International Journal of Accounting and Economics Studies, 12 (8) (2025) 232-239

International Journal of Accounting and Economics Studies

Website: www.sciencepubco.com/index.php/IJAES
https://doi.org/10.14419/6wszea16
Research paper

Predicting Youth Financial Inclusion: A Machine Learning Classification Approach Using Access, Quality, and Demographic Determinants

Om Mansatta [1] *, Dr. Sampada Kapse [2]

[1] *Research Scholar, Gujarat Technological University, Ahmedabad, Gujarat-India*
[2] *Director, Tolani Motwane Institute of Management Studies, Adipur, Gujarat-India*
**Corresponding author E-mail: om.mansatta@gmail.com*

Received: October 25, 2025, Accepted: November 28, 2025, Published: December 8, 2025

Abstract

Financial inclusion has become a cornerstone of inclusive growth in India, yet its outreach among youth remains uneven and underexplored. This study measures the determinants of financial inclusion among the youth of Kutch district by applying both Logistic Regression and Chi-square Automatic Interaction Detection (CHAID) models. The analysis is based on primary data collected from 410 respondents across diverse demographic and occupational backgrounds. The logistic regression model achieved an overall prediction accuracy of 90.2% and a Nagelkerke R^2 value of 0.805, signifying strong explanatory power. The findings reveal that Access and Quality dimensions—derived from the Reserve Bank of India's Financial Inclusion framework—are the most significant predictors of inclusion. Occupation and income also exhibit meaningful influence, indicating that employment and financial capability enhance formal participation. The CHAID model further validates these results, uncovering hierarchical relationships between access, quality, and socio-economic variables. The combined interpretation of both models provides a comprehensive view of youth inclusion behavior and highlights that accessibility, service quality, and economic opportunity jointly determine inclusion outcomes. The study concludes with policy recommendations aimed at strengthening digital availability, improving service reliability, and promoting youth employment to accelerate financial inclusion in semi-urban and rural India.

Keywords: *Access and Quality Dimensions; CHAID Model; Financial Inclusion; Logistic Regression; Socio-economic Determinants.*

1. Introduction

Financial inclusion has emerged as a critical pillar for achieving inclusive growth in modern economies. Its importance gained momentum after being recognized as a key enabler for nine out of seventeen Sustainable Development Goals (SDGs) 2030 (Demirguc-Kunt et al., 2017). According to the UN, financial inclusion has the potential to accelerate progress across multiple SDGs by expanding access to essential financial services. Broadly, financial inclusion refers to the availability and affordability of basic financial services to the marginalized groups who are otherwise excluded from the formal financial system. (Singh & Roy, 2015). In the Indian context, the first major institutional step was taken in 2021, when the Reserve Bank of India (RBI) introduced a comprehensive Financial Inclusion Index (FI-Index) for the first time. The composite index is built on 97 indicators structured under broader dimensions of Access, Usage, and Quality, which further encompass sub-dimensions such as Banking, Insurance, Credit, and Digital Finance. (RBI, 2021). The index ranges from 0 to 100 and is updated every year to show the progress of financial inclusion in India, which has risen to 67 in 2025 from 53.9 in 2021.

Most Indian studies on financial inclusion have relied heavily on descriptive statistics or on the construction of composite indices, which would have limited coverage. Also, these studies are based on the collection of secondary data sources from RBI publications, the Global Findex database, or CRISIL Inclusix. Some of these datasets are very recent, while some of these are outdated, creating inconsistency in comparability and limiting the ability to capture the fast-changing financial behavior. Another limitation is the lack of focus on youth populations. Although several studies address rural households, women, or the poor in general, very few focus on youth, despite their central role in the adoption of digital technologies and future economic participation.

The present studies aim to move beyond the descriptive and index-based approaches to a predictive modelling approach that classifies the youths into financially included or excluded. Specifically, the study used data-driven usage dimensions to define inclusion status, and then employed machine learning classification methods such as Logistic Regression and Decision Tree to predict the outcome. The objective is two-fold: first, to examine whether machine learning methods can achieve higher accuracy than traditional statistical methods in classifying financial inclusion status, and second, to identify the most critical variables – across the Access, Usage, Quality dimensions, and key demographic variables – that influence youth participation in financial inclusion. By focusing on young individuals in the Kutch district, the study also addresses the major gap in Indian literature, where youth-specific financial inclusion is limited. The ultimate aim is to generate insights that can guide policymakers in designing target interventions for accelerating inclusive growth.

This study contributes to three levels. Theoretically, it extends the Reserve Bank of India's Financial Inclusion framework of Access, Usage, and Quality by integrating it with predictive analytics. In doing so, it shifts the focus from merely describing the inclusion levels to understanding the determinants that predict whether a young individual is financially included or excluded. Methodologically, the research uses the application of machine learning models – specifically Decision Tree – to study financial inclusion. While logistic regression has also been used, the adoption of these advanced algorithms provides stronger classification accuracy and reveals the relative importance of variables in determining inclusion. Practically, the findings offer actionable suggestions for policymakers and financial institutions by identifying youth segments who are more likely to remain excluded. This enables the design of focused interventions, such as financial literacy programs, digital access initiatives, or credit support schemes, to accelerate inclusive growth in regions like Kutch and beyond.

2. Literature Review

2.1. RBI's financial inclusion framework (2021): access, usage, and quality

As stated above, in 2021, the Reserve Bank of India (RBI) released its Financial Inclusion Index as a composite score that monitors the progress of financial inclusion in India. The index was based on 97 indicators across three dimensions—Access, Usage, and Quality—which were developed in consultation with multiple financial sector regulators. Weights were assigned to the dimensions—35% for Access, 45% for Usage, and 20% for Quality—reflecting greater importance of financial behavior rather than mere access to financial services. (RBI, 2021). This framework provides the basis for many recent academic and policy studies on inclusion.

2.2. Global evidence on the three dimensions

Access (Banking Reach and Account Ownership): Access refers to the availability of affordable and convenient financial services such as bank branches, ATMs, business correspondents, and account ownership. Globally, account ownership has expanded significantly—from 51% in 2011 to 79% in 2025—accelerated largely by digitalization and government-led initiatives. (Klapper et al., 2025). In India, programs like Pradhan Mantri Jan-Dhan Yojna (PMJDY) have improved the basic account availability, while the RBI framework emphasizes that access is a necessary but not sufficient condition for deeper inclusion. (RBI, 2021).

Usage (Active engagement with financial services): Usage captures how actively individuals use accounts for saving, borrowing, making payments, and holding insurance. Global data highlights the surge in the use of digital payment adoption, particularly through mobile and internet channels. (Demirgüç-Kunt et al., 2021a). In the Indian context, the FI-index allocates the highest weight, 45%, to usage, underscoring that inclusion cannot be achieved through account ownership alone. Recent evidence shows that government transfers through accounts and UPI-based digital payments have increased the active use of financial services among youth. (Asli Demirgüç-Kunt et al., 2017; Demirgüç-Kunt et al., 2021b).

Quality (literacy, trust, and consumer protection): The quality dimension evaluates how effectively financial services meet user needs, with sub-dimensions such as financial literacy, grievance redressal, transparency, and equitable distribution. International frameworks such as the OECD/INFE toolkit emphasize financial knowledge, financial behavior, and financial attitude as critical elements of quality. (OECD, 2023). Similarly, CGAP highlights the significance of consumer protection in safe digital finance. (Partridge et al., 2022). The RBI, in its FI-Index, allocates 20% weightage to it, signaling that without trust, financial literacy, and consumer protection, access and usage may not translate into significant inclusion. (RBI, 2021).

2.3. Financial inclusion of youth: global evidence

Global evidence highlights that young people face distinct obstacles to financial inclusion—age-based restrictions, limited collateral, low financial literacy, and restricted access to formal services. International studies show that, across Africa, Asia, and the MENA region, youth inclusion is strongly influenced by education, income, digital access, and employment status (FAO, 2023; Ikosa, 2025; Isaga, 2025; Elmasmari, 2024). These findings consistently emphasise that youth are early adopters of digital finance but remain disproportionately excluded, mainly due to structural and regulatory barriers rather than behavioural reluctance.

2.4. Financial inclusion of youth in India

Research specifically focusing on youth (roughly 18-35 years) is very limited in India. National-level financial inclusion assessment studies such as CRISIL Inclusix, RBI FI-Index, and Global Findex also use aggregate or all-adult samples that are also based on secondary data and descriptive or index-based methods rather than predictive modeling. As a result, there is a gap in identifying youth-related drivers for financial inclusion and developing predictive tools that can help us identify young individuals at risk of exclusion.

Demographically, India has the largest youth population in the world. The Ministry of Statistics and Programme Implementation (MoSPI) estimates that the youth population is about 27.2% of the total population in 2021. The earlier definition of "youth" as defined in the National Youth Policy (2003) defined "youth" more broadly as those aged from 13-35 years (Shahin Sultana, 2014). This earlier framework recognized social, economic transitions such as education, employment, and family formation often extend well into mid-thirties.

Building on this broader understanding, the present study adopts the age group of 18-35 years. This aligns with the legal threshold of financial autonomy (18 years) and captures a segment that is both economically active and financially decision-capable. Individuals below the age of 18 years are legally restricted from operating the account and are dependent on a parent for finance, whereas those in their early to mid-thirties remain in formative finance stages – such as pursuing higher education, entering the job market, or starting a business, managing their finances, and householding. Therefore, adopting 18-35 years not only reflects the spirit of the earlier National Youth Policy but also provides a financially meaningful age window for studying inclusion and exclusion patterns.

Digital Capability, a cornerstone of modern financial participation, shows progress but persistent inequality. National surveys highlight the expansion of internet availability with market divides. The NSS 75th round (2017-18) reported that only 20-25% of individuals aged above 5 years and above could use the internet with significant rural-urban gaps. NFHS-5 (2019-21) found that merely 33.3% of women aged 15-49 had ever used the internet, revealing disparities by age, residence, and education. The internet user base of India has grown to 821-886 million in 2023-24, with rural households forming the majority. Thus, while digital infrastructure has expanded impressively, digital literacy and consistent usage among youth remain uneven, particularly along gender and regional lines – an essential context for financial inclusion efforts.

Most existing Indian studies have examined financial inclusion descriptively, focusing on the availability and usage of financial services and comprehensive indices derived from secondary databases. (CRISIL Inclusix, 2013; RBI, 2021; Sarma & Pais, 2011). Even recent analyses on Global Findex 2021 primarily focused on relationship findings rather than predictive insights. (Demirgüç-Kunt et al., 2021). Consequently, there exists a clear research gap in youth-focused, primary-data studies employing predictive techniques such as decision trees to classify inclusion outcomes and rank determinants—a gap this study aims to fill.

2.5. Machine learning & financial inclusion

Recent progress in the field of social finance research has overseen a shift from traditional econometric techniques to Machine Learning (ML) methods, which can handle non-linear relationships with complex data. Classical tools such as Logistic Regression have been used to estimate the possibilities of financial participation or loan default, particularly in microfinance and PMJDY uptake studies (Banerjee et al., 2017; Bapat, 2020; Kaur & Kapuria, 2023). Logistic models provide interpretable coefficients and are suitable for binary outcomes such as included vs. excluded or account holder vs. non-holder. Studies in Peru, Belize, and global datasets demonstrate that ML models improve prediction accuracy and help identify exclusion risks (Hersh, 2021; Maehara, 2024; Bazarbash, 2019). More recent work shows that ML-based lending and alternative-data scoring can expand credit availability for underserved groups (Daida & Kumar, 2024; Sreeram & Kumar, 2024). However, existing ML studies focus largely on adults, credit portfolios, or national-level indicators—very few examine youth-specific financial inclusion using micro-level access and quality measures.

Building on this foundation, Decision Tree algorithms offer a transparent, rule-based classification structure widely applied in consumer finance and credit scoring (Ghosh & Dutta, 2019; Nwankwo et al., 2021). They enable visual representation of decision paths, making them particularly valuable for policymakers and practitioners seeking to understand which demographic or behavioral variables drive inclusion.

Given these strengths, applying ML models such as Logistic Regression and Decision Trees to financial inclusion research provides both predictive and interpretive value. They allow not only estimation of inclusion probability but also identification of key determinants—offering actionable insights for targeted interventions among youth populations.

2.6. Research gap

Empirical studies are largely descriptive when it comes to financial inclusion despite the fact of its policy significance and global institutional attention. Most studies only cover the level of financial inclusion in India and use some basic regression tools, and do not include any predictive analysis (Sarma & Pais, 2011; RBI, 2021; World Bank, 2022). Only a few explore classification or forecasting frameworks, and virtually none focus on youth-specific financial inclusion using ML techniques.

This gap is particularly important because youth—representing a digitally connected yet financially diverse segment—display heterogeneous inclusion behaviors that cannot be captured through linear models alone. Existing studies seldom address who among youth is at risk of exclusion or which factors most strongly predict inclusion.

Therefore, the present study addresses this lacuna by collecting primary data from youth in the Kutch district and applying logistic regression and CHAID decision-tree models to predict their financial inclusion status. This design integrates RBI's Access and Quality dimensions at the individual level with modern predictive analytics, allowing the study to (i) classify youth into financially included and excluded groups, (ii) quantify the relative importance of structural and socio-economic predictors, and (iii) identify high-risk youth segments for targeted policy intervention in a semi-urban and rural Indian context.

3. Objectives

- To examine the key determinants of financial inclusion among the youth of Kutch district using the dimensions of Access, Usage, and Quality proposed by the Reserve Bank of India.
- To apply Logistic Regression and CHAID models to predict and classify financially included and excluded youth based on socio-economic and demographic factors.
- To interpret and compare the results of both models to provide practical insights and policy recommendations for enhancing youth financial inclusion in semi-urban and rural India.

4. Research Methodology

The study adopts the descriptive, quantitative, and predictive research design. The study aims to measure and classify the individuals in the financially included or financially excluded groups. The individuals belonging to the age group of 18-35 years (youth) are from the Kutch district. The financial inclusion is measured on the basis of a multidimensional framework, which includes Access, Usage, and Quality, given by RBI in 2021. The study further applies Machine Learning models such as Decision Tree and Logistic regression to predict the inclusion outcomes and key determinants.

The target population comprises youth aged from 18-35 years and are residents of the Kutch district, Gujarat. This age range aligns with the National Youth Policy (2003), which reflects the financially active segment of society. A sample of 410 respondents was selected using a convenience sampling technique, which meets the sample size as recommended by Hair et al. (2019), which is sufficient for predictive analysis. The data was collected through a structured questionnaire, which was administered online and offline. Each Dimension has further sub-dimensions, and the individual indicators were normalized using the Min-Max method. And further, a final score was derived, giving the weightage as mentioned by (RBI, 2021).

Although the past research shows that Usage – active engagement with financial services – plays an important role in achieving financial inclusion (Allen et al., 2012; Barajas et al., 2020; Pandey et al., 2025), the present study uses only access and quality in the predictive modelling stage. Usage is already incorporated in the composite Financial Inclusion Index to maintain alignment with the RBI framework; however, the analytical objective of the study is to examine structural, supply-side determinants of inclusion rather than behavior adoption. Since Usage typically becomes relevant only after Access barriers are resolved (Allen et al., 2012; Pandey et al., 2025), modelling Access and Quality separately provides cleaner insights into the systematic constraints faced by youth. The dependent variable, financial inclusion, was categorized as financially included or financially excluded using FII scores. Independent variables include demographic variables and individual indicators of Access and Quality dimensions. The data was analyzed using SPSS. The study is limited to the Kutch district and self-reported data, which may introduce response bias. However, careful design and validation were employed to minimize such effects.

5. Analysis and Interpretation

This chapter presents the analysis and interpretation of the primary data collected from respondents aged 18-35 years (youth) of the Kutch region. The main aim is to measure and understand the level of financial inclusion among respondents by examining three key dimensions – Access, Usage, and Quality – as defined by RBI.
The demographic composition of the sample revealed that 59.3 % of the respondents were male and 40.7 % were female. The majority of them fall into the age group of 22-25 years (34.1%), followed by 26-29 years (28.4%). For occupation, 41.7% were students, followed by private employees, who were 27.6%. The average value of Access was 0.62, for Usage was 0.58, and for Quality it was 0.55, indicating a moderate level of financial inclusion in the Kutch district.

5.1. CHAID decision tree

5.1.1. Tree overview

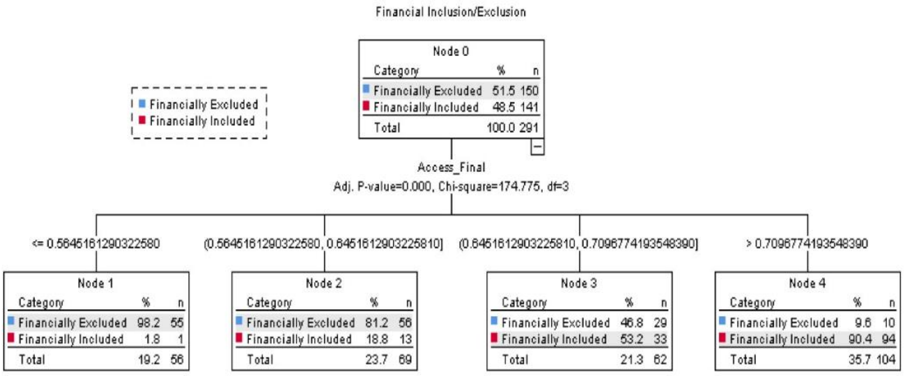

Fig. 1: Training Samples

Figure 1. CHAID decision tree predicting Financial Inclusion/Exclusion using the Access dimension (Access_Final) for the training sample (n = 291). To further classify the respondents into financially included or financially excluded and to identify the key determinants among youth of Kutch district, a Classification Tree Analysis using the CHAID algorithm was conducted. Financially Included or Excluded was the dependent variable, while age, gender, occupation, place of residence, educational qualification, marital status, average monthly household income, Access, and Quality were independent variables. Figure 1 revealed that only Access emerged as a significant splitter (χ^2 = 174.775, $\rho < 0.001$), indicating that it is the most decisive factor in differentiating the individual into financially excluded or financially included.
There are four nodes in Figure 1. Node 0 shows overall distribution – 54.6% of the respondents were financially included, while 45.4% of the respondents were financially excluded. The decision tree then divided them into four terminal nodes based on Access. Respondents with low scores (≤ 0.5846) formed Node 1, where 95.0% were financially excluded, demonstrating a clear concentration of exclusion among those with poor access. In contrast, Node 4, representing respondents with high access levels (> 0.7069), showed 88.9% inclusion, highlighting that greater access significantly enhances financial inclusion levels. Intermediate groups (Node 2 and Node 3) displayed gradual improvement in inclusion percentages (16.7% and 57.1%, respectively).

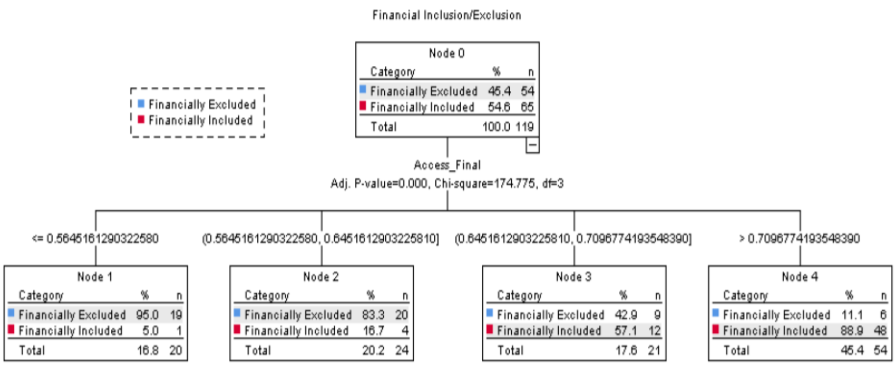

Fig. 2: Test Samples.

Figure 2. CHAID decision tree predicting Financial Inclusion/Exclusion using the Access dimension (Access_Final) for the test sample (n = 119). The model structure replicates the training sample, confirming Access_Final as the dominant predictor and demonstrating robust classification performance across four terminal nodes. The CHAID model achieved an overall classification accuracy of 84.4% in the test sample, with a risk estimate of 0.16, suggesting strong predictive performance.
In summary, the CHAID analysis substantiates that Access plays a critical role in determining the financial inclusion status of youth. The probability of being financially included rises sharply with higher levels of availability of formal financial services such as banking outlets, ATMs, digital facilities, and simplified documentation. These results underscore the need for policy interventions aimed at improving the reach and ease of financial access for young individuals in the Kutch district.

5.1.2. Splits & terminal nodes

Table 1: Target Category: Financially Excluded

Gains for Nodes							
Sample	Node	Node N	Percent	Gain N	Percent	Response	Index
Training	1	56	19.2%	55	36.7%	98.2%	190.5%
	2	69	23.7%	56	37.3%	81.2%	157.4%
	3	62	21.3%	29	19.3%	46.8%	90.7%
	4	104	35.7%	10	6.7%	9.6%	18.7%
Test	1	20	16.8%	19	35.2%	95.0%	209.4%
	2	24	20.2%	20	37.0%	83.3%	183.6%
	3	21	17.6%	9	16.7%	42.9%	94.4%
	4	54	45.4%	6	11.1%	11.1%	24.5%
Growing Method: CHAID							
Dependent Variable: Financial Inclusion/Exclusion							

Table 1 presents the Gains Table for the CHAID model, summarizing node-wise sample distribution, response rates, and index values for both the training and test samples. As the availability of formal financial services increases, exclusion drops sharply—from 98 % in low-access to 10 % in high-access groups. The Index column (e.g., 190.5 % for Node 1) shows how much higher that node's share of exclusion is compared with the overall sample average. Thus, Node 1 has almost 1.9 times the average exclusion rate, confirming that poor access is the strongest predictor of exclusion.

Table 2: Target Category: Financially Included

Gains for Nodes							
Sample	Node	Node N	Percent	Gain N	Percent	Response	Index
Training	4	104	35.7%	94	66.7%	90.4%	186.5%
	3	62	21.3%	33	23.4%	53.2%	109.8%
	2	69	23.7%	13	9.2%	18.8%	38.9%
	1	56	19.2%	1	0.7%	1.8%	3.7%
Test	4	54	45.4%	48	73.8%	88.9%	162.7%
	3	21	17.6%	12	18.5%	57.1%	104.6%
	2	24	20.2%	4	6.2%	16.7%	30.5%
	1	20	16.8%	1	1.5%	5.0%	9.2%
Growing Method: CHAID							
Dependent Variable: Financial Inclusion/Exclusion							

As shown in Table 2, Node 4 yields the highest response rate and index values for financially included youth, indicating this segment is the most likely to be included. The inclusion rate grows dramatically, from 1.8 % in the lowest group to 90 % in the highest. The Index (> 100) indicates above-average inclusion strength in those nodes. Node 4's Index = 186.5 % shows inclusion, which is nearly double the sample average.

5.1.3. Model accuracy

Table 3: Risk

Risk		
Sample	Estimate	Std. Error
Training	.182	.023
Test	.168	.034

Table 4: Classification

Classification				
Sample	Observed	Predicted Financially Excluded	Financially Included	Percent Correct
Training	Financially Excluded	111	39	74.0%
	Financially Included	14	127	90.1%
	Overall Percentage	43.0%	57.0%	81.8%
Test	Financially Excluded	39	15	72.2%
	Financially Included	5	60	92.3%
	Overall Percentage	37.0%	63.0%	83.2%

Taken together, Table 3 and Table 4 show that the CHAID model performs consistently well across both datasets, reflecting low misclassification risk and high overall classification accuracy.

5.1.4. Segment interpretation

Both samples show low risk (< 0.20), confirming that the model predicts accurately and generalizes well to new data. The CHAID model correctly classifies over 80 % of respondents in both samples, with particularly strong performance for identifying financially included youth (≈ 90 % accuracy). The high test accuracy shows the model is not overfitting.

The CHAID classification tree (Access→ Financial Inclusion/Exclusion) produced four terminal nodes, each representing a distinct level of access. The model achieved high predictive accuracy (Training = 81.8 %; Test = 83.2 %) with low risk (0.18 and 0.17, respectively). As access scores increased, the likelihood of financial inclusion rose sharply—from 1.8 % in the lowest access group to 90.4 % in the highest. The chi-square value for the first split was significant, χ^2 (3) = 174.78, p < .001.

5.2. Logistic regression

5.2.1. Model fit

Table 5: Logistic Regression Model Summary

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	186.796[a]	.602	.805

The logistic model exhibits high explanatory strength, as highlighted in Table 5, where the Nagelkerke R^2 (.805) and Cox & Snell R^2 (.602), indicating that approximately 80 % of the variation in financial inclusion among youth respondents is accounted for by the selected predictors. The 2 Log Likelihood value of 186.796 reflects a good model fit when compared with the null model. These results confirm that the combination of access, quality, and socio-economic factors provides a robust explanation of financial inclusion behaviour among youth in Kutch.

5.2.2. Classification accuracy

Table 6: Classification Accuracy

Classification Table					
	Observed		Predicted Financial Inclusion/Exclusion		
			Financially Excluded	Financially Included	Percentage Correct
Step 1	Financial Inclusion/Exclusion	Financially Excluded	204	20	91.1
		Financially Included	20	166	89.2
	Overall Percentage				90.2
a. The cut value is .500					

As shown in Table 6, the logistic regression model correctly classified 91.1% of financially excluded youth and 89.2% of financially included youth, resulting in an overall accuracy of 90.2%.

Using a default cut-off value of 0.5, the model demonstrates strong discriminatory power between included and excluded youth. This suggests that the selected predictors—particularly Access and Quality dimensions along with occupation and income—effectively distinguish financially included individuals from excluded ones.

5.2.3. Significant predictors

Table 7: Significant Variables in Equation

Predictor Variable	B	S.E.	Wald	Sig. (p)	Exp(B)	95 % C.I. for Exp(B) – Lower	95 % C.I. for Exp(B) – Upper
Access	21.914	3.013	52.904	.000	3.29×10^9	8.97×10^6	1.21×10^{12}
Quality	12.562	1.892	44.089	.000	2.86×10^5	7.00×10^3	1.16×10^7
Occupation Self-Employed/Business	3.486	1.287	7.333	.007	32.640	2.619	406.787
Occupation Student	3.542	1.318	7.221	.007	34.521	2.608	456.981
Occupation Government Employee	4.011	1.871	4.595	.032	55.201	1.410	2161.296
Occupation Private Employee	3.197	1.259	6.443	.011	24.457	2.072	288.717
Income ₹20,001–50,000	1.268	0.617	4.227	.040	3.553	1.061	11.899

5.2.4. Interpretation

As shown in Table -7, logistic regression analysis revealed that Access, Quality, Occupation, and Income were the primary factors influencing financial inclusion among youth respondents in the Kutch district. The coefficient for Access ($B = 21.914$, $p < 0.001$) is positive and very significant. This means that being able to get to financial services more easily—like having bank branches nearby, ATMs available, and digital platforms—greatly increases the chances of being financially included. The odds ratio (Exp B = 3.29×10^9) shows that even small improvements in the availability of financial services can greatly raise the chances of inclusion.

The Quality dimension ($B = 12.562$, $p < 0.001$) also turned out to be a strong predictor. Youth who perceive elevated levels of service reliability, transparency, and customer satisfaction are approximately 2.86×10^5 times more likely to achieve financial inclusion. This discovery emphasizes that inclusion encompasses not only the availability of financial services but also trust, efficiency, and the quality of service provided by financial institutions.

Occupation exhibited a significant correlation with inclusion among demographic variables ($p < 0.05$). Compared to homemakers (the reference category), youth who are self-employed or running a business (Exp B = 32.64), students (Exp B = 34.52), government employees (Exp B = 55.20), and private employees (Exp B = 24.46) show much higher chances of being financially included. These results suggest that individuals involved in income-generating or formal activities interact more frequently with financial systems, thereby exhibiting higher levels of inclusion.

Income also has a positive and statistically significant effect. People whose monthly household income is between ₹ 20,001 and ₹ 50,000 ($B = 1.268$, $p = 0.04$) are 3.55 times more likely to be financially included than people whose monthly income is less than ₹ 10,000. This indicates that an increase in earning capacity correlates with heightened ability and motivation to utilize formal financial products.

The large coefficients and wide confidence intervals observed in the logistic regression are a known outcome of using min–max normalized predictors. When the range of an independent variable is compressed between 0 and 1, logistic regression estimates naturally produce larger coefficient magnitudes and inflated odds ratios, even though the direction, significance, and model fit remain unaffected (Hosmer et al., 2013; Menard, 2002). As recommended in logistic regression literature, coefficient interpretation should therefore be understood on the transformed scale rather than in absolute numerical terms. Importantly, the model shows excellent calibration and predictive accuracy, indicating that these large estimates do not reflect instability in the analysis.

Overall, the model shows that structural factors like Access and Quality (which are similar to the Reserve Bank of India's Financial Inclusion Index dimensions) and socio-economic factors like Occupation and Income have a big impact on the chances of young people being financially included. These results show that policies that make services easier to get to, improve the customer experience, and get young people involved in the economy can have a big impact on the overall financial inclusion landscape in Kutch.

5.3. Summary of key findings

Across both models, Access emerges as the strongest and most consistent determinant of financial inclusion among youth. Quality also contributes significantly to the logistic model. Occupation and household income enhance inclusion likelihood, though their effects are secondary. The CHAID model affirms Access as the primary segmentation variable, identifying clear youth clusters with low, moderate, and high inclusion probabilities. Together, these results show that structural access to financial services is the central driver of inclusion for young individuals in Kutch.

5.4. Comparison of logistic regression and CHAID models

Logistic regression identifies which variables significantly predict financial inclusion, whereas CHAID shows how different subgroups of youth are segmented based on these predictors. The logistic model highlights Access and Quality as the strongest positive predictors, supported by Occupation and household income. In contrast, the CHAID tree reveals that Access alone forms the primary split, with progressively higher Access scores corresponding to higher inclusion probabilities across distinct youth subgroups. Thus, logistic regression provides statistical significance and marginal effects at the variable level, while CHAID visually maps the hierarchy of predictors and identifies high-risk segments for targeted interventions.

6. Findings, Policy Implications & Future Scope

6.1. Findings

The study finds that Access and Quality, as defined in the RBI framework, are the dominant predictors of youth financial inclusion. Logistic regression demonstrates strong predictive power, with Access and Quality significantly improving the likelihood of inclusion. CHAID segmentation shows Access as the primary differentiator that separates high- and low-inclusion youth cohorts. Socio-economic variables such as occupation and household income also support inclusion but are less influential than structural access barriers. Overall, the results underline that youth exclusion is primarily due to a lack of availability of financial services and inconsistent service experience rather than behavioral unwillingness.

6.2. Policy implications

The findings indicate that strengthening Access and Quality should be the priority for policymakers and financial institutions. Efforts such as improving digital infrastructure, simplifying documentation, enhancing service responsiveness, and expanding outreach in semi-urban regions can substantially raise inclusion levels. Collaboration with FinTech platforms can reduce onboarding friction, support e-KYC, and provide youth-friendly micro-savings and micro-credit solutions. Educational institutions and local bodies can integrate financial education modules to support first-time users and sustain inclusion over time. Collaboration with FinTech firms presents a high-impact pathway to advance youth financial inclusion in India. With their strong digital presence, simplified onboarding processes, and data-driven customer profiling, FinTechs have become central players in engaging young users. Partnerships between banks and FinTech platforms can help reduce documentation barriers, deliver low-cost digital accounts, enable micro-savings and micro-credit products, and improve the overall service experience for youth. Integrating FinTech-led e-KYC, AI-based credit assessment, and gamified financial education tools can further support first-time users in entering and staying within the formal financial system. Strengthening such collaborations aligns with RBI's digital finance strategy and has the potential to bridge last-mile gaps, particularly for semi-urban and rural youth.

6.3. Future scope

Future research may explore longitudinal tracking of youth inclusion to understand transitions over time. Incorporating behavioral indicators (e.g., risk attitudes, digital financial habits) and psychometric variables may improve predictive accuracy. Applying advanced machine learning techniques—such as random forests, gradient boosting, or neural networks—could identify additional hidden patterns. Additionally, comparative studies across districts or states can strengthen generalizability and offer broader policy insights.

7. Conclusion

This study applied logistic regression and CHAID models to examine the determinants of financial inclusion among youth in Kutch. Both models point to the same core insight—Access and Quality are the strongest drivers of inclusion, consistently distinguishing included from excluded youth. Logistic regression showed high explanatory power (Nagelkerke $R^2 \approx 0.80$), while CHAID visualized clear segmentation patterns based on access levels. Socio-economic variables, particularly occupation and income, further enhanced inclusion likelihood.

Together, these results show that young people are not excluded due to a lack of interest but due to structural barriers, especially the limited availability of financial services and inconsistent service quality.

From a policy perspective, these results highlight the urgent need to strengthen banking outreach in semi-urban and rural pockets, simplify digital onboarding processes, and invest in customer-centric service models. Encouraging entrepreneurship and youth employment can further expand inclusion by enhancing financial capability and trust in formal systems. Thus, the combined application of logistic regression and CHAID analysis provides not only quantitative validation but also actionable insight—bridging the gap between financial policy and the real experiences of young people striving for inclusion in Kutch's evolving financial landscape.

References

[1] Allen, F., Demirgüç-Kunt, A., Klapper, L., & Martinez Peria, M. S. (2012). *The foundations of financial inclusion: Understanding ownership and use of formal accounts* (World Bank Policy Research Working Paper No. 6290). https://doi.org/10.1596/1813-9450-6290.

[2] Barajas, A., Beck, T., Belhaj, M., & Ben Naceur, S. (2020). *Financial inclusion: What have we learned so far? What do we have to learn?* (IMF Working Paper WP/20/157). https://doi.org/10.5089/9781513553009.001.

[3] Bazarbash, M. (2019). *FinTech in financial inclusion: Machine learning applications in assessing credit risk* (IMF Working Paper No. 2019/109). International Monetary Fund. https://doi.org/10.5089/9781498314428.001.

[4] CRISIL. (2013). *CRISIL Inclusix: An index to measure India's progress on financial inclusion.* CRISIL.

[5] Daida, S., & Kumar, K. N. (2024). Financial inclusion through AI: Predictive loan approvals for underbanked populations. *European Economic Letters, 14*(4), 1359–1372.

[6] Demirgüç-Kunt, A., Klapper, L., & Singer, D. (2017). *Financial inclusion and inclusive growth: A review of recent empirical evidence.* World Bank. http://econ.worldbank.org. https://doi.org/10.1596/1813-9450-8040.

[7] Demirgüç-Kunt, A., Klapper, L., Singer, D., & Ansar, S. (2021). *Financial inclusion, digital payments, and resilience in the age of COVID-19.* World Bank.

[8] Demirgüç-Kunt, A., Klapper, L., Singer, D., Ansar, S., & Hess, J. (2017). *The Global Findex Database 2017: Measuring financial inclusion and the fintech revolution.* World Bank. https://doi.org/10.1596/978-1-4648-1259-0.

[9] Elmasmari, H. (2024). Analysis of the impact of financial inclusion and FinTech on youth labour force participation: Evidence from the MENA region. *Economics and Business Review.* https://doi.org/10.18559/ebr.2024.4.1668.

[10] Food and Agriculture Organization. (2023). *Assessing the state of youth financial inclusion in developing contexts.* FAO.

[11] Hersh, J. (2021). Where are the unbanked in Belize? Using machine learning–based small area estimation to improve financial inclusion. *Inter-American Development Bank Working Paper Series.* https://doi.org/10.18235/0003381.

[12] Hosmer, D. W., Lemeshow, S., & Sturdivant, M. (2013). *Applied logistic regression* (3rd ed.). Wiley. https://doi.org/10.1002/9781118548387.

[13] Ikosa, S. I. (2025). *Determinants of financial inclusion among the youth in Namibia* (Master's thesis). University of Namibia.

[14] Isaga, N. (2025). Determinants of Tanzanian youth entrepreneurs' financial inclusion. *African Journal of Economic Review.*

[15] Klapper, L., Singer, D., Starita, L., & Norris, A. (2025). *The Global Findex Database 2025: Connectivity and financial inclusion in the digital economy.* World Bank. https://doi.org/10.1596/978-1-4648-2204-9.

[16] Maehara, R. (2024). Predicting financial inclusion in Peru using machine learning models. *Journal of Risk and Financial Management, 17*(1), 34. https://doi.org/10.3390/jrfm17010034.

[17] Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). Sage. https://doi.org/10.4135/9781412983433.

[18] OECD. OECD. (2023). OECD/INFE 2023 international survey of adult financial literacy. https://www.oecd.org/financial/education/oecd-infe-2023-international-survey-adult-financial-literacy.pdf

[19] Pandey, A., Chauhan, Y., & Singh, R. (2025). Examining financial inclusion success in North India via usage, digitization, and socio-demographic factors. *Future Business Journal, 11*(1), 1–10. https://doi.org/10.1016/j.joitmc.2025.100542.

[20] Partridge, A., Morongwa, C. H., & Cenfri. (2022). *Consumer protection in digital finance.* www.cenfri.org

[21] RBI. (2021). *Financial Inclusion Index for India* (September Issue). Reserve Bank of India.

[22] Sarma, M., & Pais, J. (2011). Financial inclusion and development. *Journal of International Development, 23*(5), 613–628. https://doi.org/10.1002/jid.1698.

[23] Shahin Shahin Sultana, A. (n.d.). A critical exploration of the National Youth Policy of India – 2003 and 2014. https://www.youthpolicy.org/national_youth_policies/india_2003_2014_analysis.pdf

[24] Singh, R., & Roy, S. (2015). Financial inclusion: A critical assessment of its concepts and measurement. *Asian Journal of Research in Business Economics and Management, 5*(1), 12–24. https://doi.org/10.5958/2249-7307.2015.00002.X.

[25] Sreeram, V., & Kumar, M. (2024). AI-enabled alternative data credit scoring for financial inclusion: A predictive modelling approach. *International Journal of Intelligent Finance and Informatics.*