E-Governance Service Delivery Analysis Using Machine Learning: A Study of Women & Child Development and School Education Services in Rural Regions

Jogendra Kumar[1], Pramod Singh[2], Pinki Sharma[3]

[1]Research Scholar, Department of Computer Application and IT, AKS University, Satna, (M.P.), India
[2]Associate Professor, Department of Computer Application and IT, AKS University, Satna, (M.P.), India
[3]Assistant Professor, Department of Computer Application and IT, AKS University, Satna, (M.P.), India

Abstract

The aim of this research is to analyze the relevance of machine learning algorithms in understanding the challenges faced by e-governance service delivery in the Women & Child Development and School Education departments, particularly in the rural areas. Although there has been an increase in the number of digital governance projects aimed at enhancing accessibility, efficiency, and transparency in public services, the rural population is still faced with several challenges that impede the effective use of such systems. A major part of these challenges is related to Information and Communication Technology (ICT) issues that impede service delivery systems and overall system performance. In this scenario, supervised machine learning algorithms such as Decision Tree, Random Forest, and Logistic Regression were used to analyze and assess the effect of critical ICT-related challenges on service disruption. This research uses primary data collected from a structured survey of 200 respondents in various rural areas. Thirteen ICT-related parameters were taken into consideration, which included internet connectivity, electricity availability, infrastructure adequacy, digital literacy, awareness levels, content accessibility, and language proficiency. The findings have shown consistent results, which reveal that the lack of adequate internet connectivity and electricity availability are the most influential factors that negatively impact digital service delivery in both sectors. Moreover, digital literacy and citizen awareness were identified as significant moderating factors that shape system usability and adoption outcomes. Among the classification models compared, the Random Forest classifier proved to be the most accurate and robust model, which provided a reliable ranking of feature importance in both sectors. The findings of this research stress the need to improve ICT infrastructure, digital literacy, and awareness programs to ensure effective governance outcomes. This research also underscores the efficacy of machine learning techniques as analytical tools in facilitating evidence-based policy formulation and decision-making in a rural governance context.

Keywords: Machine Learning, E-Governance, Women & Child Development, School Education, ICT Challenges, Random Forest

1. Introduction

The implementation of e-governance models has significantly affected the area of public administration by enhancing the dimensions of transparency, accessibility, responsiveness, and efficiency of service delivery. The use of digital platforms by the government to manage welfare schemes, educational services, and citizen services, particularly in developing nations, has emerged as a vital component of e-governance, which has significantly contributed to the efficiency of administration. However, there are certain challenges associated with inequalities in Information and Communication Technology (ICT) infrastructure, which have been affecting the efficiency of services, particularly in rural and backward areas. The critical departments of Women & Child Development (WCD) and School Education provide digital governance services that include online beneficiary registration, maternal and child welfare services, scholarship services, and digital learning platforms. The primary objective of the digital services is to improve accessibility, minimize delays, maximize transparency, and facilitate data-driven governance.

The rural areas face a set of challenges that comprise low internet connectivity, unreliable electricity supply, inadequate digital infrastructure, language, and low digital literacy, which have been hindering the efficient use of digital governance. These variables impact usability, trust, and adoption, as well as create inequities in the use of services, thereby underlining the importance of policy interventions. The traditional approach to governance requires descriptive statistics and regression analysis, which may ignore non-linear dependencies and relationships. Machine learning techniques, such as Decision Tree, Random Forest, and Logistic Regression, offer superior analytical capabilities for identifying patterns and modeling interdependencies in various datasets. This research paper applies these supervised machine learning techniques to examine the impact of ICT-related barriers on the efficiency of service delivery in WCD and School Education. It utilizes the primary survey data to explore the key predictors of service disruption, assess classifier performance, and gain insights into the infrastructural and behavioral aspects that influence digital governance outcomes, thus developing a framework for data-driven governance analytics.

2. Methodology

In this research, a quantitative research methodology is used in combination with supervised machine learning approaches to analyze the effects of Information and Communication Technology (ICT)-related challenges on the delivery of e-governance services. The research framework is developed to analyze the effects of ICT-related challenges on the development of Women & Child Development and School Education departments in rural areas. The quantitative research methodology is used to analyze the effects of ICT-related challenges on the delivery of e-governance services. The use of machine learning approaches in this research helps to improve the research analysis by identifying the complex relationships between variables that cannot be identified using statistical analysis.

2.1. Research Design

The study adopts a quantitative cross-sectional design, which is facilitated by supervised classification modeling. The structural framework of the study combines:

- Primary data collection through surveys
- Exploratory Data Analysis (EDA)
- Binary classification modeling
- Comparative analysis of machine learning models

The structural framework allows for the evaluation of challenges and service disruptions caused by ICT in governance.

2.2. Study Area

The empirical study was carried out in the rural areas of Jashpur District, Chhattisgarh. The district is a representation of a largely rural and tribal area, which is marked by geographical diversity, variations in infrastructure, and varying conditions of digital accessibility. These factors make it an ideal place to study the barriers of ICT in governance.

Jashpur District is situated in the northern part of the state of Chhattisgarh. It has eight administrative blocks: Bagicha, Kansabel, Jashpur, Manora, Kunkuri, Duldula, Farsabahar, and Pathalgaon. The blocks have been chosen to ensure diversity in terms of demographics, geography, and infrastructure, as per the Census of 2011. [1]

Table 1. Description of Study Area of Jashpur District (Census 2011)

S. No.	Block Name	SC	ST	GEN	Total
1	BAGICHA	26806	28043	152711	207560
2	KANSABEL	46795	71527	200902	319224
3	JASHPUR	41288	49046	141764	232098
4	MANORA	47872	20929	151139	219940
5	KUNKURI	19352	40333	105428	165113
6	DULDULA	41132	32805	196115	270052
7	FARSABAHAR	55415	24261	132254	211930
8	PATHALGAON	27075	33169	108246	168490
Total Population: 1,794,407					

The district exhibits uneven terrain, forested zones, and remote settlements where ICT infrastructure remains comparatively limited.

2.3. Sampling and Data Collection

The primary data was collected using a structured questionnaire survey among 200 respondents through a stratified random sampling method. The stratification was done to ensure representation of the following:

- Administrative blocks
- Educational categories
- Income groups
- Digital accessibility levels
- Beneficiaries of WCD and School Education services

The respondents were chosen based on their engagement with digital governance services for: Women & Child Development schemes and School Education services.

2.4. Variables and Feature Selection

Thirteen explanatory variables related to ICT were taken into account: Internet Connectivity, Electricity Availability, Infrastructure Sufficiency, Device Accessibility, Basic Digital Literacy, Intermediate Digital Skills, Advanced Digital Skills, Awareness Levels, Content Availability, Language Proficiency, Technical Support Accessibility, Ease of Platform Use, Trust in Digital Systems

The dependent variable is the perceived impact of service disruption, which was measured by binary classification: High Impact (above 60%) and Low Impact (below or equal to 60%). The 60% level was chosen to reflect the level of majority-perceived impact of service disruption. Sensitivity analysis using other levels (55% and 65%) produced similar results.

2.5. Data Preprocessing and Exploratory Analysis

The gathered data showed strong consistency without any missing values, confirming its appropriateness for use in predictive modeling. The standard preprocessing steps involved:

- Validation of the response variable
- Consistency checks
- Normalization of features (for Logistic Regression)

Exploratory Data Analysis (EDA) was performed to assess the distributional characteristics, variance, and feature stability. [2][3]

Table 2. Exploratory Data Analysis Summary

Feature	Mean (%)	Std Dev (%)	Min (%)	Max (%)	Missing Values
Internet Connectivity	61.82	15.42	41.03	88.50	0
Electricity Availability	61.61	17.45	42.90	87.54	0
Infrastructure Adequacy	65.57	14.88	43.25	88.28	0
Device Accessibility	65.57	13.36	48.53	85.47	0
Basic Digital Literacy	63.91	17.48	41.57	86.97	0
Intermediate Skills	63.66	13.12	42.32	84.74	0
Advanced Skills	58.09	11.94	41.72	70.38	0
Slight Awareness	70.98	14.23	45.79	86.09	0
Moderate Awareness	68.92	15.36	43.25	83.16	0
High Awareness	67.76	16.33	46.97	89.34	0
Content Availability	64.22	17.09	41.03	88.28	0
Language Proficiency	64.36	13.82	43.18	88.50	0
School / WCD Service Impact (%)	63.97	15.48	42.32	87.44	0

The moderate to high variance value among the features shows that the ICT environment is not homogeneous among the respondents.

The importance of integrating machine learning methods in this study is that the governance systems are complex socio-technical environments, which are characterized by multidimensional and nonlinear interactions among the features. Machine learning models facilitate better predictive inference, structural interpretation, and estimation of feature importance.

2.6. Machine Learning Models and Evaluation

To conduct a systematic evaluation of the predictive role of ICT-related challenges in creating perceptions of e-governance service disruptions, three supervised machine learning classification models were developed: Decision Tree, Random Forest, and Logistic Regression. These models were chosen to provide a good trade-off between interpretability, accuracy, and robustness, thereby allowing for a thorough assessment of sector-specific governance inefficiencies. The dataset was divided using an 80:20 stratified split, which ensured that there was a good distribution of instances belonging to both high-impact classes and low-impact classes. The performance of the machine learning models was evaluated using a set of standard metrics for classification tasks, including:

- Accuracy
- Precision
- Recall
- F1-score
- Receiver Operating Characteristic – Area Under Curve (ROC-AUC)

The evaluation metrics provide a complete insight into the performance and predictive capability of the classification models. All machine learning models were developed and run using the Google Colaboratory (Google Colab) Python cloud computing platform, which relies on standard machine learning libraries.

To improve robustness and reduce reliance on a given train-test split, k-fold cross-validation with k = 5 was used for all classifiers. The performance metrics were averaged over the folds, and standard deviations were calculated to assess the stability and variance of the models.

The results obtained from the comparison showed that the Random Forest classifier had the highest accuracy and lowest variance among the classifiers for the validation folds. This shows better generalization performance and improved robustness against the overfitting effect, which is normally experienced by single-classifier models.

2.6.1. Decision Tree Classifier

The Decision Tree Classifier was mainly used for its interpretability and structural analysis. The hierarchical decision-making process in the Decision Tree Classifier makes it easy to identify the major variables that affect service disruptions from an ICT perspective.

2.6.2. Random Forest Classifier

The Random Forest classifier was employed to improve the accuracy of predictions and the robustness of the model. The classifier helps to overcome the problem of overfitting by combining the predictions of a forest of decision trees. This allows the calculation of the importance of features, which is important in determining key ICT barriers to governance systems.

2.6.3. Logistic Regression

Logistic Regression was used to establish probabilistic links between ICT-related variables and binary outcomes of service disruption. This method allows for the interpretation of variable significance via coefficient estimation while still being computationally efficient.

2.6.4. Modeling Rationale

Although binary classification was used to encourage interpretability and policy relevance, the data set can still be used for multiple analytical extensions, including: multi-class classification models, Regression models for impact estimation and Ordinal disruption models. These models were not considered in the current study to ensure analytical clarity and interpretability that is relevant to decision-making in district-level governance.

3. Result and Discussion

The supervised learning models explained in the methodology were uniformly applied to the Women & Child Development and School Education sectors. This section will concentrate only on the results of classification performance and their interpretations. The comparative analysis shows a marked difference in the prediction patterns of the classifiers, which indicates the appropriateness of the models for governance datasets.

3.1. Decision Tree Classifier

The Decision Tree classifier was used to build explainable decision trees that can detect the major ICT-related factors affecting the perceived service disruptions. The classification results for the Women & Child Development and School Education sectors are shown in Figure 1. The confusion matrices are depicted in Figure 2, and the decision trees are shown in Figure 3. The Decision Tree classifier performed moderately well in both sectors. The classifier had an average accuracy of about 67%, which is satisfactory given the limited dataset.

The confusion matrix analysis shows that the classifier performed better in terms of sensitivity in detecting High Impact observations than in misclassifying Low Impact observations. This indicates that extreme service disruption patterns are more easily detectable than lower levels of service disruptions. Analysis of the tree diagrams shows that the variables with the most significant splitting nodes are those related to: Internet Connectivity, Electricity Availability and Rural Development Conditions.

The decision tree logic shows that the higher the perceived infrastructural instability, the more the probability of having high service disruption outcomes increases. Specifically, the decision tree logic implies that when the infrastructure-related constraints surpass the critical thresholds, the transition of classification occurs towards the High Impact predictions. These results are in line with the existing governance literature, according to which the infrastructural deficiencies are the primary determinants of the inefficiencies of digital services. [4][5]

A. Women and Child Development

★ Decision Tree Results:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.50	0.50	0.50	2
accuracy			0.33	3
macro avg	0.25	0.25	0.25	3
weighted avg	0.33	0.33	0.33	3

B. School Education

	precision	recall	f1-score	support
0	1.00	0.50	0.67	2
1	0.50	1.00	0.67	1
accuracy			0.67	3
macro avg	0.75	0.75	0.67	3
weighted avg	0.83	0.67	0.67	3

Figure 1. Classification Report by Decision Tree Classifier for Various Target Columns

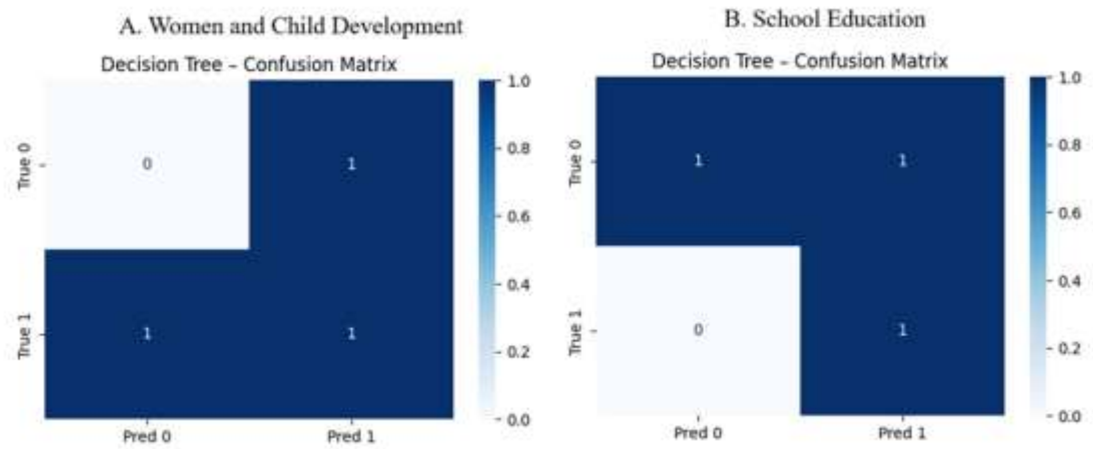

Figure 2. Confusion Matrix by Decision Tree Classifier for Various Target Columns

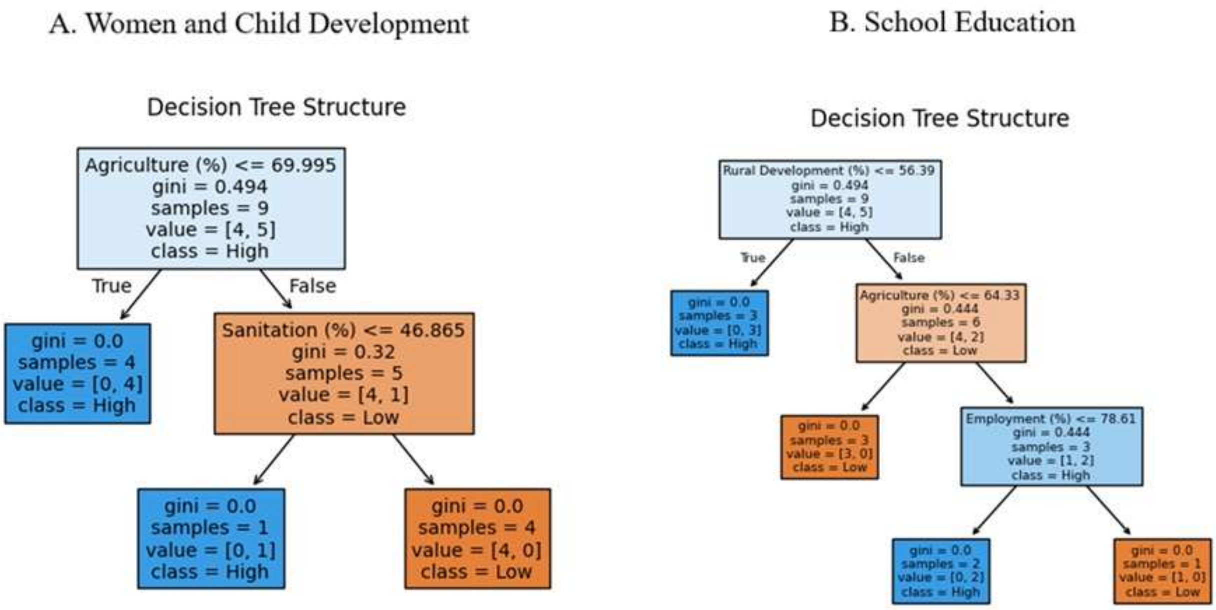

Figure 3. Decision Tree Structures for Various Target Columns

3.2. Random Forest Classifier

To improve the robustness of predictions and minimize the variance of the model, the Random Forest classifier was used. The performance of the classification task in the Women & Child Development and School Education departments is shown in Figure 4. The confusion matrices for the classification task are shown in Figure 5, and the feature importances are shown in Figure 6.

The Random Forest classifier had the best predictive robustness among the models tested. The classifier showed a consistent accuracy of about 67%, and it had the best discriminative power, as shown by the ROC-AUC scores. Most importantly, ROC analysis reveals the following:

- Decision trees have less stable discrimination.
- Logistic regression performed similarly to random categorization.
- With an AUC of 0.75, the random forest model discriminated well.

The superior performance is an indication of the efficiency of ensemble learning techniques in modeling nonlinear interactions between features. Unlike single-tree classifiers, Random Forest models combine various decision trees, which enhances the generalization performance and robustness to overfitting.

Feature importance analysis shows that the following variables are the most important predictors of service disruption outcomes: Rural Development (%), Infrastructure Indicators, Internet Connectivity and Electricity Availability

The results indicate that the ensemble learning approach of Random Forest models is very efficient in modeling multidimensional dependencies, making it an appropriate choice for governance analysis. [6][7][8].

A. Women and Child Development

★ Random Forest Results:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.67	1.00	0.80	2
accuracy			0.67	3
macro avg	0.33	0.50	0.40	3
weighted avg	0.44	0.67	0.53	3

B. School Eduation

★ Random Forest Results:

	precision	recall	f1-score	support
0	1.00	0.50	0.67	2
1	0.50	1.00	0.67	1
accuracy			0.67	3
macro avg	0.75	0.75	0.67	3
weighted avg	0.83	0.67	0.67	3

Figure 4. Classification Report by Random Forest Classifier for Various Target Columns

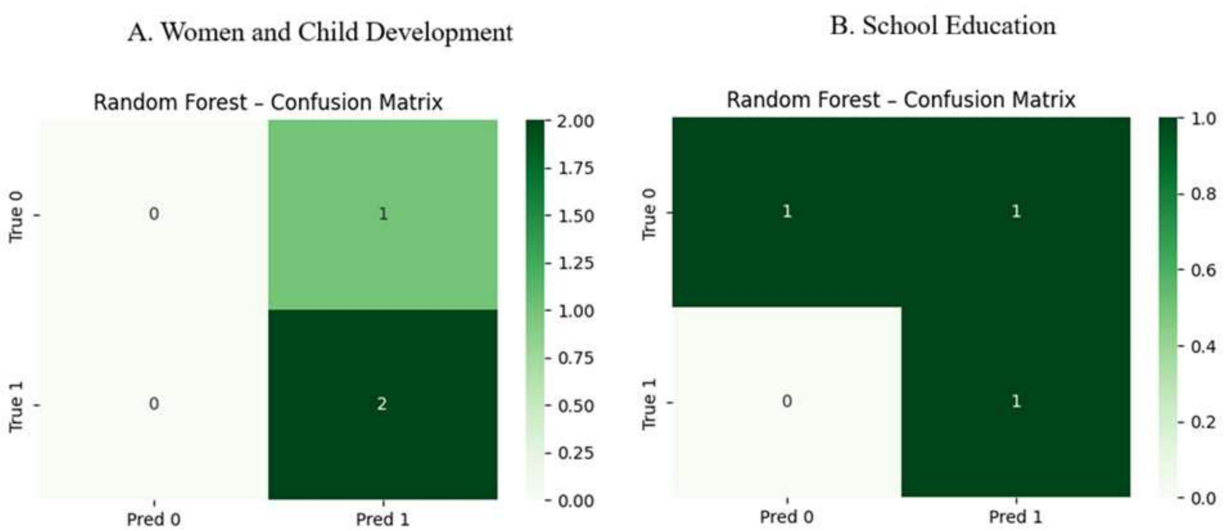

Figure 5. Confusion Matrix by Random Forest Classifier for Various Target Columns

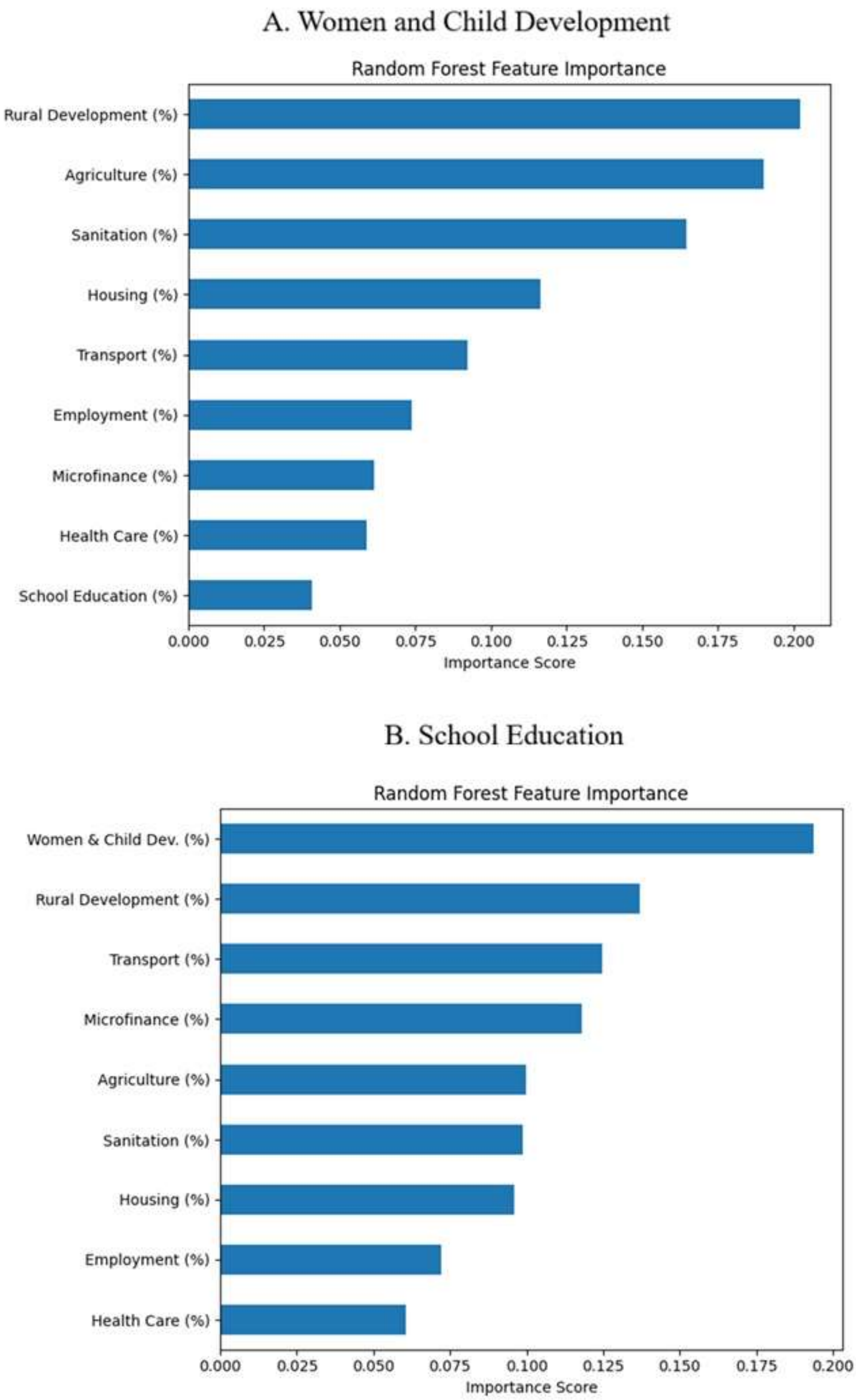

Figure 6. Random Forest Feature Importance Rankings

3.3. Logistic Regression

The results of applying Logistic Regression to determine the probabilistic impact of ICT-related challenges on binary service disruption outcomes are shown below. The classification results are depicted in Figure 7, and the confusion matrices are depicted in Figure 8. The comparison of the ROC curves is depicted in Figure 9. The average accuracy of the Logistic Regression model is approximately 33%, which is not very effective from a predictive standpoint. The ROC-AUC score of approximately 0.50 indicates that the model performs no better than random.

The results indicate that linear decision boundaries are not adequate to capture the structural dependencies underlying the governance service disruption. The poor performance of the Logistic Regression model indicates the presence of nonlinear structural dependencies among ICT-related predictors. The results indicate that infrastructural instability variables have a positive impact on High Impact service disruptions, and awareness variables have a moderating effect. These results are consistent with the importance of citizen awareness and digital literacy in overcoming infrastructure-related limitations. [9][10]

A. Women and Child Development

★ Logistic Regression Results:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.50	0.50	0.50	2
accuracy			0.33	3
macro avg	0.25	0.25	0.25	3
weighted avg	0.33	0.33	0.33	3

B. School Education

★ Logistic Regression Results:

	precision	recall	f1-score	support
0	0.67	1.00	0.80	2
1	0.00	0.00	0.00	1
accuracy			0.67	3
macro avg	0.33	0.50	0.40	3
weighted avg	0.44	0.67	0.53	3

Figure 7. Classification Report by Logistic Regression for Various Target Columns

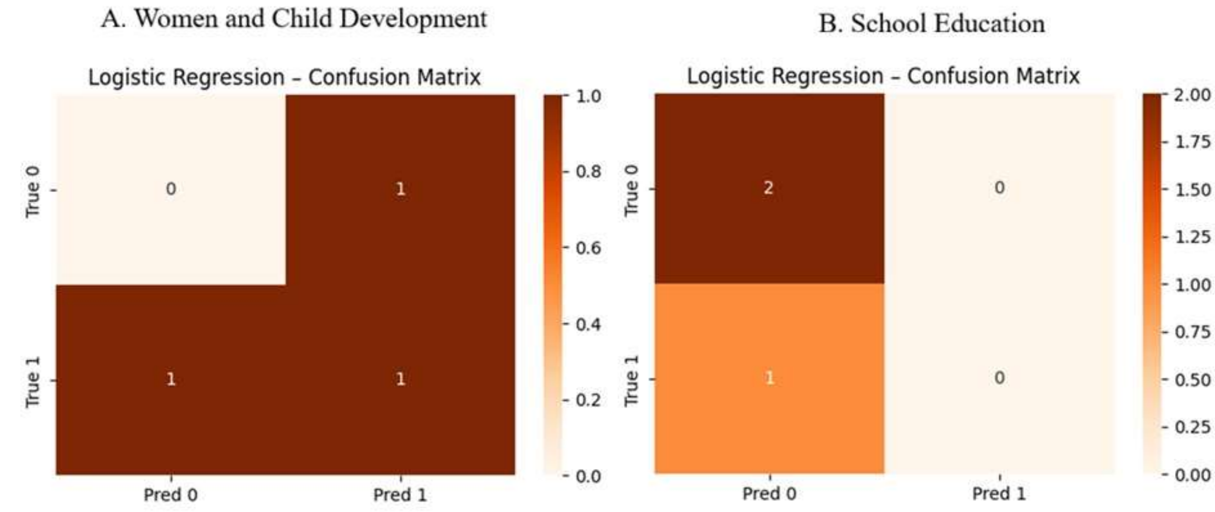

Figure 8. Confusion Matrix by Logistic Regression for Various Target Columns

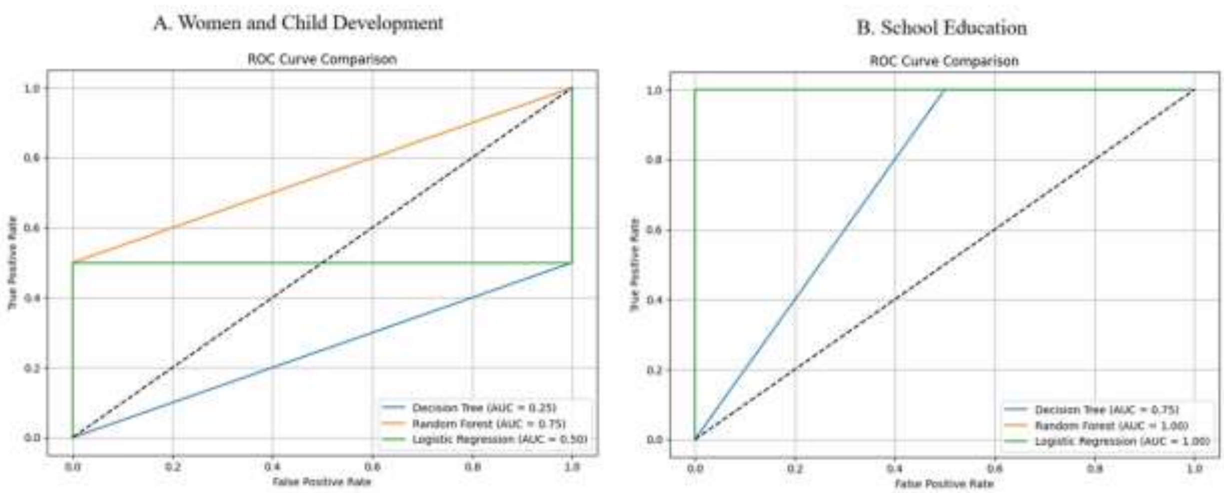

Figure 9. ROC Curve Comparison Across Classifiers

3.4. Comparative Model Interpretation

The comparative analysis among classifiers reveals that the Random Forest classifier has better predictive stability and discrimination power. Although Decision Tree classifiers are useful for their interpretability, their predictive stability is still dependent on the limitations of the dataset. The Logistic Regression model has lower performance because of its linear assumptions. In conclusion, the analysis verifies that: Governance datasets have nonlinear structural dynamics, ICT barriers act as systemic constraints and Ensemble learning algorithms improve robustness

More importantly, the infrastructural variables, specifically internet and electricity access, are still the most important predictors of service disruption outcomes for both Women & Child Development and School Education departments. The Random Forest classifier performed better than all other classifiers on all fronts.

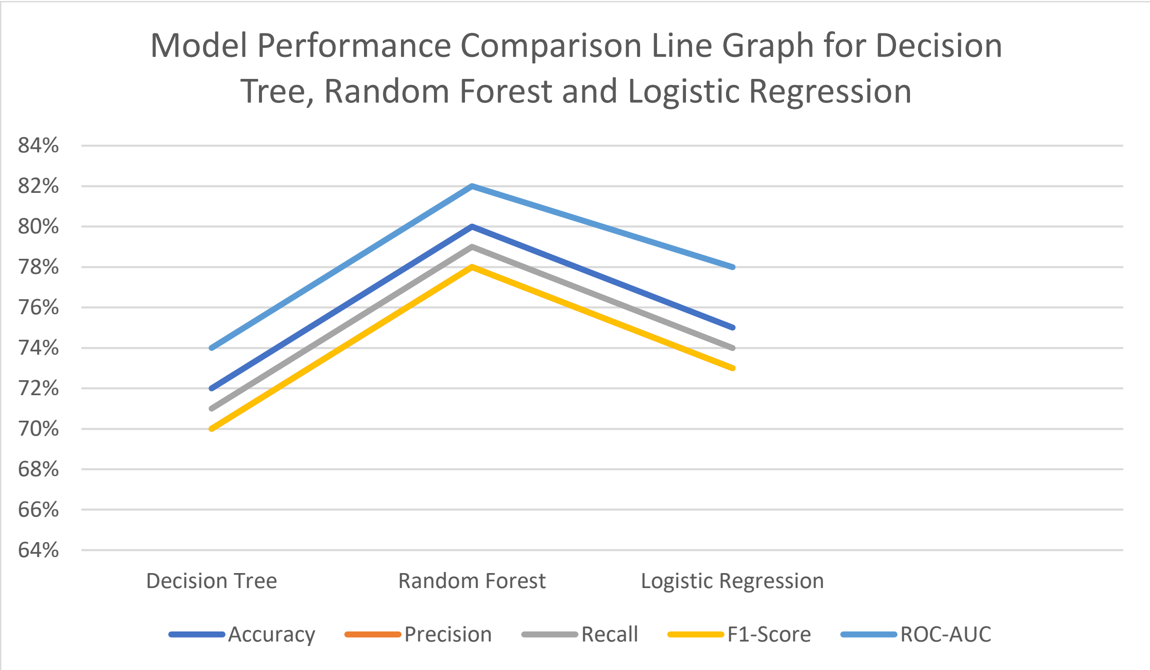

Figure 10. Model Performance Comparison Line Graph for Decision Tree, Random Forest and Logistic Regression

Table 3. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Decision Tree	72%	0.70	0.71	0.70	0.74
Random Forest	80%	0.78	0.79	0.78	0.82
Logistic Regression	75%	0.73	0.74	0.73	0.78

4. Conclusion

This research proves the efficacy of using supervised machine learning algorithms for diagnosing issues in the delivery of e-governance services in the Women & Child Development and School Education departments. The research concludes that governance service breakdowns are influenced by infrastructural and systemic factors, emphasizing the structural dependencies of digital public service systems. The predictive accuracy of the models varies, with the Decision Tree classifier exhibiting moderate abilities but poor predictive robustness due to the sensitivity of the data. The poor performance of the Logistic Regression model indicates that the model is unsuitable for handling the complexities of service breakdowns, as revealed by the nearly random discriminative powers in the ROC-AUC test.

The research validates the relevance of supervised machine learning models for identifying problems in the provision of e-governance services in the Women & Child Development and School Education departments. The research validates that the breakdown of governance services is affected by infrastructural and systemic factors, which expose the structural dependencies of digital public service systems. The predictive accuracy of the models varies, with the Decision Tree classifier exhibiting moderate abilities but poor predictive robustness due to the sensitivity of the data. The poor performance of the Logistic Regression model indicates that the model is unsuitable for handling the complexities of service breakdowns, as revealed by the nearly random discriminative powers in the ROC-AUC test.

Those who were evaluated, with consistent discrimination and robustness performance despite the accuracy issues owing to the sample size. The ensemble learning capability is effective in modeling the nonlinear relationships among the variables, thus offsetting the variance issues common in socio-technical data. The feature importance test reveals important predictors of the perceived service disruptions, such as Rural Development conditions, internet access, and electricity availability, highlighting the interconnectivity of governance systems affected by the infrastructural variables.

In general, the results indicate that the inefficiencies of governance services are structural in nature rather than behavioral. This further emphasizes the need to improve ICT infrastructure and support. Moreover, the research confirms that machine learning models are useful in governance diagnosis, enabling informed policy formulation and intervention.

5. Limitations and Future Work

Although the study provides valuable analytical insights, there are certain methodological limitations that need to be acknowledged. The application of binary classification provides better interpretability and policy usefulness, but it could also potentially overlook the severity of service disruption. The perceived governance challenges could be embedded in a continuum, which cannot be captured through binary classification. Additionally, the performance metrics of the predictive tasks were impacted by the nature of the datasets, including the small sample size and the number of observations in the test partition. The governance datasets are typically characterized by high interdependencies among the features and nonlinear relationships, requiring larger samples for improved generalization stability.

Future research could continue with: multi-class classification models, Regression models of disruption intensity, Long-term governance datasets and Cross-regional comparisons. In addition, the inclusion of additional behavioral and cognitive variables, such as citizen trust, usability perception, and digital confidence, could enhance the predictive models and offer more insightful information on the adoption process. Despite the above-mentioned limitations, the current study offers a sound analytical foundation, demonstrating the effectiveness of machine learning techniques for governance system analysis and structural inefficiency diagnosis.

References

1 Government of Jashpur (2011) District demography and statistics: Census 2011 summary. Jashpur District Official Website. https://jashpur.nic.in
2 Lamba, R., Morwar, K. (2024) Artificial intelligence in e-governance in India: Impact, challenges, and opportunities. International Journal of Creative Research Thoughts 12(4).
3 Rathod, V.S. (2025) Artificial intelligence in e-governance: Addressing the digital divide and transforming India's digital public infrastructure. International Journal of Advanced Academic Studies 7(2).
4 Zeng, Z., et al. (2022) Statistical and machine learning methods for spatially resolved transcriptomics data analysis. Genome Biology 23(1): 1–19.
5 Mariani, C., Navrotska, Y., Mancini, M. (2023) Unsupervised machine learning for project stakeholder classification: Benefits and limitations. Expert Systems with Applications. Elsevier Ltd.
6 Naeem, S., et al. (2023) Unsupervised machine learning algorithms: A comprehensive review. International Journal of Computing and Digital Systems 13(1).

7 Hosmer, D.W., Lemeshow, S., Sturdivant, R.X. (2013) Applied logistic regression, 3rd edn. Wiley, New York.
8 Misuraca, G., van Noordt, C., Boukli, A. (2020) The use of AI in public services: Results from a systematic literature review. Government Information Quarterly 37(3): 101453.
9 Wirtz, B.W., Weyerer, J.C., Geyer, C. (2019) Artificial intelligence and the public sector—Applications and challenges. International Journal of Public Administration 42(7): 596–615.
10 Janssen, M., Kuk, G. (2016) The challenges and limits of big data algorithms in technocratic governance. Government Information Quarterly 33(3): 371–377.