Original Researcher Article

Predicting Financial Inclusion with Machine Learning: Insights from Rural Area of Sambalpur and Bargarh District

Soni Dubey*[1], Tushar Kanti Das[2], Monalisha Pattnaik[3], Abinash Kar[4]

*[1]Dept. of Business Administration Sambalpur University India-768019 https://orcid.org/0009-0005- 0692-1921

[2]Dept. of Business Administration Sambalpur University India-768019 https://orcid.org/0000-0002-9039-3236

[3]Dept. of Statistics Sambalpur University India-768019
https://orcid.org/0000-0002-6528-2294

[4]Sambalpur University India-768019

Corresponding author:

Soni Dubey

Email:ksony948@suniv.ac.in..

ABSTRACT

Purpose - This paper explores how financial inclusion depends on demographic factors such as, age, gender, education, occupation and income, in the Sambalpur and Bargarh districts of Odisha, India, and uses the latest machine learning to study the subject.

Design/Methodology/approach –The multistage sampling was conducted on a population of 813 respondents, of which 70 percent were used as a training sample, and 30 percent as a testing sample. Different models were used to analyze the data, among them are "decision trees, artificial neural networks (ANN), regression trees, regression model and deep neural networks (DNN)". "Root Mean squared error (RMSE) and Mean Absolute errors (MAE)" were used as measurements of performance.

Findings - Second ANN model had high performance in forecasting financial inclusion as the RMSE and MAE of the training and testing data are lower. Moreover, the 1st, 3rd, and 6th DNN models were very accurate in boosting financial inclusion prediction. The analysis of the decision tree showed the high income is a very strong predictor of financial inclusion, but other attributes like age, occupation, and gender complicate the situation with the majority of the population. The regression tree also found the income as the strongest variable, whereas the relationships were more nuanced, including the age, occupation, and education variables, on the lower-income groups.

Originality/Value - The reading has added worth to current works on applying AI-based approaches to forecasting financial inclusion as it reveals its possible usage in other geographical markets..

Keywords: Financial inclusion, AI-driven models, Western Odisha, RMSE, MAE.

INTRODUCTION:

The cross point of demographic factors and financial inclusion in the rural society is a vital frontier in the study of economic development. The study investigates the relationship between salary, gender, age, education, occupation and Income on Financial inclusion rate in rural areas of Sambalpur and Bargarh by using an analytical approach that is comprehensive. The study should help identify subtle trends and associations that may be leveraged to implement specific interventions to improve the financial accessibility of underserved communities through the application of advanced "statistical and machine learning" methods such as "decision trees, regression trees, artificial neural networks (ANN), deep neural networks (DNN), and regression models".

A number of theories have informed the studies on the influence of demographic factors on "financial inclusion" in rural settings. It was postulated by the life cycle hypothesis that maximize the accumulation and expenditure habits of people variably in various stages of their lives impacted their interest in using financial services (Modigliani and Brumberg, 1954). The theory of planned behavior, proposed by Ajzen, sheds light on the impact of attitudes, subjective norms, and perceived control over behavior on financial decision-making (Ajzen, 1991). Also, the theory of diffusion of innovations by Rogers provided the conceptual model of explaining the rise of using financial technologies and services in villages (Rogers, 2003). The capability approach which was introduced by Sen underlined the importance of personal capabilities such as education and income in creating more financial liberties (Sen, 1999). The theoretical basis for the analysis of the interactions of many variables associated with the demographic factors and financial inclusion within the rural setting of the Sambalpur and Bargarh districts was represented by these theories along with such advanced methods of analysis like the "decision trees, regression trees, artificial neural networks (ANN), deep neural networks (DNN) and regression models".

The demographic variables affecting the concept of financial inclusion in rural India comprise age, gender,

education, occupation, and income and the studies indicate that these factors influence the access to and utilization of financial services significantly (Sharma and Jain, 2020). The modern tendencies reveal that the "digital financial inclusion" is being simplified by the government through Digital India and PMJDY, and the introduction of new possibilities to review the intricate socio-economic trends is made by the implementation of the latest models, in particular decision trees, regression trees, ANN, and DNN (Verma et al., 2021; Kaur and Gupta, 2020). Nonetheless, the problems like the presence of chronic gender inequality, in which women have little economic freedom (Singh and Naik, 2021), and low literacy, which limits the quality of digital literacy and technology connectivity (Chaudhary and Kumar, 2020), remain to be potential barriers to improvement. Income disparity is another fact that hinders rural populations to access formal financial services (Bapat, 2019), and the lack of possibilities to occupy jobs between agricultural and non-agricultural workers leads to the existence of unmet financial needs (Mishra, 2021). One of the challenges is to bridge the digital gap because in rural areas, there is a problem in the use of financial technology due to poor internet infrastructure and digital illiteracy (Rai & Nath, 2022). Besides, data-driven models such as ANN and DNN are limited to the challenge of gathering big and trustworthy data in rural areas (Kaur and Gupta, 2020).

The study will examine how the demographic factors (age, gender, education, occupation, and income) would contribute to financial inclusion in Sambalpur and Bargarh districts rural regions. Although rural areas are facing increased attention to include digital financial services and thus data-driven policymaking, there are still major problems in rural areas such as access to formal institutions, financial literacy, and technological factors. The proposed research aims to reveal the non-obvious interdependences between demographic variables and financial inclusion by utilizing such advanced methods of analysis as "Decision Trees, Regression Trees, Artificial Neural Networks (ANN), Deep Neural Networks (DNN) and Regression Models". Results will be used to come up with specific solutions in improving financial issues and literacy in rural communities, reducing gender gaps, and closing the digital divide. The study is of great importance to the present scenario in India where the country is currently planning to encouragement of inclusive development and financial empowerment of rural areas.

Rest of research paper is systematized as: In Section: 2, there is summary of important literature. Section 3 discusses techniques utilized, including datasets, metrics for performance evaluation, and a detailed look at techniques such as "Decision tree", "Regression tree", "ANN", "DNN" and "Regression models". Section 4 details the experimental findings and the results for each model. In conclusion, Section 5 offers a comprehensive analysis of these results, their practical applications, and final observations.

Related Work

Demographic variables significantly influence financial inclusion across various contexts. Higher education, professional activity, and age (25+) correlate positively with account ownership, while gender showed no significant differentiation in some studies (Waliszewski, 2023). Income levels impact financial inclusion, particularly in certain European countries (Waliszewski, 2023). In South Asia, age, gender, education, income, and workforce participation significantly influence financial inclusion, with higher income and education levels increasing the likelihood of accessing formal financial services (Khushboo and Pradhan, 2024). In Türkiye, education is the strongest determinant of Digital Financial Inclusion (DFI), with women being more financially excluded than men (Güz and Poyraz, 2024). Studies in India reveal that age, education, marital status, occupation, dependents, and income significantly influence women's awareness of government schemes for financial inclusion (Kaviyarasu V and Siddiq, 2024), while among university lecturers in Indonesia, only income and financial literacy significantly affected financial inclusion (Soejono and Mendari, 2022). Research on Scheduled Tribes in Telangana State, India, found that demographic variables significantly impact various dimensions of financial inclusion (Nandru et al., 2021). Property ownership positively influences financial inclusion, while lower socio-economic status, illiteracy, unemployment, and homemaking negatively affect it (Behera, 2024). In Pakistan, females face reduced likelihood of financial inclusion, while higher age, income, and education increase inclusion (Bashir, 2023). Kenyan studies show that being male, young, educated, employed, and wealthy enhances accessibility to digital financial services (Mose et al., 2024). Studies in D-8 countries reveal that gender, education, and income status/working status has a great impact on financial inclusion (Wardani et al., 2023). Several studies include several demographic variables as the determinant of increased digital financial literacy and inclusion (Ocharave and Iworiso, 2024; Pattnayak and Sahoo, 2024; Kaur and Bansal, 2024). Financial inclusion is more affected by income and education whereas there is no linear relationship between age and financial inclusion indicators (Kumar and Pradhan, 2024). Financial literacy also depends on demographic factors, in which the social-economic factor is paramount (Nursjanti and Amaliawiati, 2024). Lastly, the research demonstrates positive correlations amongst financial inclusion and education, income, and employment and nonlinear correlation to age and differences in access to financial technologies by education and income level (Debuque-Gonzales and Corpus, 2024).

Machine learning (ML) is now being applied to financial services to improve credit scoring, risk assessment, fraud detection, customer segmentation and personalization, in underserved populations specifically. Recent innovations take advantage both of alternative data sources, like transactional histories, mobile use, and social media activity, as well as of algorithms, e.g., decision trees, random forests, gradient boosting, LightGBM and neural networks, which are much more effective than the traditional credit score systems, and produce high accuracy rates, e.g., LightGBM with 89.6 percent in credit scoring and Random Forests with 87.6 percent in fraud detection (Kiralioglu, 2024; Shak et al., 2024). Such

innovations enhance scalability, efficiency, and access to loans, enabling microfinance institutions, digital banks, and other alternative lenders to lend and provide customized services, overcoming the issue of data privacy, algorithmic bias, and compliance with guidelines (Shak et al., 2024; Das, 2024; Agu et al., 2024; Lee, 2024). In India, ML-driven solutions such as automated KYC, fraud detection, and digital payment systems—supported by initiatives like IndiaStack—have helped bridge the $380 billion credit gap, enabling platforms like Capital Float and Aye Finance to expand access for rural populations, small-scale farmers, and entrepreneurs (Das, 2024; Gupta, 2022). Beyond credit scoring, ML also supports predictive analytics, natural language processing, and personalization of financial products, which enhance customer interaction, financial literacy, and ethical decision-making, especially in emerging markets such as the United States and South Asia (Omogbeme et al., 2024; Amarnath and Thimothy, 2024). Additionally, ML optimizes investment strategies, anticipates market trends, and improves client experiences through virtual assistants and personalized recommendations, ultimately strengthening financial inclusion and economic growth (Varadarajan, 2022).

Demographic and sociodemographic characteristics are consistently shown to shape financial inclusion across regions, with studies employing both machine learning and traditional statistical models to analyze these determinants. In Mexico, education, income, savings capacity, future-oriented preferences, and smartphone access were identified as key predictors of financial inclusion, suggesting policy opportunities to expand formal financial services (Rioja and Valdez, 2024). Similar findings in Peru, where decision trees and random forests highlighted demographic variables as strong predictors, demonstrated the usefulness of machine learning in uncovering patterns relevant for underserved rural areas (Maehara et al., 2024). In Africa, age, phone ownership, and household size significantly influenced account ownership, with logistic regression outperforming random forest models (Li, 2022). Research in Eswatini, Rwanda, and Madagascar also confirmed the influence of demographics alongside service quality and remittances, with Gaussian Naive Bayes models enhancing prediction (Dlamini et al., 2023). In Malaysia, income, employment, and education determined insurance uptake, with decision tree and random forest models pointing to opportunities for tailoring Takaful products to low-income and unemployed groups (Lee et al., 2023). Evidence from rural China emphasized household size, education of the household head, income, employment, and financial literacy as critical drivers of inclusion, with probit and OLS models revealing significant regional disparities (Jin et al., 2024). Similarly, in Zimbabwe, household size, off-farm income, and access to agricultural extension services emerged as significant determinants, with logistic regression results highlighting necessity to diminish transaction expenses and progress rural access to financial services (Mhlanga, 2022).

Research Methodology

An extensive literature review was conducted, analyzing studies, articles, and reports on effect of socio-economic characteristics on "financial inclusion". Databases such as Google Scholar, ProQuest, ResearchGate, and ScienceDirect were utilized as primary sources. Subsequently, a survey instrument was developed and validated by experts. Upon approval, it was distributed to collect the perspectives and experiences of rural inhabitants. The data collected underwent analysis through the application of descriptive techniques and advanced statistical methods to identify trends and correlations. Various AI-driven machine learning models, including "Regression Tree," "Decision Tree," "ANN," and "DNN," were employed for financial inclusion datasets. The findings suggest that both the ANN and Regression models for financial inclusion outperform all other models, respectively.

Data and Preliminary Analysis

This paper is concentrated in the Sambalpur and Bargarh District of Western Odisha, two developing region within the state. A substantial percentage of the population in Sambalpur & Bargarh remains underdeveloped and requires integration into inclusive financial systems. Numerous studies have endeavored to assess and quantify financial inclusion in "India". This study's objective is to evaluate the influence of socio-economic characteristics on "financial inclusion" & to identify causal factors. Data were collected through in-person surveys conducted in the rural areas of Sambalpur and Bargarh District. The survey comprised two sections: demographic and general information, and participants' understanding and experiences with financial inclusion. The rural population was the primary focus, identified through visits to the District Rural Development Agency, Odisha Rural Development and Marketing Society, and Mission Shakti Office in Sambalpur and Bargarh. Employing a multistage sampling technique, 813 respondents were selected from the districts. The dataset is partitioned such that 70% is utilized for the purpose of training, while the rest 30% is reserved for the purpose of testing all the models. Figure: 1 demonstrates the current study area within the Sambalpur District, and Figure 2 depicts the studied blocks and villages within the district.

Source: https://www.mapchart.net/india.html
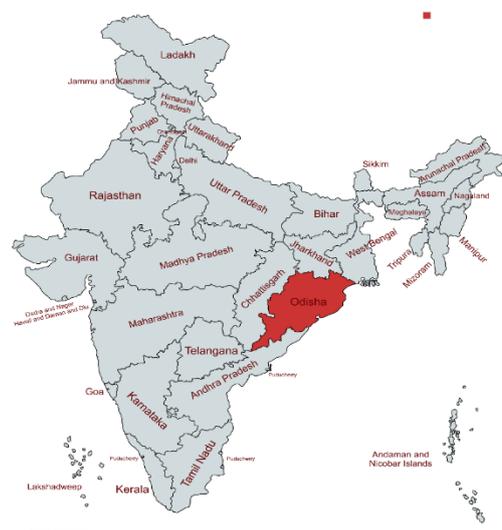
Figure 1: Odisha State in India

Figure 2: Sambalpur and Bargarh District in Odisha

Table I: Descriptive Statistics of the Demographic Variables, Financial Inclusion and its Variables and Social, Economic and Human development of Sambalpur and Bargarh District

Sl. No.	Variables	Variable types	Minimum	Q1	Median	Mean	Q3	Standard Deviation	Maximum	Skewness	Kurtosis
1	Gender	Int	1.000	1.000	1.000	1.456	2.000	0.4983963	2.000	0.1753316	1.030741
2	Age	Int	18.00	26.00	35.00	37.73	46.00	13.438124 3	73.00	0.4723166	2.515677
3	Education	Int	1.000	2.000	2.000	1.998	2.000	0.7191907	4.000	0.2024032	2.491535
4	Occupation	Int	1.000	2.000	4.000	4.114	7.000	2.3437270	8.000	0.1218581	1.438205
5	Income	Int	1.000	1.000	1.000	1.565	2.000	0.9622555	5.000	1.9158801	6.175845
6	Fin_Aware	Num	1.000	1.600	1.800	2.033	2.400	0.5688829	3.750	0.8459534	3.292836
7	Fin_Access	Num	1.300	2.200	2.200	2.488	3.000	0.6207407	4.000	0.5000194	2.630453
8	Fin_Avail	Num	1.440	2.000	2.210	2.179	2.330	0.2923457	2.830	-0.4090045	3.527118

9	Fin_Usab	Num	1.110	1.670	1.830	2.162	2.780	0.6869060	3.670	0.6637453	2.311673
10	Eco_develop	Num	2.080	2.330	2.750	2.699	2.920	0.4241293	3.580	0.5208096	2.263392
11	Soc_develop	Num	1.500	2.000	2.000	2.244	2.400	0.5012864	3.500	0.9852398	3.434391
12	Hum_develop	Num	1.430	2.210	2.500	2.458	2.570	0.4342077	3.640	0.6237555	4.846443
13	Fin_Inclu	Num	1.280	1.880	2.010	2.175	2.580	0.4430943	3.130	0.2409294	2.462347

Performance Assessment Metrics

"Root Mean Square Error (RMSE)" & "Mean Absolute Error (MAE)" are applied to measure usefulness of diverse predicting frameworks on financial inclusion datasets (Chakraborty et al., 2020). Equations for 2 "performance assessment metrics" are delineated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2} \text{ and } MAE = \frac{\sum_{i=1}^{n}|x_i - \hat{x}_i|}{n}$$

The quantity of data points is represented by n, with x_i denoting the true value and $\hat{x}_i$ indicating the estimated value. A lower value of the performance metrics indicates a more effective model.

Artificial Neural Networks

"Artificial Neural Networks (ANNs)" are computational frameworks that draw inspiration from the structure of biological neural networks (Roheen Qamar and Baqar Ali Zardari, 2023). These networks are composed of interconnected nodes arranged in layers, which process input data through intermediate layers to generate an output (Kumar and Thakur, 2012). ANNs are trained to change the connection strengths based on erroneous prediction performances by backpropagation (Huang, 2009). These networks are ideal at recognizing patterns, classifying, and predicting them, so they can be useful in several spheres, including image recognition, natural language processing, and financial prognostication (Schmidgall et al., 2024).

Regression trees

Regression trees are non-parametric models that are applied to predict continuous consequences. They divide the feature space into sub-sets recursively, and, as a result, generate a tree-like representation with every leaf node denoting a predicted result (Gama et al., 2005). The algorithm then uses the optimal split points and removes the contribution of the squared error of each subset (Loh, 2011). Regression trees are also interpretable and non-linear with the ability to model complex interactions among variables hence can be used during exploratory analysis and predictive modeling processes in numerous

research disciplines (Buja and Lee, 2001; Jena and Dehuri, 2020).

Decision trees

"Decision trees" are categorical models that are applied to classification & regressions tasks (Abedinia and Seydi, 2024). They divide the feature space into areas through recursive binary divisions forming a tree form (Quan and Valdez, 2018). Internal nodes are the decision rule, and the leaf nodes hold the forecasted result (Dai et al., 2016). Decision trees can be readily interpreted, are used to capture non-linear relationships, and can capture sophisticated interactions between variables, thus they are useful in both the exploration of research data and prediction of research outcomes within methodology (Chandrasekar et al., 2017).

Deep Neural Networks

"Deep Neural Networks (DNNs)" are complex "machine learning" architectures consisting of several layers of networks. The reason is that they do well at learning complicated patterns involving big datasets, which is why they are applicable to different tasks, such as image recognition and natural language processing (Chung et al., 2017; Fan et al., 2021). The hierarchical features can also be identified automatically through DNNs and manual feature engineering is not required. They are effective methods of research analysis because of their capacity to extract complex relationships in their data (Cai et al., 2022; Sze et al., 2020).

Regression models

Regression models are used to study relationships between dependent and independent variables and approximate the effects and make forecasts. Machine learning applications augment the standard regression, and they incorporate regularization, feature selection, and cross-validation (Lewis et al., 2023; Osman et al., 2017; Yang et al., 2023). These are utilized to enhance model performance, to work with high dimensional data and to reduce over fitting.

Experimental Advances and Findings of AI-led Financial Inclusion

The way financial inclusion datasets are characterized and defined is a significant challenging task. In the Data sets of financial Inclusion of the two Districts lightly different like Sambalpur and Bargarh investigated. The current study aims at changing data-based prediction of inclusion of financial and banking services at Sambalpur and Bargarh district in relation to the demographical variables in terms of age, gender, education, occupation, income and source of income with the help of Decision tree model, Regression tree model, ANN model and DNN model. Table 1 presents the description of the various components in regards to financial inclusion of Sambalpur and Bargarh district.
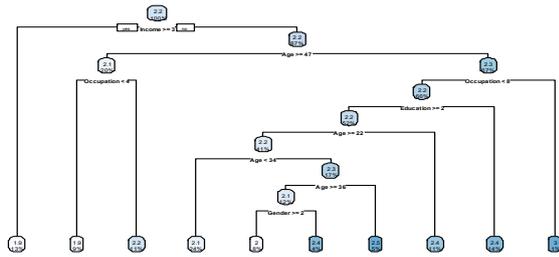
Figure 3: Decision Tree for Prediction of Financial Inclusion

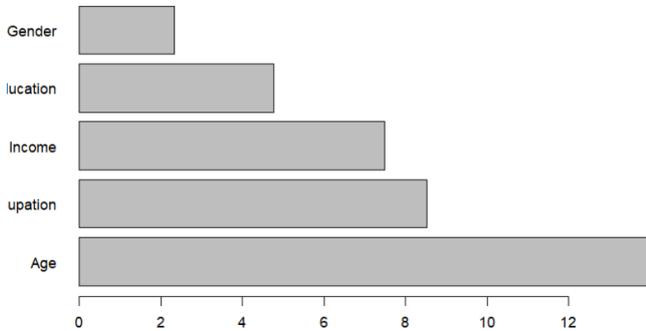*Decision Tree Model for Financial Inclusion*

The "Decision Tree (DT) models" are effective classifiers since they are based on a tree-like format in defining the relationship among features and possible results. When one comes to the last decision, the tree ends in leaf nodes otherwise known as terminal nodes which reveal the action according to the path that was taken in decision-making process. In predictive modeling, leaf nodes signify the expected results derived from the progression of events traversing tree. Choice to utilize a "decision tree model" for pinpointing crucial input variables from five demographic factors, along with financial inclusion in the rural regions of Sambalpur and Bargarh districts, was driven by framework's straightforwardness, ease of interpretation, & high level of precision. Optimal "decision tree model" was employed on dataset containing 813 regular records to recognize potential causal variables that might predict inclusion levels in Sambalpur and Bargarh. "Decision tree" analysis was conducted using "rpart" package in R software, with the control parameter "minsplit" set at ten percent of total data. Framework's predictive accuracy was assessed utilizing "RMSE". Most effective "decision tree" was constructed using five input variables with a "minsplit" of five, assigning equal costs to each variable. Figure 4 presents the list of variable importance, while Figure 3 illustrates the fitted tree. The analysis using the decision tree identified that, among the five potential input variables, three were deemed highly significant.

The utilization of the "rpart" package facilitates the examination of various levels of cost-complexity. To evaluate the errors at each stage, "rpart" performs a ten-fold cross-validation by determining errors on validation dataset. Figure 3 illustrates the optimal "decision tree", which consists of 9 internal nodes & 10 terminal nodes, & is built using five variables for its model. By setting the complexity parameter (cp) to zero, this tree can be expanded into full tree with 10 terminal nodes, as shown in Figure 3. In Figure 4, y-axis represents cross-validation error, while lower x-axis shows values related to cost complexity, & upper x-axis indicates number of terminal nodes. As tree grows beyond 10 terminal nodes, decrease in errors becomes less significant due to increased depth. In predicting financial inclusion in the Sambalpur and Bargarh Districts, the decision tree identified five significant variables from the total of five available. Each of the 813 daily data points from the five demographic factors is evaluated using the decision tree shown in Figure 3. Within this tree, each data point is assessed at a node, proceeding left side for a "Yes" response or right side for a "No" response.

If the population income is 1,00,001 or more, than the Financial Inclusion (FI) is quantified at 1.9, with a probability of occurrence at 13%. If the population income is less than 1,00,001, age is more than equal to 47 and occupation is student, homemaker and farmer than the Financial Inclusion (FI) is quantified at 1.9, with a probability of occurrence at 9%, but if occupation is daily wage worker, private employee, govt. employee, self-employed, others than the Financial Inclusion (FI) is quantified at 2.2, with a probability of occurrence at 11%. If the population income is less than 1,00,001, age is between 22 to 34, occupation is student, homemaker, farmer, daily wage worker, private employee, govt. employee, and self-employed, education is matriculate, graduation and post-graduation than the Financial Inclusion (FI) is quantified at 2.1, with a probability of occurrence at 24%. If the population income is less than 1,00,001, age is between 36 to 46, occupation is student, homemaker, farmer, daily wage worker, private employee, govt. employee, and self-employed, education is matriculate, graduation and post-graduation, gender is female than the Financial Inclusion (FI) is quantified at 2, with a probability of occurrence at 8%, but if gender is male than the Financial Inclusion (FI) is quantified at 2.4, with a probability of occurrence at 4%. If the population income is less than 1,00,001, age is between 34 to 35, occupation is student, homemaker, farmer, daily wage worker, private employee, govt. employee, and self-employed, education is matriculate, graduation and post-graduation than the Financial Inclusion (FI) is quantified at 2.5, with a probability of occurrence at 5%. If the population income is less than 1,00,001, age is less than 22, occupation is student, homemaker, farmer, daily wage worker, private employee, govt. employee, and self-employed, education is matriculate, graduation and post-graduation than the Financial Inclusion (FI) is quantified at 2.4, with a probability of occurrence at 11%. If the population income is less than 1,00,001, age is less than 47, occupation is student, homemaker, farmer, daily wage worker, private employee, govt. employee, and self-employed, education is illiterate than the Financial Inclusion (FI) is quantified at 2.4, with a probability of occurrence at 14%. If the population income is less than 1,00,001, age is less than 47, occupation is others than the Financial Inclusion (FI) is quantified at 3, with a

probability of occurrence at 1%. In conclusion, while high income is a straightforward predictor of a high FI score, financial inclusion for the majority of the population is a more complex issue influenced by a combination of age, occupation, and gender. The model identifies specific demographic segments with distinct FI scores.

Figure 4: Variable Importance for Prediction of Financial Inclusion
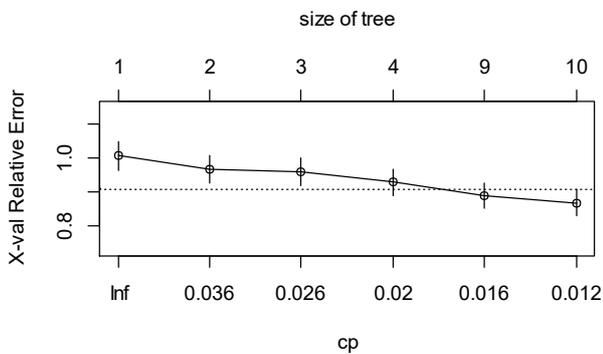
**Figure 5**: Complex Parameter for Prediction of Financial Inclusion
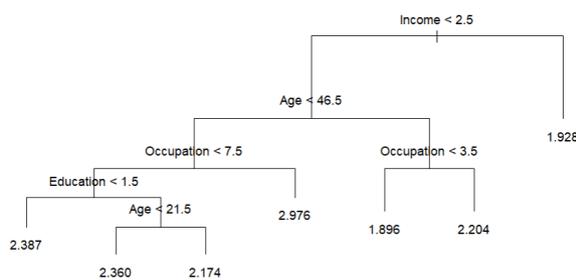
Figure 6: Regression tree for Prediction of Financial Inclusion

Regression Tree Model for Financial Inclusion

The rpart algorithm is utilized to assess the spectrum of cost complexity. To evaluate the error linked to each cost complexity value, rpart performs 10-fold cross-validation, calculating error on reserved validation data. Figure 6 shows a "regression tree" with six internal nodes, leading to seven terminal nodes. This tree essentially divides based on five determinants to form its framework. By setting the "Complexity Parameter" (CP) to 0, a tree with

seven terminal nodes can be long-drawn-out into full tree (Figure 5). In Figure 5, the y-axis shows the cross-validation error, the lower x-axis represents the cost complexity value, and the upper x-axis indicates number of terminal nodes. Beyond seven terminal nodes, there is a noticeable decline in the reduction of error as the tree becomes deeper.

To forecast the financial inclusion of the Sambalpur and Bargarh districts in Odisha, all five variables are processed through a Regression Tree (Figure 6). At each node, the variables are evaluated, and the path proceeds to the left if the response is "yes" or to the right if the response is "no." If the population's income is less than 50,001, age is less than 46.5 years, and occupation is categorized as student, homemaker, farmer, daily wage worker, private employee, or government employee, and education level is illiterate, then the Financial Inclusion (FI) is quantified at 2.387. If the population's income is less than 50,001, occupation is student, homemaker, farmer, daily wage worker, private employee, or government employee, education level is matriculate, graduate, or post-graduate, and age is less than 21.5 years, then the FI is quantified at 2.360. However, if the population's age is between 21.5 and 46.5 years, the FI is quantified at 2.174. If the population's income is less than 50,001, age is less than 46.5 years, and occupation is self-employed or other, then the FI is quantified at 2.976. If the population's income is less than 50,001, age is equal to or greater than 46.5 years, and occupation is student or homemaker, then the FI is quantified at 1.896. However, if the occupation is farmer, daily wage worker, private employee, government employee, self-employed, or other, then the FI is quantified at 2.204. Conversely, if the population's income is 50,001 or more, the FI is quantified at 1.928. The regression tree identifies income as the most influential variable, as an income of 50,001 or more leads to a direct prediction. For lower incomes, the model employs a more complex set of rules involving age, occupation, and education to determine the financial inclusion score, indicating a more nuanced relationship with these variables.
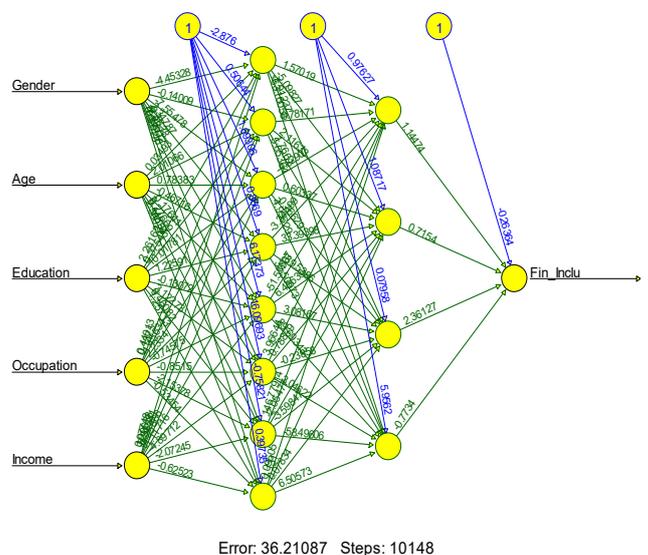
Figure 7: ANN with Two Hidden Layer Model to Predict Average Financial Inclusion of Bargarh & Sambalpur District

"Artificial Neural Network" & "Regression Models" for Financial Inclusion Prediction

Research uses Artificial Neural Networks (ANN) and regression models to examine impact of the demographic variables, such as age, gender, education, occupation, & income, on the "financial inclusion" in Sambalpur & Bargarh district. Financial inclusion prediction of population of Sambalpur and Bargarh district is analyzed. The initial artificial neural network (ANN) model processes five demographic inputs and is structured as a three-layer feed-forward neural network, featuring a hidden layer with eight neurons and a single output neuron. Similarly, the second "ANN model" also incorporates five input variables and is designed as four-layer feed-forward neural network. This configuration includes 2 hidden layers, containing four & eight neurons,

respectively, along with 1 output neuron. The third "ANN model", which also utilizes five input variables, is arranged as four-layer feed-forward neural network. This model comprises 2 hidden layers with four & two neurons, respectively, & a single output neuron. Likewise, the fourth "ANN model" uses five input variables and it is set as three-layered feed forward neural network, the model has only one hidden layers with four neurons each, and a single output neuron. One regression model is developed to predict financial inclusion at Sambalpur and Bargarh district. In this "regression model", all 5 demographic variables are taken. Table 2 presents the notations used in different AI-driven models. The study focuses on financial inclusion prediction for Sambalpur and Bargarh district of Western Odisha. In Equations 1, 2, 3 and 4, the ANN models are presented.

Table II: Notations used in Different Prediction Models

Terms		Description
Age	:	Age of the participants
Gen	:	Gender of the participants
Edu	:	Educational qualification of the participants
Occup	:	Occupation of the participants
Inc	:	Annual income of the participants
FinInclu	:	Delivery of appropriate financial products & services at an affordable cost – on timely basis to vulnerable groups (Financial Inclusion)
α	:	Constant in "regression model"
β_i	:	Coefficients in "regression model" (i=1, 2......k)
ε	:	Random error
$FinInclu_{(ANN,8)}$	:	A three-layer feedforward neural network comprising five input layer and hidden layer with 8 neurons, & single target of *Financial Inclusion*
$FinInclu_{(ANN,8-4)}$	:	A four-layer feedforward neural network comprises five input layer, two hidden layers with 8 & 4 neurons respectively, & single output target of *Financial Inclusion*
$FinInclu_{(ANN,4-2)}$	:	A four-layer feedforward neural network comprising five input layer and two hidden layer with 4 & 2 neurons, and single target of *Financial Inclusion*
$FinInclu_{(ANN,4)}$	:	A three-layer feedforward neural network comprises five input layer, hidden layers with 4 neurons respectively, and a single output target of *Financial Inclusion*

"ANN and the regression models" are advanced to inspect effect of demographic variables namely age, gender, education, occupation, & income on financial inclusion at Sambalpur and Bargarh district. The analysis is conducted for financial inclusion prediction of population of Sambalpur and Bargarh district of Western Odisha. The dataset, comprising 813 data points, is partitioned in a 70:30 ratio, with 70% designated for training the network & the remaining 30% for testing purposes. Four models of

"artificial neural networks (ANN)" are established. The first framework features a hidden layer with eight neurons, while second model includes 2 hidden layers with eight & four neurons, respectively. Additionally, first model has two hidden layers with 4 and 2 neurons each. Both models utilize five input variables. In the initial "artificial neural network (ANN)" model, the five demographic variables are processed through three-layer feed-forward neural network, which comprises eight

neurons in hidden layer & a single output neuron. Likewise, 2nd "ANN model" also employs five input variables and is configured as four-layer feed-forward neural network, incorporating hidden layers with eight & four neurons, along with 1 output neuron. The 3rd "ANN model", using five input variables, is structured as four-layer feed-forward neural network, featuring two hidden layers with four & two neurons, respectively, & 1 output neuron. Likewise, the fourth "ANN model" uses five input variables and it is set as three-layers feed forward neural network, that is, its setting consists of four neurons in one hidden layer, each with four neurons, respectively, and one output neuron. A single regression equation is created to forecast the financial Inclusion in the Sambalpur district. All the 5 demographic variables are taken in this regression model. Figures 7 depict the ANN models to forecast the financial Inclusion at Sambalpur and Bargarh district. Equations 1 and 2 illustrate structure of four "ANN models" with either 1 or 2 hidden layers. Table 3 presents accuracy of both training & testing datasets for these four ANN models, evaluated using RMSE and MAE. The data indicates that 2nd "ANN model" achieves lower "RMSE and MAE" values for both training and testing datasets compared to other models. Specifically, 2nd "ANN model" demonstrates superior "RMSE and MAE" performance, with values of [0.3748397, 0.3046555] for the training data and [0.3800821, 0.3080368] for the testing data, in predicting financial inclusion in the Sambalpur and Bargarh districts.

$$FinInclu_{(ANN,8)} = f(Age, Gen, Edu, Occup, Inc) \quad (1)$$

$$FinInclu_{(ANN,8-4)} = f(Age, Gen, Edu, Occup, Inc) \quad (2)$$

$$FinInclu_{(ANN,4-2)} = f(Age, Gen, Edu, Occup, Inc) \quad (3)$$

$$FinInclu_{(ANN,4)} = f(Age, Gen, Edu, Occup, Inc) \quad (4)$$

The regression model is formulated to examine how demography (age, gender, education, occupation and income) affect the financial Inclusion within Sambalpur and Bargarh district. The data is divided 70:30, 70 percent serving as the training data to the network and the rest 30 percent is used as the testing data. The regression model will have a structure as shown on equation 5. Table 3 represents the levels of accuracy of the training data and the testing data with "RMSE and MAE" of the "regression model". The research indicates that "RMSE and MAE" values of the training data [0.4110656, 0.430545] and testing data [0.3466942, 0.3607383] are obtained in explaining the financial Inclusion in Sambalpur district when the regression model is used.

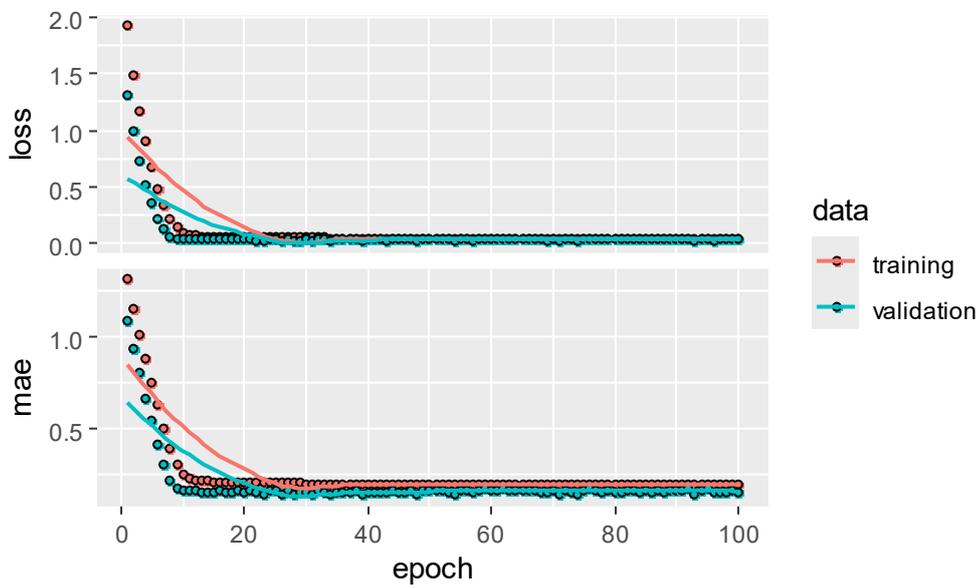$$FinInclu = \alpha + \beta_1 Age + \beta_2 Gen + \beta_3 Edu + \beta_4 Occup + \beta_5 Inc + \varepsilon \quad (5)$$

Figure 8: DNN with One Single Hidden Layer Model to Predict Financial Inclusion
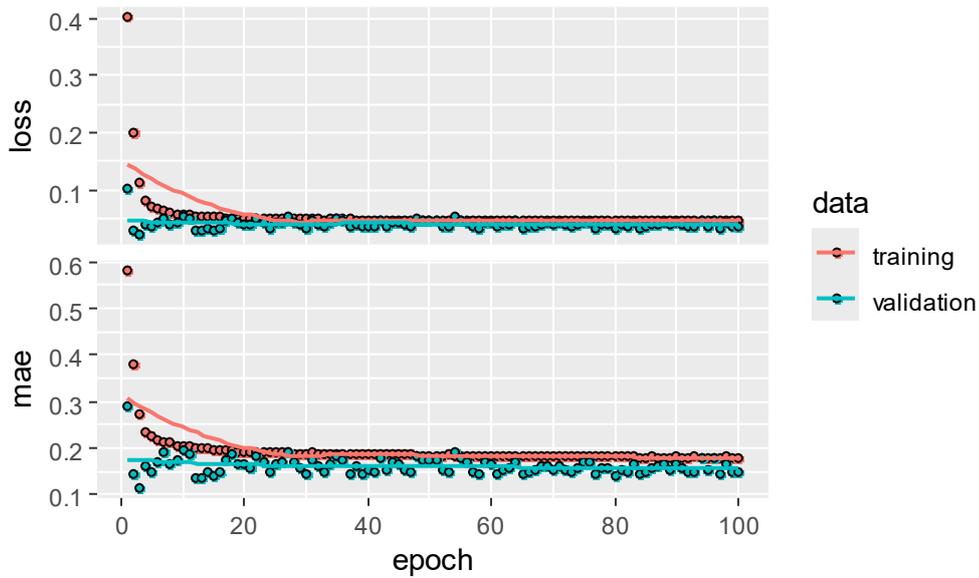
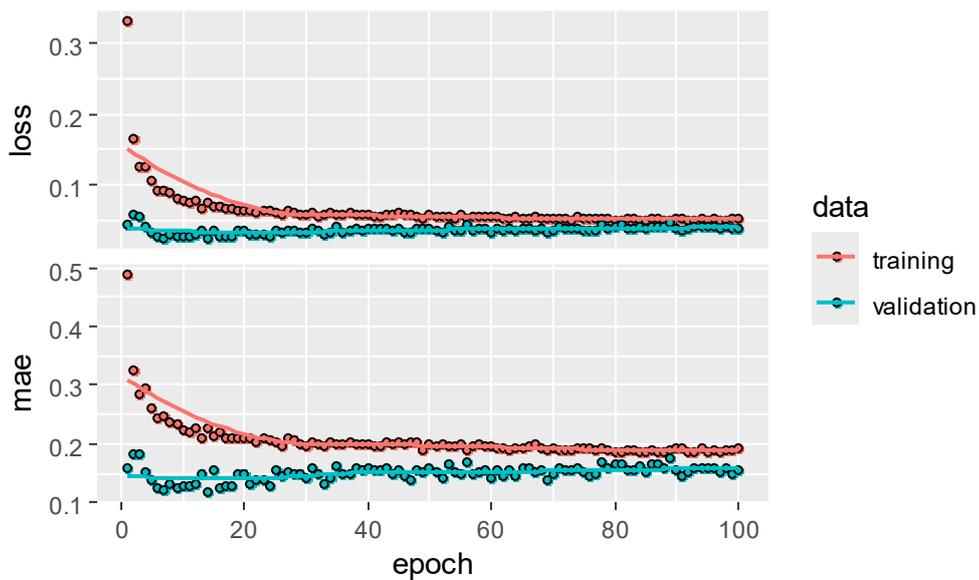**Figure 9: DNN with One Single Hidden Layer Model to Predict Financial Inclusion**

Figure 10: DNN with three Single Hidden Layer Model to Predict Financial Inclusion

Deep Neural Network Model for Financial Inclusion

Six instances of Deep Neural Network (DNN) models, are developed to classify and predict the financial inclusion of Sambalpur and Bargarh District. A total of 6 instances of the deep learning models are developed to classify & predict the financial inclusion accurately. In this study, 70% of entire dataset is allocated for training the models, while remaining 30% is reserved for testing. ReLU activation function is employed for hidden layer, & "sigmoid" function is utilized for output layer. To train these deep learning models, the "adam" optimizer, "accuracy" metrics, and "categorical_crossentropy" loss function are implemented. The architecture of the six distinct DNN models includes configurations of [8-1], [16-1], [64-1], [16-8-1], [32(0.4)-16(0.3)-8(0.2)-1], and [64(0.4)-32(0.3)-16(0.2)-1] for the hidden layers and output, corresponding to five input variables: "age," "gender," "education," "occupation," and "income." Table 3 presents the accuracy values for both the training and testing datasets, which are used to classify and predict financial inclusion in the Sambalpur and Bargarh Districts. It is observed that the 1st and 3rd DNN models show better accuracy results with the accuracy of [0.1820488, 0.19038792], [0.1732129, 0.1929447] and [0.1753766, 0.1960646] for the test datasets respectively. Figure 8 illustrates the top-performing first DNN, which features a single hidden layer and utilizes five inputs. Figure 9 depicts the third DNN model, also with one hidden layer and five inputs. Figure 10 presents the sixth

DNN model, characterized by three hidden layers and five inputs

.

Table III: Performance Measure of Different Financial Inclusion Prediction Models

AI-Driven Model	Neurons in Input, Hidden and Output Layers	RMSE		MAE	
		Training (70%)	Testing (30%)	Training (70%)	Testing (30%)
Regression	5-1	0.4110656	0.3466942	0.430545	0.3607383
ANN	5-8-1	0.3748397	0.4032486	0.3046555	0.331931
ANN	5-8-4-1	0.3748397	0.3800821	0.3046555	0.3080368
ANN	5-4-2-1	0.3748397	0.4140354	0.3046555	0.3424748
ANN	5-4-1	0.3748397	0.4228438	0.3046555	0.348445
DNN	5-8-1/100-32-0.2	0.2151458	0.2297273	0.1820488	0.19038792
DNN	5-16-1/100-32-0.2	0.2121082	0.2345056	0.1781334	0.195385
DNN	5-64-1	0.2067071	0.2317137	0.1732129	0.1929447
DNN	5-16-8-1/100-32-0.2	0.217069	0.2356593	0.1829066	0.1954742
DNN	5-32(0.4)-16(0.3)-8(0.2)-1/100-32-0.2	0.2151749	0.2387213	0.1825827	0.2003215
DNN	5-64(0.4)-32(0.3)-16(0.2)-1/100-32-0.2	0.2090163	0.236027	0.1753766	0.1960646

Conclusion

This study investigated the impact of demographic factors on financial inclusion using various machine learning techniques. A multistage sampling approach was employed to collect data from 813 respondents in the Sambalpur and Bargarh districts. The dataset was analyzed using descriptive techniques and advanced statistical methods, with 70% used for training and 30% for testing. Several AI-based models were used such as decision trees, artificial neural networks (ANN), regression trees, regression models, and deep neural networks (DNN). The second ANN model was the best performing with lower values of RMSE, as well as the MAE in both the training and testing set of data. The 1st, 3rd and 6th DNN models were also indicated to have high predictive accuracy in "financial inclusion". The results of decision tree examination specified that though high pay is a great predictor of financial inclusion, other variables include age, occupation, and gender that exhibit a complex relationship with a majority of the population. Regression tree gave income as the most motivated variable whereby, other demographic variables contributed to more detailed forecasts among low-income groups. Results of these implications are enormous in regards to local administration, national government and financial institutions in Sambalpur, Bargarh and possibly in other rural regions. The predictive model is an important contribution to the stakeholders and it provides a basis of future research, interregional comparisons, and how to formidable plans to increase financial inclusion.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used Machine Learning in order to analyze the data. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

REFERENCES

1.   Abedinia, A. and Seydi, V. (2024), "Building semi-supervised decision trees with semi-cart algorithm", International Journal of Machine Learning and Cybernetics, Vol. 15 No. 10, pp. 4493-4510.
2.   Agu, E.E., Abhulimen, A.O., Obiki-Osafiele, A.N., Osundare, O.S., Adeniran, I.A. and Efunniyi, C.P. (2024), "Utilizing AI-driven predictive analytics to reduce credit risk and enhance financial inclusion", International Journal of Frontline Research in Multidisciplinary Studies, Vol. 3 No. 2, pp. 020-029.
3.   Ajzen, I. (1991), "The theory of planned behavior", Organizational Behavior and Human Decision Processes, Vol. 50 No. 2, pp. 179-211.
4.   Amarnath, D.D. and Thimothy, S.J. (2024), "Revolutionizing the financial landscape", IGI Global, pp. 55-75.

5. Bapat, D. (2019), "Financial inclusion and economic development: Evidence from India", Economic and Political Weekly, Vol. 54 No. 33, pp. 45-53.

6. Bashir, A. (2023), "The state of financial inclusion and digitization in Pakistan: Insights from the global financial inclusion index 2021", Pakistan Journal of International Affairs, Vol. 6 No. 2.

7. Behera, B.K. (2024), "Understanding financial inclusion and its socio-economic determinants: Evidence from India", Review of Development and Change.

8. Buja, A. and Lee, Y.-S. (2001), "Data mining criteria for tree-based regression and classification", pp. 27-36.

9. Cai, H., Zhu, L., Han, S., Lin, Y., Lin, J., Tang, H., Liu, Z. and Wang, H. (2022), "Enable deep learning on mobile devices: Methods, systems, and applications", ACM Transactions on Design Automation of Electronic Systems, Vol. 27 No. 3, pp. 1-50.

10. Chakraborty, A., Singh, B., Rao, C. and Kumar, D. (2020), "Evaluation of predictive models for financial inclusion using RMSE and MAE", Journal of Financial Data Science, Vol. 8 No. 3, pp. 145-162.

11. Chandrasekar, P., Bhattacharya, P., Qian, K. and Shahriar, H. (2017), "Improving the prediction accuracy of decision tree mining with data preprocessing", COMPSAC 2017, Vol. 2, pp. 481-484.

12. Chaudhary, S. and Kumar, A. (2020), "Digital literacy and financial inclusion: A study of rural households in India", Journal of Rural Development, Vol. 39 No. 3, pp. 321-338.

13. Chung, I.-H., Ramabhadran, B., Chauhari, U., Picheny, M., Gunnels, J., Sainath, T.N., Austel, V. and Kingsbury, B. (2017), "Parallel deep neural network training for big data on Blue Gene/Q", IEEE Transactions on Parallel and Distributed Systems, Vol. 28 No. 6, pp. 1703-1714.

14. Dai, Q.-Y., Zhang, C.-P. and Wu, H. (2016), "Research of decision tree classification algorithm in data mining", International Journal of Database Theory and Application, Vol. 9 No. 5, pp. 1-8.

15. Das, S. (2024), "Bridging India's financial divide: The power of artificial intelligence and machine learning", International Journal for Multidisciplinary Research, Vol. 6 No. 5.

16. Debuque-Gonzales, M. and Corpus, J.P. (2024), "Understanding financial inclusion in the Philippines", Philippine Journal of Development, Vol. 48 No. 01, pp. 01-38.

17. Dlamini, J.S., Marshall, L. and Modupe, A. (2023), "Financial inclusion in Sub-Saharan emerging markets: The application of deep learning to improve determinants", Springer Science+Business Media, pp. 219-236.

18. Fan, F.-L., Xiong, J., Li, M. and Wang, G. (2021), "On interpretability of artificial neural networks: A survey", IEEE Transactions on Radiation and Plasma Medical Sciences, Vol. 5 No. 6, pp. 741-760.

19. Gama, J., Rodrigues, P. and Medas, P. (2005), "Learning decision trees from dynamic data streams", 13 March.

20. Gupta, P. (2022), "AI-powered credit scoring models: Transforming financial inclusion in rural India", Innovative Research Thoughts, Vol. 8 No. 4.

21. Guz, T. and Poyraz, G. (2024), "Digital financial inclusion and its determinants: Evidence from Turkiye", Ekonomi, Politika & Finans Arastirmalari Dergisi, Vol. 9 No. 4, pp. 700-714.

22. Huang, Y. (2009), "Advances in artificial neural networks - Methodological development and application", Algorithms, Vol. 2 No. 3, pp. 973-1007.

23. Jena, M. and Dehuri, S. (2020), "Decision tree for classification and regression: A state-of-the-art review", Informatica, Vol. 44 No. 4.

24. Jin, S., Gan, C. and Le Trang Anh, D. (2024), "Financial inclusion toward economic inclusion: Empirical evidence from China's rural household", Agricultural Finance Review.

25. Kaur, L. and Bansal, D. (2024), "Financial inclusion status, measurement, determinants and gender gap: An empirical analysis of Punjab state in India".

26. Kaur, R. and Gupta, P. (2020), "Machine learning techniques in financial inclusion: A review", International Journal of Management and Applied Science, Vol. 6 No. 5, pp. 12-17.

27. Kaviyarasu V, P.A. and Siddiq, A. (2024), "Impact of demographic factors on financial inclusion of women through government schemes of unorganized sector in Mysore district", EPRA International Journal of Economic and Business Review, pp. 44-51.

28. Khushboo, M. and Pradhan, K.C. (2024), "Examining the determinants of financial inclusion in South Asian countries", The Indian Economic Journal.

29. Kiralioglu, T. (2024), "Investigating the use of machine learning in automating credit scoring for microfinance", Human-Computer Interaction, Vol. 8 No. 1, p. 29.

30. Kumar, K. and Thakur, G.S.M. (2012), "Advanced applications of neural networks and artificial intelligence: A review", International Journal of Information Technology and Computer Science, Vol. 4 No. 6, pp. 57-68.

31. Kumar, S. and Pradhan, K.C. (2024), "Socioeconomic and demographic determinants of financial inclusion in South Asia: Integrated policy for targeted groups of population", Journal of Policy Modeling.

32. Lee, H.S., Kuang, K.S., Liew, P.-X., Md Nassir, A. and Ab Razak, N.H. (2023), "Machine learning models for financial inclusion in Malaysia: Opportunity for insurance or Takaful in achieving financial inclusion", pp. 981-985.

33. Lee, L. (2024), "Enhancing financial inclusion and regulatory challenges: A critical analysis of digital banks and alternative lenders through digital platforms, machine learning, and large language models integration", Social Science Research Network.

34. Lewis, M.J., Spiliopoulou, A., Goldmann, K., Pitzalis, C., Mckeigue, P. and Barnes, M.R. (2023), "nestedcv: An R package for fast implementation of nested cross-validation with embedded feature selection designed for transcriptomics and high-dimensional data", Bioinformatics Advances, Vol. 3 No. 1.

35. Li, C. (2022), "Analysing the determinations of financial inclusion in Africa based on random forest model and logistic regression model", 2022 International Conference on Machine Learning and Intelligent Systems

Engineering (MLISE).

36.   Loh, W. (2011), "Classification and regression trees", WIREs Data Mining and Knowledge Discovery, Vol. 1 No. 1, pp. 14-23.

37.   Maehara, R., Benites, L., Noguera Talavera, A.J., Aybar-Flores, A. and Muñoz, M. (2024), "Predicting financial inclusion in Peru: Application of machine learning algorithms", Journal of Risk and Financial Management.

38.   Mhlanga, D. (2022), "An analysis of the influence of socioeconomic and demographic factors on financial inclusion in underdeveloped regions", International Journal of Research in Business and Social Science, Vol. 11 No. 6, pp. 341-349.

39.   Mishra, A. (2021), "Financial inclusion and occupational diversity: Evidence from rural India", Indian Journal of Finance and Banking, Vol. 10 No. 2, pp. 55-66.

40.   Modigliani, F. and Brumberg, R. (1954), "Utility analysis and the consumption function: An interpretation of cross-section data", in Kurihara, K.K. (Ed.), Post-Keynesian Economics, Rutgers University Press, New Brunswick, NJ, pp. 388-436.

41.   Mose, N., Gitonga, M. and Fumey, M.P. (2024), "Socio-economic determinants of digital financial inclusion in Kenya", Journal of the Kenya National Commission for UNESCO, Vol. 5 No. 1.

42.   Nandru, P., Rentala, S. and Suresh, V. (2021), "Demographic discriminators in the adoption of banking services: Evidence from the primitive tribal households", Palgrave Macmillan, Singapore, pp. 3-25.

43.   Nursjanti, F. and Amaliawiati, L. (2024), "Demographic and socio-economic determinants of financial literacy: A literature review", Gema Wiralodra, Vol. 15 No. 3.

44.   Nwaimo, C.S., Adegbola, A.E. and Adegbola, M.D. (2024), "Predictive analytics for financial inclusion: Using machine learning to improve credit access for under banked populations", Computer Science & IT Research Journal, Vol. 5 No. 6, pp. 1358-1373.

45.   Ocharive, A. and Iworiso, J. (2024), "The impact of digital financial services on financial inclusion: A panel data regression method", International Journal of Data Science and Analysis.

46.   Omogbeme, A.O., Phil-Ugochukwu, A.I., Nwabufo, I. and Nwabufo, J.O. (2024), "The role of artificial intelligence in enhancing financial inclusion: A review of its impact on financial services for the unbanked population in the United States", World Journal of Advanced Research and Reviews, Vol. 23 No. 2, pp. 2184-2192.

47.   Osman, H., Nierstrasz, O. and Ghafari, M. (2017), "Automatic feature selection by regularization to improve bug prediction accuracy", pp. 27-32.

48.   Pattnayak, N.C. and Sahoo, R. (2024), "A study on socio-demographic determinants of digital financial literacy in India", Theoretical and Practical Research in Economic Fields, Vol. 15 No. 4, p. 1012.

49.   Quan, Z. and Valdez, E.A. (2018), "Predictive analytics of insurance claims using multivariate decision trees", Dependence Modeling, Vol. 6 No. 1, pp. 377-407.

50.   Rai, S. and Nath, V. (2022), "Digital divide and financial inclusion in rural India", Journal of Financial Economic Policy, Vol. 14 No. 1, pp. 76-90.

51.   Rioja, I.C. and Valdez, M.I.C. (2024), "Estudio de causalidad sobre problemas de inclusion financiera con tecnicas de ciencia de datos: El caso de Mexico", The Anahuac Journal, Vol. 24 No. 1.

52.   Roheen Qamar, R. and Baqar Ali Zardari, B. (2023), "Artificial neural networks: An overview", Mesopotamian Journal of Computer Science, pp. 124-133.

53.   Rogers, E.M. (2003), Diffusion of Innovations, 5th ed., Free Press, New York.

54.   Schmidgall, S., Hajiseyedrazi, S.P., Achterberg, J., Kirsch, L., Eshraghian, J. and Ziaei, R. (2024), "Brain-inspired learning in artificial neural networks: A review", APL Machine Learning, Vol. 2 No. 2.

55.   Sen, A. (1999), Development as Freedom, Alfred A. Knopf, New York.

56.   Shak, M.S., Uddin, Md.A., Rahman, M.H., Anjum, N.A., Bony, M.N.V.A., Alam, M., Helal, M.A., Khan, A.A., Das, P. and Pervin, T. (2024), "Innovative machine learning approaches to foster financial inclusion in microfinance", Journal of Global Education and Research, Vol. 05 No. 11, pp. 6-20.

57.   Sharma, R. and Jain, S. (2020), "Demographic determinants of financial inclusion in rural India", Asian Journal of Economics and Finance, Vol. 2 No. 4, pp. 67-80.

58.   Singh, P. and Naik, G. (2021), "Gender and financial inclusion in India: Challenges and prospects", International Journal of Gender Studies, Vol. 5 No. 1, pp. 22-35.

59.   Soejono, F. and Mendari, A.S. (2022), "Financial literacy, individual characteristic and financial inclusion: A study on lecturers in Palembang", International Journal of Multidisciplinary, Vol. 3 No. 2, pp. 269-277.

60.   Sze, V., Emer, J.S., Yang, T.-J. and Chen, Y.-H. (2020), "Efficient processing of deep neural networks", Synthesis Lectures on Computer Architecture, Vol. 15 No. 2, pp. 1-341.

61.   Varadarajan, M.N. (2022), "Unlocking financial intelligence: The role of artificial intelligence and machine learning in the industry", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 11 No. 11, pp. 1-14.

62.   Verma, K., Yadav, R. and Singh, M. (2021), "Artificial intelligence in financial inclusion: Applications of decision tree and deep learning models", International Journal of Data Science, Vol. 7 No. 2, pp. 145-160.

63.   Waliszewski, K. (2023), "Socio-economic determinants of financial inclusion in Central and Eastern Europe during the COVID-19 pandemic", Przeglad Wschodnioeuropejski, Vol. 14 No. 1, pp. 25-41.

64.   Wardani, D.T.K., Khusniati, N. and Darsono, S.N.A.C. (2023), "Sociodemographic effects on financial inclusion: Implications from online transaction in Developing-8 countries", Economics and Finance in Indonesia.

65.   Yang, L., Zhu, D., Liu, X. and Cui, P. (2023), "Robust feature selection method based on joint L2, 1 norm minimization for sparse regression", Electronics, Vol. 12 No. 21, p. 4450