El objetivo es identificar los atributos más relevantes de la titularidad de una cuenta móvil, con el fin de mejorar la inclusión financiera digital. Para la clasificación, los resultados destacan la edad y los ingresos como factores relevantes, pero también la posesión de una tarjeta de débito, el acceso a internet en el hogar y el haber ahorrado en el último año como factores significativos independientemente del país de origen (lo que significa que son muy similares) o atributos típicamente relevantes para las finanzas tradicionales como la educación o el género.

Palabras clave

License

How to cite this article

Introduction

Financial inclusion represents one of the main ways to increase financial resilience, limit household vulnerability, and boost sustainable and inclusive growth at global and local levels. In this sense, its promotion becomes more relevant for policy-makers, given that many investigations demonstrated the direct link between financial exclusion and poverty in developing nations (Kling, Pesqué-Cela, Tian, & Luo, 2022; Levine, 2021; Sharma & Kukreja, 2013).

During extreme situations like pandemics, the world acknowledges the importance of having a widespread financial system capable of reaching the most vulnerable populations to provide essential information and economic or medical assistance. However, developing countries continue to face persistent gaps in gender, income, and location (urban or rural areas) regarding access to and use of digital financial services (Orazi, Martinez & Vigier, 2023; Tay, Tai & Tan, 2022).

Cash in an economy leads to higher transaction costs and insecurity problems and sustains the existence of informal markets. A transition to a widespread financial system, using available technology, would lower costs, stimulate economic activity, and formalize the economy (Bastante, 2020). But the use of technology in financial and non-financial domains presents its own challenges, mainly related to age, educational level, familiarity with other technologies, and confidence in handling personal data (Piotrowska, 2024).

Latin America is currently experiencing high mobile phone penetration and accelerated growth in electronic transactions. This represents a complement or alternative to traditional financial instruments, which have shown little permeability to certain sectors, especially in rural areas due to the lack of access points, and among disadvantaged populations such as women, younger individuals, those with lower educational levels, and those engaged in informal jobs. Studying digital financial inclusion allows to understand the benefits of digital technologies in facilitating access to financial services for segments of the population that have historically been excluded from the traditional financial system.

We analysed mobile account ownership and other socio-economic variables of the population using World Bank data collected in the 2021 Global Findex survey, when the effect of the pandemic on the economy was at its peak. Various models were used to classify survey participants in four Latin American countries, based on their socio-economic characteristics, ownership of other instruments, and country of origin. The selected countries (Argentina, Brazil, Peru, and Colombia) share a similar income level (GDP per capita) and cultural and historical traits and are the largest in terms of geographic extension. The latter is particularly relevant for promoting digital financial inclusion, aiming to overcome the lack of physical access points to traditional financial services in large geographic areas.

The methodology applied consists of diverse classification models (two trees and a random forest) that use the attributes of individuals to distinguish those with a mobile account. In this sense, we sought to identify the characteristics contributing the most to the classification. Will the classic attributes observed in the literature on educational level, income or age prevail? Or will it be more important for the model to differentiate individuals by country of origin? This methodology as being machine learning, do not get biased by the researcher's background.

This work poses several challenges, such as detecting the model that best fits the data for prediction while maintaining explainability. To propose inclusion policies or strategies, it is necessary to understand which attributes affect the probability of having a mobile account. We analysed individual-level variables such as age, gender, income level, education level, and employment, all of them have been used in similar works pertaining to financial inclusion (Fungáčová & Weill, 2015; Zins & Weill, 2016; Yangdol & Sarma, 2019; Martinez, Scherger, Guercio & Orazi, 2020). Also, there are going to be distinguish by country of origin and the use of other financial services to let the models determine which are more relevant in the classification of the data.

Unlike other studies that focus on traditional banking access, this article addresses mobile account ownership, a topic that has gained relevance in recent years. Combining a financial inclusion approach with classification models adapted to informal contexts provides a different approach that contributes to the field of future research or specific strategies to promote mobile account use.

The structure of the article is as follows. Section 2 shows the literature review related to digital financial inclusion. Section 3 provides the database considered for the analysis, followed by the methodology detailed in Section 4. Then, the results are presented and discussed in Section 5. Finally, Section 6 summarises the conclusions of the study.

1. Literature review

In recent decades, digital financial inclusion has emerged as a critical issue in the global economic landscape. In an increasingly digitally connected world, the ability to access and use financial services through digital platforms has become a key factor in driving economic growth and reducing inequality (World Bank, 2021). Mobile money services, internet banking, and other advances in financial technology offer several benefits: (i) greater access to other financial services, (ii) expense control, (iii) safety and convenience, and (iv) visualisation of credit history, which facilitates access to credit (Tay et al., 2022; Simplice, Biekpe & Cassimon, 2021; ENIF, 2020; Gomber, Koch & Siering, 2017).

While there is no standard definition of digital finance, it is widely agreed that it encompasses a range of products, services, technologies, and infrastructure that enable individuals' and businesses' access to payment, savings, and credit facilities through online channels. This eliminates the need for physically visiting a bank branch or directly engaging with financial service providers, thus aiding in cost-cutting and achieving more efficient internal processes (Gomber et al., 2017; Ozili, 2018).

When implemented ethically and sustainably in a well-regulated environment, digital financial inclusion fosters development and accelerates progress towards achieving the Sustainable Development Goals (SDGs), a set of global objectives established by the United Nations to address various social, economic, and environmental challenges by 2030 (Bastante, 2020; Ozili, 2018). The first step into greater digital financial inclusion is mobile money accounts. They can be linked to a bank account or not, with the issuer being a non-banking financial institution.

The disruptions caused by the COVID-19 pandemic accelerated the digitization of financial services. For more isolated and financially disadvantaged populations, especially in developing countries, digital banking, particularly mobile money, has proven to be the foundation for financial inclusion (Tram, Lai & Nguyen, 2021). In the aftermath of the pandemic, the rapid surge in demand for digital solutions from governments, businesses, and the general public is expected to create more opportunities for digital channels to promote global financial inclusion.

However, there still exist several obstacles to establishing an inclusive financial system in developing countries. In this sense, Tay, Tai & Tan (2022) conducted a systematic literature review on digital financial inclusion, revealing a significant gap in Asian developing countries, particularly in gender, income levels, and urban-rural disparities in accessing and utilising digital financial services. To address these challenges, the authors recommended enhancing digital infrastructure, streamlining banking processes, and emphasising the significance of financial education.

In addition, Parvin and Panakaje (2022) addressed the advantages, benefits, constraints, and disadvantages of digital financial inclusion. Their analysis noted the pivotal role of digital financial inclusion in driving socio-economic progress, fostering sustainable and inclusive prosperity, reducing costs, and bolstering the efficiency and competitiveness of service providers. Nevertheless, there are key difficulties such as inadequate financial literacy, limited rural access to technology, and concerns regarding trust and data privacy.

To understand digital financial inclusion, Aziz and Naima (2021) indicated that comprehending the social dynamics of financial interaction with new technologies requires departing from both simplistically analysing individual adoption or non-adoption and exclusively prioritising a supply-focused financial infrastructure. While digital services have facilitated access to financial services and reduced physical barriers, their underutilization is often attributed to insufficient connectivity, limited financial literacy, and inadequate social awareness.

In Latin America, digital financial services have quickly gained traction, mainly in payments and alternative finance. Cantú and Ulloa (2020) found that digital finance in Latin America is at a turning point, indicating a crossroads for how they can transform financial services. To fully leverage the benefits of digital financial inclusion, the authors considered it necessary to develop a robust regulatory framework and institutional capacity, increase investment in cybersecurity and data protection, improve rural connectivity, and bolster financial education.

The work of Ioannou and Wójcik (2022) observed that digital finance in Latin America has primarily operated on the periphery of the financial industry, contributing to the region's existing high concentration of financial services. However, its impact on financial inclusion has been limited. Similarly, Agufa Midika (2016) highlighted that banking institutions adopted digital financial services to lower operating costs associated with opening and managing branches, aiming to improve profitability and financial performance rather than to foster financial inclusion.

This work sought to address a gap in the literature by focusing on Latin American countries and analysing key characteristics of the population to determine their access to mobile money accounts (as a proxy of digital financial inclusion). In addition, we used artificial intelligence to identify the most relevant attributes for recognizing this group of excluded people.

2. Database

This paper used data from the 2021 Global Findex survey conducted by the World Bank. The Global Findex is a comprehensive, nationally representative survey designed to measure financial inclusion and the use of financial services worldwide. The survey, which has been conducted every three years since 2011 in collaboration with Gallup, Inc., collects data from more than 140 economies through face-to-face and telephone interviews. Respondents are selected using stratified random sampling to ensure a representative sample of each country's adult population. The data collection process adheres to rigorous quality control measures, including multi-stage sampling and post-stratification weighting to adjust for population demographics.

Its main objective was to measure the proportion of people that hold different financial instruments. It also included variables related to individual characteristics such as age, gender, education level (categorized into three categories), income level (divided into quintiles), and employment status, as shown in Table 1. These variables were selected based on their relevance, as demonstrated in the literature review (Tay et al., 2022; Fungáčová & Weill, 2015; Zins & Weill, 2016; Yangdol & Sarma, 2019; Martinez et al., 2020). Additionally, the survey does not provide further details regarding the individuals' socioeconomic background.

Given the extensive and complex nature of the dataset, Python was widely used for data cleaning, transformation, and analysis in this work.

This study analysed individuals from Argentina, Brazil, Colombia, and Paraguay, since they are among the largest nations in terms of geographic expanse in Latin America. This aspect is crucial for fostering digital financial inclusion by addressing barriers to traditional financial services, which arise from the lack of physical access points across vast geographical regions. The country of origin was depicted by four binary variables.

Table 1.
Descriptive statistics of the decision variable and attributes.

	Description	Average	Standard deviation	Min	Max
Target variable					
Mobile Account	0=Does not have a mobile account; 1=Has a mobile account	0.31	0.46	0	1
Attributes					
Gender	0=Male; 1=Female	0.48	0.50	0	1
Age	Continuous Variable	41.08	16.41	15	98
Education	1= Completed primary education; 2=Completed secondary education; 3= Completed higher education	2.05	0.59	1	3
Income quintiles	based on decision trees	3.4	1.4	1	5
Employment	1= Employed; 0 = Not employed	0.20	0.4	0	1
Mobile	1= Has a mobile phone; 0= Does not have a mobile phone	0.06	0.23	0	1
Internet	1= Has internet; 0= Does not have internet	0.19	0.39	0	1
Financial Account	1= Has a financial account;0= Does not have a financial account	0.28	0.45	0	1
Debit card (fin2)	1= Has a debit card; 0= Does not have a debit card	0.45	0.5	0	1
Saved	1= Saved ; 0=Not saved (in the last year)	0.56	0.5	0	1
Had a loan	1= Had a loan; 0=Did not have a loan (in the last year)	0.45	0.5	0	1
Total observations: 3935					

Source: Own elaboration based on the 2021 Global Findex Database.

3. Methodology

Decision trees and random forest models do not assume linearity in the relationships between variables. In contexts such as mobile account ownership in economies with high informality, the relationships between socioeconomic characteristics and mobile account adoption can be complex and not necessarily linear (like other methodologies assume like probit o logit models). Random forest and decision trees can effectively capture these non-linear relationships. They also allow us to obtain the importance of the variables, that is, the weight that each attribute has in the prediction. This is valuable for identifying the most relevant factors that influence mobile account ownership. Probit and logit models provide regression coefficients, but interpreting these coefficients, especially in terms of practical relevance, can be less intuitive and requires transformations.

Decision trees are graphical representations of decision-making processes employed in artificial intelligence and machine learning. They can be compared to flowcharts where decisions are made at each node, and different paths are followed based on specific conditions. The algorithm starts at the root node and recursively splits the data based on the attribute that best separates the classes. This splitting process continues until a stopping criterion is met, such as reaching a maximum depth or achieving a minimum number of samples in a leaf node. The goal is to create a tree that accurately classifies the training data while also showing a very visual graphic at the first nodes, which are the most important attributes for the

discrimination of individuals who do and do not have a mobile account, which is relevant for the design of digital financial inclusion policies.

The top node of the tree, known as 'root node,' indicates the initial decision, which is generally based on a characteristic or variable within the dataset. Further down the tree are the 'internal nodes,' each representing a decision according to a specific feature. These nodes divide the dataset into smaller subsets.

Leaves or 'leaf nodes' are located at the ends of the branches of the tree and represent the ultimate outcomes. For example, in a decision tree for predicting whether someone has a mobile account, the leaves indicate groups that are more or less homogeneous or 'pure' in terms of mobile account ownership. Each internal node imposes a rule or condition on the data. If a data instance meets that condition, it proceeds along the corresponding branch to the next node.

Decision trees find applications in diverse fields, spanning from data classification to decision making in business and medicine (Breiman, Friedman, Olshen & Stone, 1984). Their value lies in their simplicity, making them easy to interpret, thereby facilitating transparent, data-driven decision-making. This methodology has proven useful in a variety of financial and economic applications. It has been employed to predict credit risk, analyse the effectiveness of financial marketing campaigns, and segment customers based on their financial behaviour (Durica, Frnda & Svabova, 2019; Lin, Ke & Tsai, 2017; Prusak, 2018). Comparing with other methodologies, trees can be displayed graphically and are easily interpreted even by a non-expert (James, Witten, Hastie & Tibshirani, 2021). However, its role in the context of digital financial inclusion is still evolving, and its potential to uncover valuable insights has yet to be fully explored.

This paper uses several classification models based on decision trees, as it shows Figure 1. In principle, we employed a model capable of examining the first four levels, distinguishing each variable clearly from the most significant for classification (at the root of the tree) to the next four most relevant. This approach helps identify the variables that most affect an individual's likelihood of having a mobile account. In turn, assessing its metrics allows comparing the performance in classifying individuals or understanding patterns of differentiation based on their characteristics.

Then we controlled the tree growth by limiting it to a maximum of 100 samples for further splitting, enabling evaluation of the model's ongoing development and improvement in metrics. This constraint ensures a balance between model complexity and clarity, preventing excessive branching that could obscure the logical structure of the classification. Limiting the depth facilitates a more comprehensible hierarchical representation of decision boundaries, making it easier to analyze how the model differentiates between classes.

Finally, we constructed a random forest. Random forests build upon the concept of decision trees by creating an ensemble of them. Each tree in the forest is trained on a random subset of the data and a random subset of the features. When a new data point needs to be classified, each tree in the forest makes a prediction, and the final classification is determined by a majority vote. By averaging the predictions of multiple trees, random forests reduce the risk of overfitting and improve the overall accuracy and robustness of the model (James, Witten, Hastie & Tibshirani, 2021).

Given the inability to visualize how the task is performed and the aspects it considers in this last case, we provided the list of the main variables used for classification.

We employed the train_test_split method from the Python's sklearn library to build the test and test sets, with 20% of samples for testing and, in turn, stratification by country. In total, we performed the train sets with 796 individuals for Argentina, 790 for Peru, 787 for Colombia, and 775 for Brazil. While the test set consisted of 199 individuals for Argentina, 197 for Colombia and Peru, and 194 for Brazil.
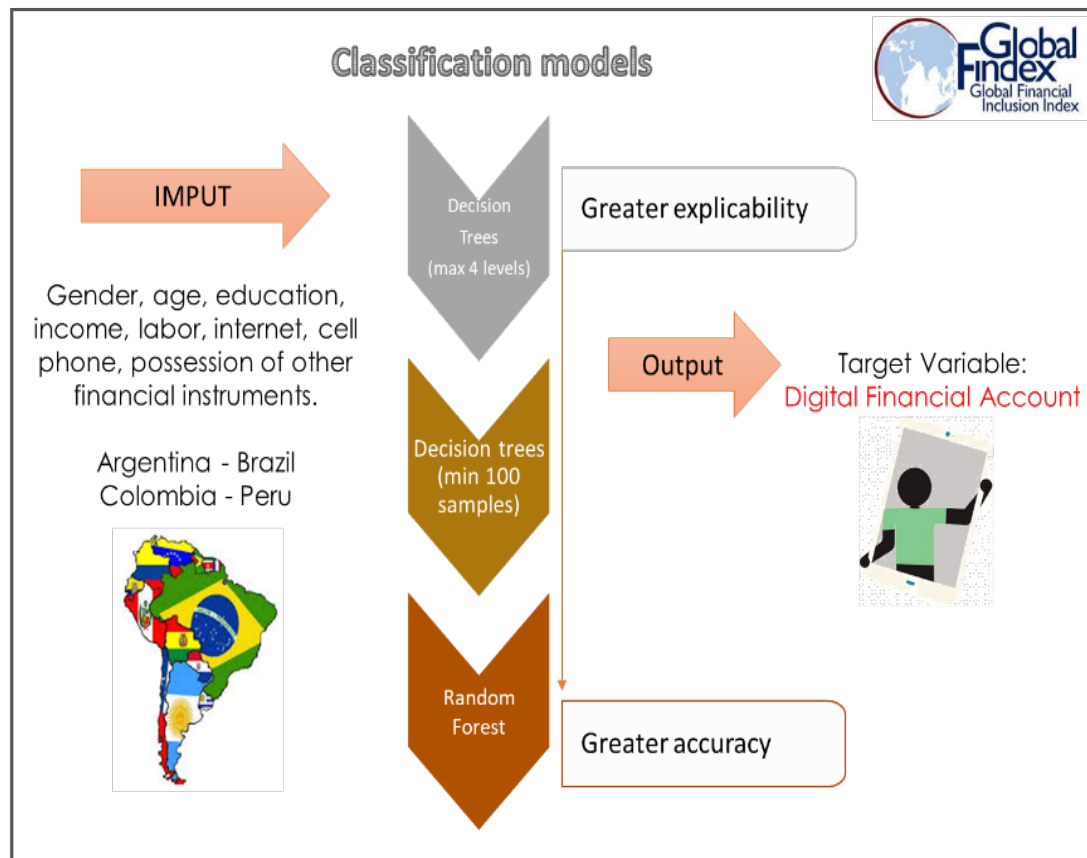
Figure 1. Summary of the methodologies applied.
Source: Own elaboration

As shown in Table 1, the proportion of the target variable ranges between 68% and 70% of individuals without a mobile account, indicating relatively unbalanced distribution. This situation mainly affects the interpretation of the model's performance measures. For instance, a dummy classifier predicting that nobody has mobile accounts would achieve an accuracy rate of 70%, suggesting that the model should demonstrate superior performance indicators. However, we also analysed other metrics that do not present this disadvantage in asymmetric samples, such as the area under the ROC curve (ROC-AUC). This goodness-of-fit measure allows comparing accuracy and precision across different models.

When developing the different classification methods, it should be noted that the decision tree was initially constructed with a maximum depth of four levels (max_depth=4) to understand how the classification is formed on the basis of the attributes that contribute most to solving the problem. It was then expanded to allow free development with a minimum of 100 observations in each leaf node (min_samples_split=100), providing it with greater flexibility and enabling performance measurement.

Next, we carried out the random forest analysis with the standardised data. Although its structure cannot be observed, the aim was to improve the accuracy of classification. To preserve the explanatory capacity of the classification process, we ranked the attributes based on their contribution to reducing the Gini index, a measure of impurity. Attributes were evaluated from most to least important in classifying individuals.

4. Results

Figure 2 shows the first decision tree, which was limited to a maximum of four levels. Darker colours indicate higher node purity, while lighter colours suggest limited ability to distinguish between individuals with or without a mobile account. Conversely, the hue or colour of the node, ranging from blue to orange, denotes whether the majority of the remaining sample in the node comprises individuals with (blue) or without (orange) a mobile account.
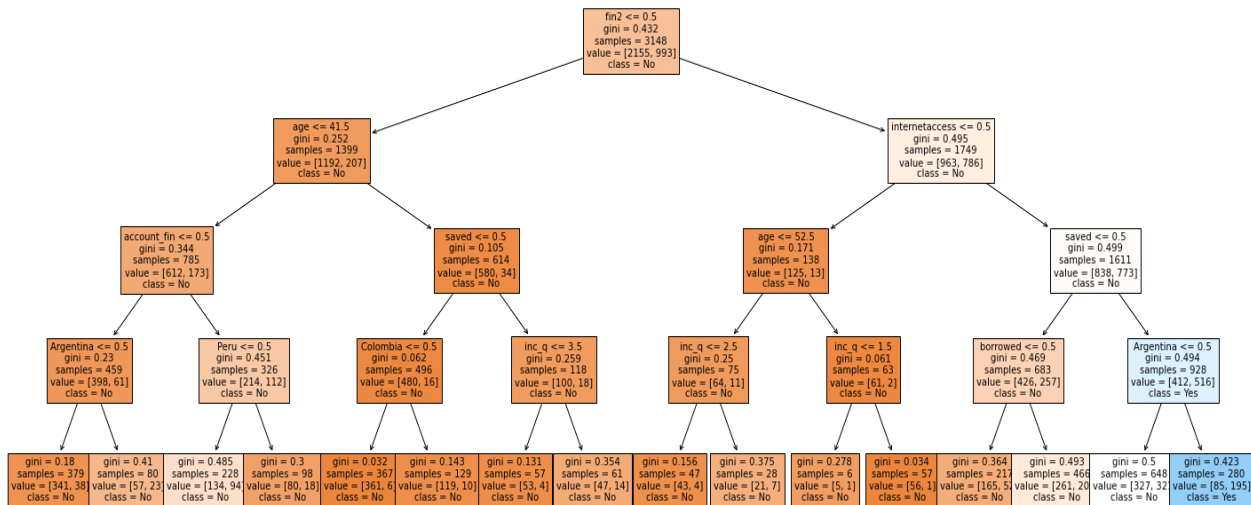
Confusion Matrix: Accuracy: 0.7204; Precision: 0.5694
Recall: 0.1782; F1 Score: 0.2715 ROC AUC: 0.5613

Figure 2. Summary of the methodologies applied.
Source: Own elaboration

Several evaluation metrics were used to assess the model's performance. Accuracy represents the proportion of correctly classified instances over the total number of instances, but it can be misleading in imbalanced datasets. Precision measures the proportion of true positive predictions among all instances predicted as positive, indicating how reliable positive classifications are. Recall (Sensitivity), on the other hand, calculates the proportion of true positive instances correctly identified by the model, highlighting its ability to capture actual positives. F1 Score is the harmonic mean of precision and recall, balancing both metrics in cases of class imbalance. The Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) measures the model's ability to distinguish between classes across different threshold settings, where higher values indicate better classification performance.

The central node identifies the variable debit card ownership as the main discriminant among the observations. Based on this criterion, it segregates individuals, placing those meeting the rule (<0.5, i.e., those with 0 in this attribute) in the left-hand branch and those not meeting this condition in the right-hand branch. Then, at the second level, it uses age and internet access criteria.

For example, individuals with a debit card are categorized on whether they have internet, saved in the last year, and belong to Argentina. This branch of the tree is unique in having a majority of individuals

with a mobile account, despite the high Gini index (0.423). The Gini Index is an impurity measure used to determine the best splits during training. It quantifies how often a randomly chosen element would be incorrectly classified if randomly labeled according to the class distribution in a node. Lower Gini values indicate purer nodes.

In the left-hand branch of the tree, individuals with a debit card are divided by age, with those under 41.5 directed to the left and those older than 41.5, to the right. At the third level, individuals are classified based on whether they have a bank account or whether they have saved in the last year. Finally, at the fourth level, other discriminants are used such as country of origin or income level.

The leaf with the lowest Gini index or impurity (0.032) comprises individuals who do not have a debit card, are older than 41 years, did not save in the last year and are not from Colombia. On the opposite side of the tree, there is another leaf with a low Gini index (0.034). Figure 3 shows the second decision tree, which was limited to a minimum of 100 samples per leaf. In this case, individuals possess a debit card but lack internet access, are older than 52 years, and do not belong to the first income quintile.

Regarding this model fit metrics, we observe an accuracy of 72.04%, which indicates the percentage of the sample that is correct, including both positives and negatives. However, due to the imbalance, it is not advisable to rely solely on this metric. By noting the distribution of the test set (523 no, 203 yes), if the classifier consistently predicts that no one has a mobile account, it would achieve 72% accuracy. The precision, on the other hand, indicates the percentage of individuals correctly classified as having a mobile account; that is, 56.94% of all positives are true positives. Another important metric is recall or completeness, which reveals the percentage of individuals with a mobile account that are correctly identified; in this case, it is 17.82%. Finally, the F1 Score combines precision and recall for ease of comparison, similarly to Area Under the Curve (ROC-AUC) measures the model's ability to distinguish between classes across different threshold settings, where higher values indicate better classification performance.

From the total test set, the confusion matrix reveals that 526 negative observations (indicating not having a mobile account) were correctly classified, while 31 negative observations were classified as positive (indicating having a mobile account). Conversely, 189 positive observations were incorrectly predicted as negative, and 41 negative observations were incorrectly predicted as positive.

These values are expected to improve in subsequent models, which, despite losing explainability, are more complex as they use deeper trees to model the output variable, that is, whether individuals have a mobile account or not.

In this case, the leaves with more observations continued to discern individuals and refine the selection of variables and thresholds to improve classification. However, considering the initial classification criteria, possessing a debit card, internet access, and having saved in the last year are key attributes for the tree to identify more nodes with a majority of individuals with a mobile account, as evidenced by the prevalence of blue nodes on the right.

With respect to age, the initial discriminant is whether the individual is younger or older than 41 years. This branch of the 'age' attribute, being at the second level, suggests its relevance. In turn, Argentina appears twice as a separating criterion at the second level, revealing distinct characteristics in comparison to the other countries.

In this model, accuracy rose to 74.2%, precision remained unchanged, but recall improved to 51.73%, resulting in a corresponding increase in F1 due to the enhanced recall. The area under the ROC curve also shows an improvement compared to the previous model. The confusion matrix evidences that, despite identifying fewer true negatives, the model more effectively discriminated true positives.

Figure 3. Summary of the methodologies applied.
Source: Own elaboration

Finally, a third classification model was performed using a random forest. Figure 4 presents the attributes in order of importance based on the model and its performance metrics. The importance of a feature is evaluated based on its contribution to reducing impurity, measured by the Gini Importance (for classification). In each tree, features that lead to the greatest decrease in impurity when used for splitting receive higher importance scores. By aggregating feature importance scores across all trees in the ensemble, Random Forest provides a robust ranking of attributes.
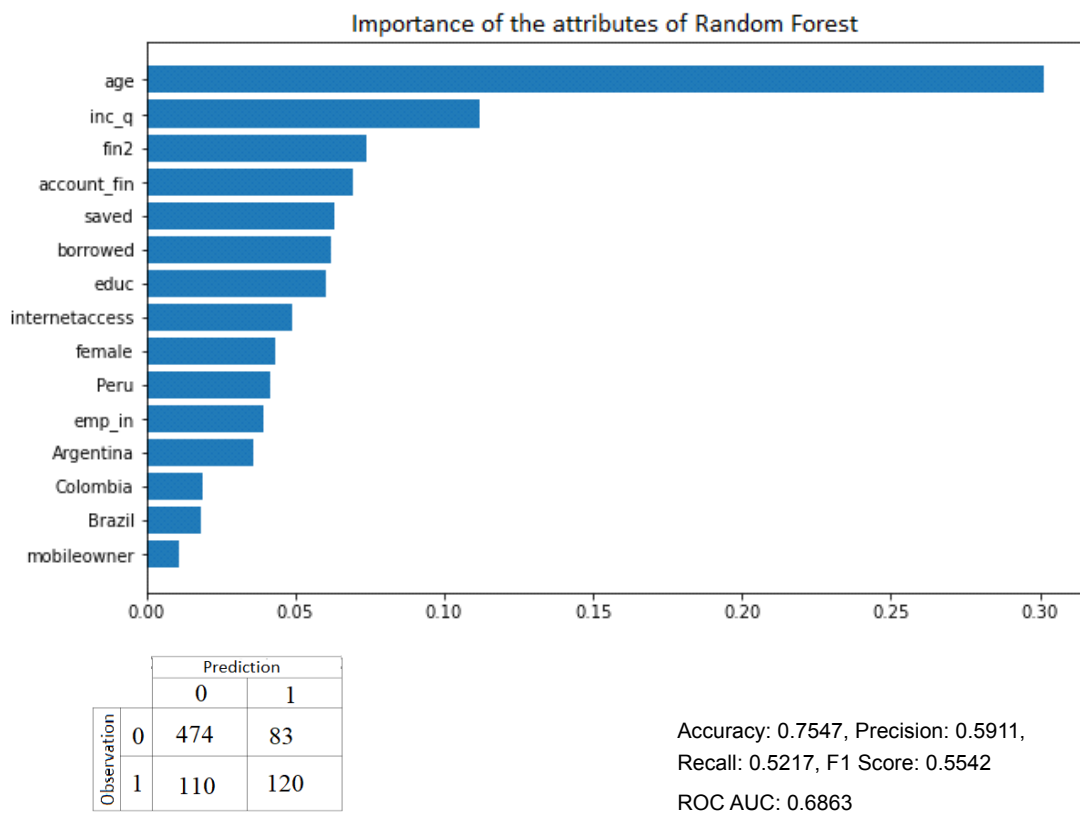

	Prediction	
	0	**1**
0	474	83
1	110	120

(Observation)

Accuracy: 0.7547, Precision: 0.5911,
Recall: 0.5217, F1 Score: 0.5542
ROC AUC: 0.6863

Figure 4. Importance of attributes in the random forest model and metrics.
Source: Own elaboration

The confusion matrix reveals a higher number of true-positive classifications compared to the previously analysed tree, although its contribution to the overall classification improvement was minor.

The variable age emerges as the most significant attribute for the classification due to its continuous nature and greater variability, enabling more effective discrimination among individuals. This is also why income, with its five categories, becomes apparent as another meaningful contributor to the classification. On the other hand, owning a debit card is the main instrument related to mobile accounts, as identified in the previous trees, followed by having a financial account, saving in the last year, or obtaining credit. Education is also among the most crucial attributes.

While this analysis does not reveal the specific classification criteria, such as age-related likelihood of having a mobile account, it is relevant for reinforcing findings from the initial trees. It also aids in attribute engineering, facilitating further exploration of those significant for understanding patterns of discrimination in digital financial inclusion.

The results confirm that individuals' demographic characteristics significantly influence their level of financial inclusion, as does the country in which they reside. While age emerges as a key determinant of digital financial inclusion, other variables-such as income level and the ownership and usage of specific financial instruments-also play a crucial role. Given these nuances, public policies in Latin American countries should be designed to target different social groups. Nonetheless, a central focus across all strategies should be the promotion of financial education. Enhancing financial literacy can generate a positive spillover effect by reducing fear and lack of awareness regarding financial tools, which in turn fosters greater use of digital financial instruments-driven by their speed and ease of use once individuals become familiar with their core features.

5. Conclusion

Identifying the factors that encourage mobile account ownership is a key objective for promoting digital financial inclusion. This market is experiencing rapid expansion and is essential for ensuring the reach of public policies aimed at equal access to other financial services. This was evident during the COVID-19 pandemic when it became necessary to provide assistance not only in healthcare but also through information and economic support to help people cope with quarantine measures and prevent contagion.

To investigate the factors influencing mobile account ownership among the population, this paper studied the characteristics of individuals in four major Latin American countries, which are highly limited in traditional financial inclusion and have a resilient informality share of the economy. To this end, we applied a simple four-level decision tree model, followed by a larger tree, and finally a random forest. Although the latter does not provide a visual representation of the classification process, it allowed us to determine the attributes most influential in reducing the Gini (or the impurity function used) throughout the forest's formation. This strengthened the results obtained from the analysed trees.

The conclusion drawn from the different models underscores the significance of debit card ownership compared to mobile account ownership. This factor may stem from individuals' familiarity with and experience in the financial market, or from the convenience of managing funds through a physical card. Internet access is essential, and having saved in the last year also plays a role in fostering interest in mobile account ownership. On the other hand, age influences the adoption of mobile accounts, with younger people generally more technologically connected and inclined toward embracing them. The main point of separation across the age continuum is 41 years. At the same time, individuals in Argentina also separate quickly from the rest of the sample, indicating that they have particular characteristics that are distinct from the rest of the countries.

Based on the findings, policy recommendations could focus on enhancing financial literacy programs to increase awareness and understanding of mobile account usage, especially among older demographics. Targeted efforts could be devoted to promote internet access and encourage saving habits, while also considering regional variations such as those observed in Argentina. Additionally, measures to facilitate the accessibility and usability of mobile banking services could be implemented to further incentivize adoption.

This study demonstrates how classification models can be applied to financial and social inclusion problems, showing an innovative way to address the problem of informality and barriers to financial access from data science. The use of decision trees and Random Forest provides greater flexibility, accuracy and robustness in predicting mobile account ownership in a context with high informality, where complex relationships between variables and noisy data make methodology based on linear models such as probit or logit less appropriate.

In future research, it would be important to continue exploring the sample of individuals, conducting separate models for each country to observe their differences, and using more powerful classification tools such

as neural networks. This is crucial for understanding their operation, as the goal is to identify population segments and their characteristics to enhance digital financial inclusion.

Moreover, along these lines, the sample size could also be expanded by including a larger number of countries and even incorporating other similar databases that offer a greater number of observations per country. Although, to date, we are not aware of the existence of databases comparable to the one developed by the World Bank and used in this study, the number of observations collected per country may be a critical factor when drawing conclusions about the promotion of financial instruments in each case.

About the authors

Sofia Orazi
Doctora en Ciencias de la Administración, Universidad Nacional del Sur, Argentina. Docente - Investigador, Instituto de Investigaciones Económicas y Sociales del Sur (IIESS-UNS-CONICET), Departamento de Economía, Universidad Nacional del Sur, Escuela de Ciencias Empresariales, Universidad Católica del Norte, Coquimbo, Chile.
sofia.orazi@uns.edu.ar   https://orcid.org/0000-0002-0611-8179

Lisana Belen Martinez
Doctora en Economia y Empresa, Universitat Rovira i Virgili - Facultad de Economia y Empresa-Reus, Espana. Docente - Investigador, Instituto de Investigaciones Económicas y Sociales del Sur (IIESS-UNS-CONICET) y Departamento de Economía, Universidad Nacional del Sur.
lbmartinez@iiess-conicet.gob.ar   https://orcid.org/0000-0001-5201-6651

Hernan Pedro Vigier
Doctor en Administración y Dirección de Empresas, Universidad Rovira i Virgili, Reus, España. Docente - Investigador, Instituto de Investigaciones Económicas y Sociales del Sur (IIESS-UNS-CONICET) y Departamento de Economía, Universidad Nacional del Sur.
hvigier@upso.edu.ar   https://orcid.org/0000-0003-0774-8620

Disclosure statement
The authors declare that there is no potential conflict of interest related to the article.

Availability of data
The authors declare that the article contains all the necessary and sufficient data for the understanding of the research.

Authors' contribution
Sofia Orazi: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing – original draft; Writing – review & editing

Lisana Belen Martinez: Conceptualization; Funding acquisition; Project administration; Resources; Supervision; Writing – review & editing.

Hernan Pedro Vigier: Conceptualization; Funding acquisition; Project administration; Resources; Supervision; Writing – review & editing.

References

1. AGUFA, Midika Michelle. The Effect of Digital Finance on Financial Inclusion in The Banking Industry in Kenya. Doctoral dissertation, University Of Nairobi, Nov. 2016. https://erepository.uonbi.ac.ke/handle/11295/98616

2. AZIZ, Abdul; NAIMA, Umma. Rethinking digital financial inclusion: Evidence from Bangladesh. In: Technology in Society, 2021, vol. 64, p. 101509. https://doi.org/10.1016/j.techsoc.2020.101509

3. BASTANTE, Marcelo. Estudio Fintech 2020. In: Banco Interamericano de Desarrollo. Ecosistema Argentino. Julio 2020. https://marcelobastante.com/wp/wp-content/uploads/2020/12/Estudio-Fintech-2020-Ecosistema-Argentino.pdf

4. BREIMAN, Leo; FRIEDMAN, Jerome; OLSHEN, R. A.; STONE, Charles J. Classification and regression trees. New York: Routledge.1984 https://doi.org/10.1201/9781315139470

5. CANTÚ, Carlos; ULLOA, Bárbara. The dawn of fintech in Latin America: landscape, prospects and challenges. Bank for International Settlements Papers. 2020. N. 112. https://www.bis.org/publ/bppdf/bispap112.htm?

6. DURICA, Marek; FRNDA, Jaroslav; SVABOVA, Lucia. Decision tree based model of business failure prediction for Polish companies. In: Oeconomia Copernicana, 2019, vol. 10, no 3, p. 453-469. https://doi.org/10.24136/oc.2019.022

7. ENIF. Estrategia Nacional de Inclusión Financiera. Ministerio de la Nación. Argentina 220. https://www.argentina.gob.ar/sites/default/files/enif_2020-23_vf _011220_con_prologo_1.pdf

8. FUNGÁČOVÁ, Zuzana; WEILL, Laurent. Understanding financial inclusion in China. In: China Economic Review. July 2015. vol. 34. p. 196-206. https://doi.org/10.1016/j.chieco.2014.12.004

9. GOMBER, Peter; KOCH, Jasha-Alexander; SIERING, Michael. Digital Finance and FinTech: current research and future research directions. In: Journal of Business Economics. 2017 vol. 8. p. 537-580. https://doi.org/10.1007/s11573-017-0852-x

10. IOANNOU, Stefanos; WÓJCIK, Dariusz. The limits to FinTech unveiled by the financial geography of Latin America. In: Geoforum, 2022, vol. 128, p. 57-67. https://www.sciencedirect.com/science/article/pii/S0016718521003183

11. JAMES, Garet; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. An Introduction to Statistical Learning: with Applications in R. New York: Springer. 2021. https://www.stat.berkeley.edu/~rabbee/s154/ISLR_First_Printing.pdf

12. KLING, Gerhard; Pesqué-Cela, Vanesa; Tian, Lihui.; Luo, Deming. A theory of financial inclusion and income inequality. In: The European Journal of Finance, 2022, vol. 28, no 1, p. 137-157. https://doi.org/10.1080/1351847X.2020.1792960

13. LEVINE, Mr Ross. Finance, growth, and inequality. International Monetary Fund. 2021. https://www.imf.org/-/media/Files/Publications/WP/2021/English/wpiea2021164-print-pdf.ashx/1000

14. LIN, Wei-Chao; KE, Shih-Wen; TSAI, Chih-Fong. Top 10 data mining techniques in business applications: a brief survey. In: Kybernetes, 2017, vol. 46, no 7, p. 1158-1170. https://doi.org/10.1108/k-10-2016-0302

15. MARTINEZ, Lisana Belén; SCHERGER, Valeria; GUERCIO, María Belén;ORAZI , Sofia. Evolution of financial inclusion in Latin America. In: Academia Revista Latinoamericana de Administración. 2020. vol. 33, no 2, p. 261-276. https://www.emerald.com/insight/content/doi/10.1108/arla-12-2018-0287/full/html

16. ORAZI, Sofía; MARTINEZ, Lisana Belén; VIGIER, Hernan Pedro. Determinants and evolution of financial inclusion in Latin America: A demand side analysis. In: Quantitative Finance and Economics, vol. 7. No 2, 187-206. https://doi.org/10.3934/QFE.2023010

17. OZILI, Peterson K. Impact of digital finance on financial inclusion and stability. In: Borsa Istanbul Review, 2018, vol. 18, no 4, p. 329-340. https://doi.org/10.1016/j.bir.2017.12.003

18. PARVIN, S. R.; PANAKAJE, Niyaz. A study on the prospects and challenges of digital financial Inclusion. In: Education (IJCSBE), 2022, vol. 6, no 2, p. 469-480. https://doi.org/10.5281/zenodo.7259013

19. PIOTROWSKA, Adrianna. Determinants of consumer adoption of biometric technologies in mobile financial applications. In: Economics and Business Review, 2024, vol. 10, no 1, p. 81-100. https://www.ceeol.com/search/article-detail?id=1239518

20. PRUSAK, Błażej. Review of research into enterprise bankruptcy prediction in selected central and eastern European countries. In: International Journal of Financial Studies, 2018, vol. 6, no 3, p. 60. https://doi.org/10.3390/ijfs6030060

21. SHARMA, Anupama; KUKREJA, Sumota. An analytical study: Relevance of financial inclusion for developing nations. In: International journal of engineering and science. 2013. vol.2 no.6. p.15-20. https://www.scirp.org/reference/referencespapers?referenceid=2755643

22. SIMPLICE, Asongu Anutechia; BIEKPE, Nicolas; CASSIMON, Danny. On the diffusion of mobile phone innovations for financial inclusion. In: Technology in Society. 2021. vol. 65, 101542. https://doi.org/10.1016/j.techsoc.2021.101542

23. TAY, Lee-Ying; TAI, Hen-Toong; TAN, Gek-Siang. Digital financial inclusion: A gateway to sustainable development. In: Heliyon, 2022, vol. 8, no 6. https://doi.org/10.1016/j.heliyon.2022.e09766

24. TRAM, Thi Xuan Huong; LAI, Tien Dinh; NGUYEN, Thi Truc Huong. Constructing a composite financial inclusion index for developing economies. In: The Quarterly Review of economics and finance, 2023, vol. 87, p. 257-265. https://doi.org/10.1016/j.qref.2021.01.003

25. World Bank, Digital Financial Inclusion, 2021.
    https://www.worldbank.org/en/topic/financialinclusion/publication/digital-financial-inclusion

26. YANGDOL, Rigzin; SARMA, Mandira. Demand-side factors for financial inclusion: A cross-country empirical analysis. In: International Studies, 2019, vol. 56, no 2-3, p. 163-185. https://doi.org/10.1177/0020881719849246

27. ZINS, Alexandra; WEILL, Laurent. The determinants of financial inclusion in Africa. In: Review of development finance, 2016, vol. 6, no 1, p. 46-57. https://hdl.handle.net/10520/EJC193922