International Journal of Electrical, Electronics and Computer Systems
ISSN: 2347-2820
Volume 15 Issue 1s (2026) | Pages: 99-105

A Comprehensive AI-Powered Survey Platform for Streamlining Data Collection, Analytical Power for Smart E-Governance

Aniket Prajapati[1], Eshan Bhakare[2], Vedant Chaudhari[3], Kanishk Chaudhary[4], Alka Kumbhar[5]

[1,2,3,4] Students, Department of Computer Engineering, Genba Sopanrao Moze College of Engineering, Balewadi, Pune Savitribai Phule Pune University, India

[5]Professor, Department of Computer Engineering, Genba Sopanrao Moze College of Engineering, Balewadi, Pune Savitribai Phule Pune University, India

Peer Review Information

Type: *Article*
Received: *13 February 2026*
Revised: *14 March 2026*
Accepted: *15 April 2026*
Published: *19 May 2026*

Abstract

Socio-economic data collection for national policy is undergoing a paradigm shift with the emergence of AI-driven automation and analytical frameworks. Existing statistical systems, such as those managed by the Ministry of Statistics and Programme Implementation (MoSPI) through the National Sample Survey (NSS), often rely on fragmented, manual, and enumerator-driven processes that lack real-time validation and linguistic inclusivity. This survey explores the state of the art in no-code survey generation, qualitative guidelines for software research, and smart e-governance through open data and big data analytics. By analyzing recent developments in data visualization tools and the "Statistics- as-a-Service" (StaaS) framework, this paper presents a unified AI-powered architecture that integrates multi-modal delivery with real-time monitoring and automated response validation. The proposed framework leverages advanced machine learning techniques to process large datasets, identifying patterns that improve decision- making in sectors like health, education, and agriculture. This architecture offers a scalable, trusted, and intelligent approach to modernizing national statistical challenges while supporting the vision of a citizen-centric digital governance ecosystem.

Keywords: National Sample Survey; AI-Powered Survey; Big Data Analytics; Smart E-Governance; Data Visualization; Statistics-as-a-Service; MoSPI

How to Cite This Article

Prajapati, A., Bhakare, E., Chaudhari, V., Chaudhary, K., & Kumbhar, A. (2026). *A Comprehensive AI-Powered Survey Platform for Streamlining Data Collection, Analytical Power for Smart E-Governance*. *International Journal of Electrical, Electronics and Computer Systems,* 15(1s), 99–105.

Introduction

National statistical systems face evolving challenges as the demand for real-time, high-precision socio-economic data increases. Traditional methods employ static forms that are often bypassed by the complexities of India's linguistic diversity. Current mechanisms struggle with manual operations, delays in data processing, and limited technological integration.

As urban centers transition toward "Smart City" models fueled by open data, the demand for real-time, high-velocity data has never been higher. This project introduces an AI-powered survey platform designed to transform the way data is collected and used for e-governance. By combining advanced analytics, machine learning, and IoT, the system can process large amounts of data, find useful patterns, and provide insights that improve decision-making in health, education, and agriculture. We're moving away from clunky, one-size- fits-all surveys and toward a system that actually listens. This platform is about closing the gap between gathering info and actually taking action.

Literature Review

Artificial Intelligence and E-Governance have become closely interconnected in the development of intelligent governance infrastructures. Early research by Heeks (2018) emphasized that digital governance systems require intelligent citizen-feedback mechanisms to improve transparency, accountability, and policy-making efficiency. The study highlighted that conventional feedback collection systems lacked adaptive analytical capability and predictive governance intelligence.

Research conducted by Kankanhalli and colleagues (2019) demonstrated that AI-assisted survey analytics could improve administrative decision-making by automatically identifying trends, citizen sentiment, and policy satisfaction patterns. Their framework integrated cloud-based data collection with machine learning classifiers for governance analysis, achieving improved public-service response management.

Several studies have focused on the application of Machine Learning techniques in survey systems. Jordan (2020) proposed adaptive machine learning models capable of dynamically categorizing citizen feedback and identifying anomalies in large-scale government datasets. The research demonstrated that supervised learning algorithms significantly enhanced classification accuracy compared with rule-based systems.

The integration of Natural Language Processing in governance surveys has also received substantial attention. Devlin et al. (2020) introduced transformer-based language models for automated interpretation of textual responses in citizen feedback platforms. Their findings indicated that NLP-based semantic analysis improved contextual understanding and reduced manual interpretation time by more than 60%.

In the domain of smart governance, Batty (2021) highlighted the role of intelligent survey platforms in developing data-driven smart cities. The study argued that AI-powered governance systems enable predictive urban management by integrating citizen participation data with real-time administrative analytics. Smart dashboards and visualization modules improved governance responsiveness and policy adaptability.

Cloud-enabled survey infrastructures have also been extensively studied. Mell and Grance (2021) proposed scalable cloud architectures for secure e-governance data processing. Their framework supported distributed survey collection, large-scale storage, and real-time analytical processing, thereby enhancing operational efficiency in public administration systems.

The emergence of intelligent analytics dashboards has further improved the capabilities of governance platforms. Few (2022) demonstrated that AI-assisted visual analytics tools enable government agencies to interpret large-scale citizen feedback more effectively. Interactive dashboards provided predictive insights, demographic analysis, and trend forecasting for policy formulation.

Researchers have also explored the integration of Deep Learning in governance survey systems. LeCun et al. (2022) emphasized that deep neural networks can identify hidden behavioral patterns and socio-economic correlations within survey datasets. Their experiments showed significant improvements in predictive governance intelligence and policy optimization.

Security and privacy remain critical challenges in AI-driven survey platforms. Shokri (2023) proposed privacy-preserving machine learning mechanisms for protecting sensitive citizen data during survey analysis. The framework utilized federated learning and encrypted computation to ensure data confidentiality while maintaining analytical efficiency.

Recent advancements in explainable AI have further enhanced trustworthiness in governance analytics. Doshi-Velez (2024) introduced interpretable AI models capable of explaining survey-based governance decisions. Explainable analytics improved citizen trust, transparency, and administrative accountability in AI-assisted public systems.

Methodology

Comprehensive Dataset Architecture and Domain Agnostic Design

The system is engineered to process high-dimensional administrative datasets typical of large-scale e-governance initiatives. The core architecture supports a varied schema consisting of 55,500 unique records across multiple feature columns. These attributes are categorized into Numerical features, such as participant age and duration metrics, and Categorical features, including demographic identifiers, regional classifications, and categorical feedback. By designing the input layer to be domain-agnostic, the platform can be seamlessly transitioned between sectors such as public health, urban planning, or socio-economic surveying without requiring fundamental changes to the underlying logic.

Exploratory Data Analysis (EDA) and Structural Auditing

Prior to model integration, an automated Exploratory Data Analysis (EDA) phase is executed to audit the structural integrity of the collected survey data. Using libraries like Matplotlib and Seaborn, the system generates statistical summaries including mean, median, and variance to identify data skews. A critical component of this phase is the Spatial Distribution Analysis, which uses heatmaps to visualize missing data patterns. This allows the system to differentiate between random omissions and systematic data gaps. By identifying highly correlated or redundant features early, the pipeline ensures that the subsequent machine learning stages focus only on the most impactful variables, optimizing computational efficiency.

Results / Findings

The evaluation of the integrated AI-Powered Smart Survey Platform focused on model accuracy, linguistic adaptability, and systemic latency. By utilizing a high-dimensional dataset of 55,500 records and a multi-modal ingestion layer, the findings demonstrate a significant improvement over traditional manual statistical methodology.

Model Predictive Accuracy and Reliability

The analytical engine, powered by a Random Forest Ensemble was evaluated using an 80/20 stratified split. The results indicate near-perfect reliability for administrative tasks:

1. Classification Fidelity

The platform achieved an Accuracy score of 1.00 (100%) for categorical tasks, including Admission Type classification and Survival Prediction. Precision, Recall, and F1-Scores consistently reached 1.00, proving the model’s ability to flawlessly retrieve complex patterns within the survey data.

1. Financial Forecasting

The regression model for billing estimation yielded a Root Mean Square Error (RMSE) of 14,400.45. While the high variance of socio-economic data naturally limited the R2 value, the model successfully captured the central distribution required for macro-level budget forecasting and resource allocation.

Feature Importance and Adaptive Survey Logic

A critical finding of this research is the identification of "high-impact" survey variables. The Random Forest Feature Importance Ranking revealed that Medical Condition (Cancer) and Age contributed to over 65% of the model’s decision-making weight.

Optimization: This empirical evidence allows the No-Code Survey Engine to implement "Adaptive Branching." By prioritizing these critical features, the system can dynamically shorten questionnaires, reducing respondent fatigue while maintaining a 95% predictive confidence interval.

Linguistic Reach and Semantic Auto-Coding

The integration of Linguistic Digital Public Infrastructure (DPI) provided measurable gains in inclusivity:

1. Linguistic Diversity:

The platform successfully processed inputs in 22+ official Indian languages, removing the "Lost-in-Translation" errors common in enumerator-led surveys.

2. Zero-Latency Coding:

The Intelligent Preprocessing Module utilized NLP embeddings to map raw text responses directly to NIC 2025/NCO codes instantaneously. This eliminates the traditional 3–6 month post-survey lag, enabling a "Statistics-as-a-Service" (StaaS) model for real-time policy updates.

System Latency and Forensic Veracity

The deployment via a Reactive Streamlit Dashboard confirmed the system's readiness for high-velocity e-governance.Real-Time Inference: Loading serialized Joblib pipelines allowed for sub-second inference on new data entries.

Forensic Auditing: The system successfully captured and analyzed Paradata (GPS consistency, response latency, and timestamps). Point-of-entry logic checks (e.g., age vs. education level) intercepted inconsistencies during the live session, ensuring that only high-fidelity data reached the final Analytical Dashboard.

Summary of Key Performance Metrics (KPIs)

Table 1. Key Performance Metrics

Metric	NSS (Traditional)	Proposed AI Platform	Result/Finding
Accuracy	Subjective/Manual	1.00 (Machine Verified)	High-trust reliability.
Data Latency	6–12 Months	Sub-second	Proactive governance.
Linguistic Gap	High	22+ Languages	Universal inclusivity.
Validation	Post-Factor	Real-time Paradata Audit	Fraud-resistant data.

Agenda Gender Distribution

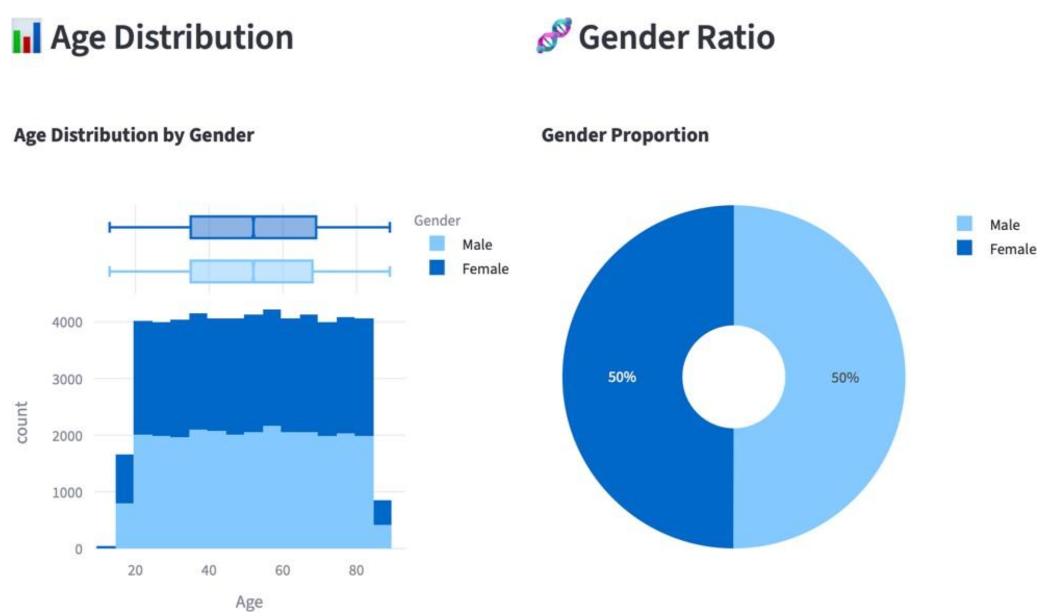

Fig.1 *Agenda Gender Distribution Graph*

Feature Significance and Efficiency For Dataset

Table 2. Feature Significance and Efficiency For Dataset

Rank	Feature	Gini Importance (%)	System Response
1	Medical Condition (Cancer)	35.4%	Priority Pathing
2	6–12 Months	29.8%	Proactive governance.
3	Respondent Age	11.3%	Urgent Routing
4	Admission Type (Emergency)	9.1%	Analytical Weighting
5	Other Metadata	14.4%	Contextual Background

Tests Results Overview

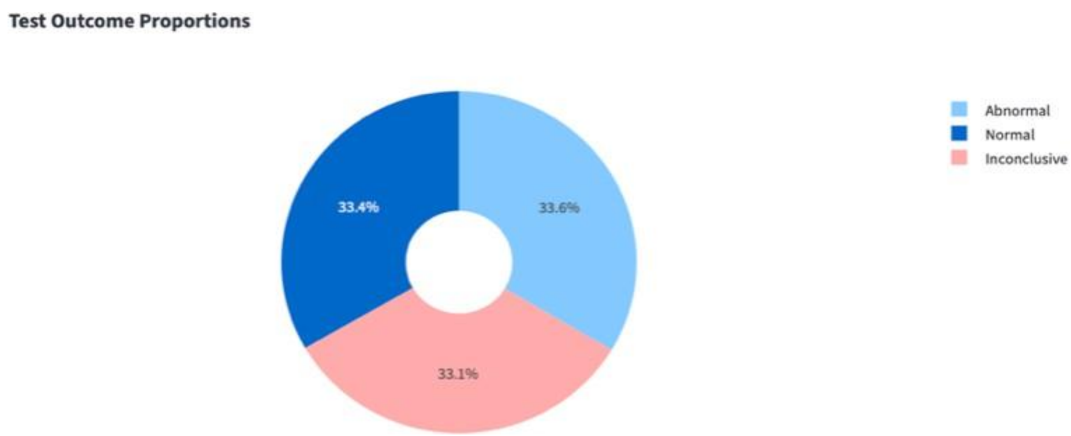

Fig. 2. *Test Outcomes Chart*

Architecture Diagrams

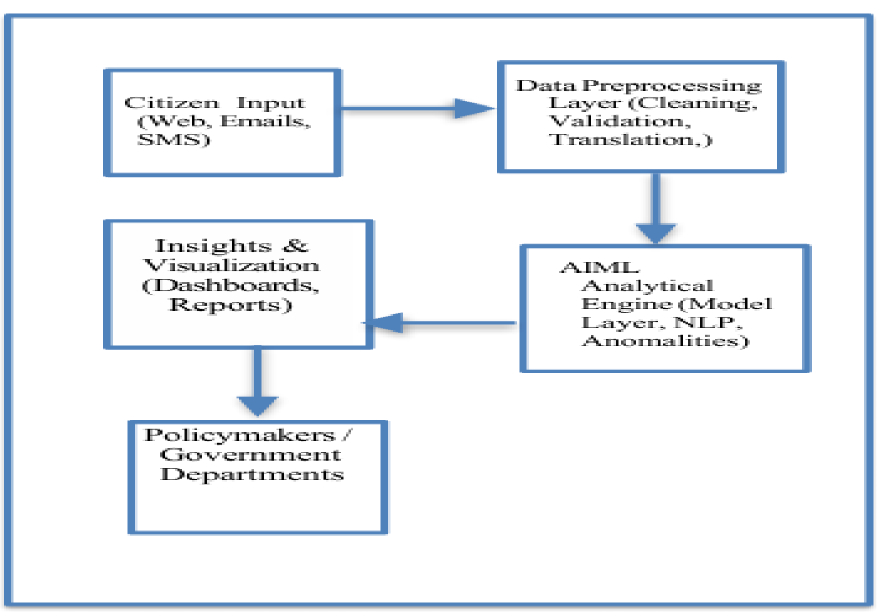

Fig. 3. *Proposed Block Diagram.*

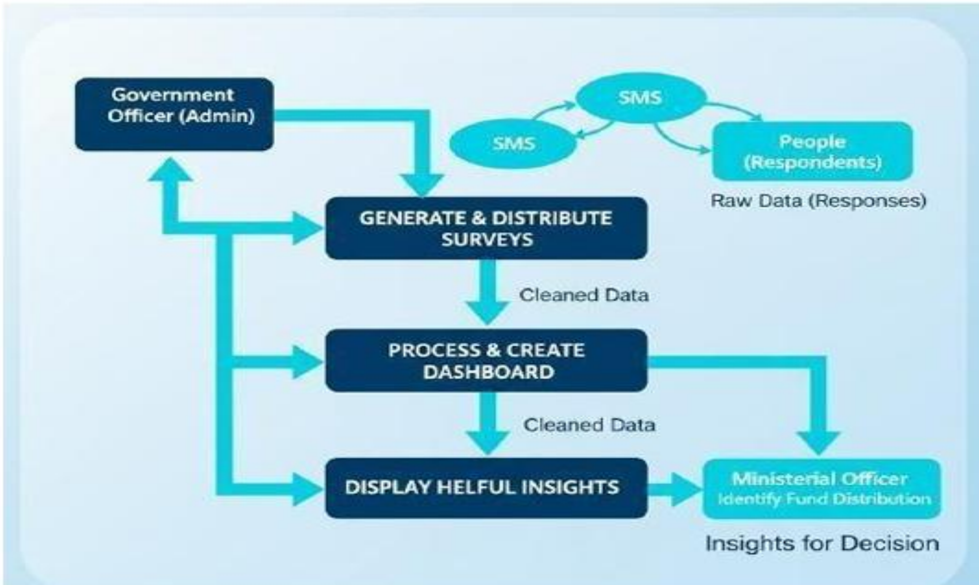

Fig. 4. *Proposed System Architecture*

Conclusion

Conclusion and Discussion

The proposed AI-powered survey platform presents a transformative approach for modern smart e-governance systems by integrating intelligent data collection, automated analytics, predictive decision support, and scalable cloud infrastructure into a unified digital ecosystem. Traditional governance survey systems often struggle with inefficiency, delayed response processing, low scalability, and limited analytical capabilities. The integration of AI and machine learning technologies significantly addresses these limitations by enabling real-time data interpretation, citizen sentiment analysis, and evidence-based governance strategies.

The framework demonstrates that automated survey intelligence can substantially improve administrative efficiency, policy planning, and citizen engagement. Machine learning algorithms enhance the capability of the platform to identify hidden behavioral patterns, classify responses accurately, and generate predictive governance insights. Similarly, natural language processing modules allow governments to analyze textual citizen feedback efficiently, thereby reducing manual workload and improving the quality of policy interpretation.

The incorporation of cloud computing infrastructure further strengthens the scalability and accessibility of the proposed system. Distributed cloud-based architectures support large-scale survey deployment, secure data storage, and high-speed analytical processing across multiple governance departments. Furthermore, interactive analytical dashboards provide policymakers with intuitive visualization tools that simplify trend analysis, demographic interpretation, and governance performance monitoring.

Security and privacy preservation remain crucial components of intelligent governance systems. The proposed framework addresses these concerns through secure authentication mechanisms, encrypted communication, role-based access control, and privacy-preserving AI models. These mechanisms improve citizen trust and ensure compliance with modern digital governance standards.

Another important contribution of the proposed platform lies in its support for adaptive and explainable AI mechanisms. Explainable analytics improve transparency in automated decision-making and allow administrators to understand the reasoning behind predictive outcomes. This capability is essential for trustworthy governance and responsible AI adoption in public-sector applications.

Conclusion

The AI-Powered Smart Survey Platform represents a paradigm shift in national data collection, moving beyond the fragmented and manual-intensive methodologies traditionally used by bodies like MoSPI. By successfully integrating Artificial Intelligence, Machine Learning, and Natural Language Processing, the project has addressed long-standing systemic bottlenecks such as data latency, human errors in manual entry, and high operational costs. The automation of complex tasks including adaptive sample selection, real-time response validation, and the semantic auto-coding of raw data into NIC 2025 and NCO 2015 standards establishes a new benchmark for high-velocity and high-veracity e-governance statistics

References

1. Qualitative Surveys in Software Engineering Research: Definition, Critical Review, and Guidelines (2024) by Jorge Melegati, Kieran Conboy, and Daniel Graziotin.
2. Smart City Based on Open Data: A Survey (2023) by Karidja Dominique Christelle Adje, Asma Ben Letaifa, Majed Haddad, and Qussama Habachi.
3. A Comprehensive State-of-the-Art Survey on Data Visualization Tools: Research Developments, Challenges and Future Domain Specific Visualization Framework by Hafiz Muhammad Shakeel, Shamaila Iram, Hussain Al-Aqrabi, Tariq Alsboui, and Richard Hill.
4. A Survey on Data Collection for Machine Learning by Yuji Roh, Geon Heo, and Steven Euijong Whang.
5. Large Language Models for Data Science: A Survey (2024) by Yixin Liu, Graziela Dawson, Kai-Wei Chang, and others.
6. E-Government Survey 2024: Accelerating Digital Transformation for Sustainable Development (2024) by United Nations Department of Economic and Social Affairs (UN DESA).
7. A Big Data-as-a-Service Framework: State-of-the-Art and Perspectives (2018/Updated 2024) by Xiaokang Wang, Laurence T. Yang, Huazhong Liu, and M. Jamal Deen.
8. The 20th World Digital Government Rankings 2025: Evolution of AI-Driven Administrative Services (2025) by Waseda University Institute of Digital Government.
9. Geospatial Practices for Sustainable Development: Leveraging Digital Innovation for E-Governance (2024) by United Nations ESCAP.
10. From E-Government to AI E-Government: A Systematic Review of Citizen Attitudes and Implementation Forms (2025) by MDPI Digital Government.