Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

SchemeImpactNet: Forecasting and Optimizing MGNREGA Employment Generation Using Machine Learning

Muskan Shaikh, Sameer Sayyad, Ashish Lakhimale, Ayush Bag

Department of Computer Engineering,Siddhant College Engineering, Pune, India

----------***----------

ABSTRACT

The Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) represents a cornerstone of India's social security infrastructure, disbursing substantial fiscal re-sources annually to sustain rural livelihoods. However, traditional management practices within public welfare administration re-main largely reactive, relying heavily on historical baselines and manual planning paradigms that exhibit structural limitations in predicting localized demand spikes and dynamic climate disruptions. This paper introduces SchemeImpactNet, a data-driven predictive and optimization framework engineered to transition public welfare administration from reactive manage-ment to evidence-based proactive policy intervention. Leveraging an extensive administrative dataset comprising 7,758 district-year records across 759 districts and 34 states/union territories from 2014–15 to 2024–25, we evaluate a series of rigorous predictive machine learning algorithms under a strictly leak-free temporal validation protocol. Gradient Boosting models achieve a cross-validated mean R2 of 0.9078 (excluding anomaly years), outperforming alternative parametric and non-parametric algorithms. Utilizing these localized predictions, we model a linear programming resource optimization engine using the PuLP framework to maximize public employment outcomes under fixed historical budgetary parameters. Experimental results establish that mathematically optimized resource reallocation across ad-ministrative subdivisions yields a simulated employment gain of 1,840.41 lakh person-days, representing a 6.38% net im-provement in nationwide execution efficiency without increasing nominal public expenditure. The architecture is containerized and deployed via an open REST API and interactive Streamlit service, establishing reproducible methodology for public policy optimization.

Index Terms—Public Policy Analytics, Predictive Modeling, Gradient Boosting, Resource Allocation Optimization, Linear Programming, MGNREGA.

I.INTRODUCTION

India coordinates one of the most comprehensive public welfare networks globally, encompassing over 400 central government schemes designed to buffer vulnerable populations against socio-economic instabilities. Foremost among these initiatives is the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), a legislative rights-based wage employment mechanism spending over Rs. 70,000–90,000 crore annually. By guaranteeing 100 days of manual wage employment per financial year to rural households, the scheme structurally directly affects hundreds of millions of citizens. Despite its critical relevance to agricultural stability, climate resilience, and economic equity, contemporary administrative infrastructure exhibits major operational inefficiencies. Alloca-tion and management methodologies remain structurally reac-tive. Resource dissemination across state and district planner cells is heavily dependent on retrospective manual account-ing paradigms. Administrators typically audit prior financial data to adjust budgets after implementation shortfalls occur, failing to model localized future demands or dynamic climate anomalies.

This manual allocation schema introduces severe spatial and temporal imbalances. Certain agrarian administrative

sub-divisions suffer acute budgetary deficits during periods of localized drought, leading to employment rationing, while adjacent subregions exhibit capital underutilization due to rigid, unoptimized fiscal limits. Furthermore, contemporary analytics systems evaluate government welfare programs in absolute operational isolation, ignoring complex inter-scheme dependencies. For example, the household penetration of agricultural income stabilizers like the Pradhan Mantri Kisan Samman Nidhi (PM-KISAN) or housing asset additions via the Pradhan Mantri Awas Yojana (PMAY-G) alter the underlying structural demand for manual wage employment under MGN-REGA. Without quantifying these systemic relationships, state interventions encounter persistent over-allocation or localized delivery deficits.

To mitigate these analytical limitations, this research intro-duces SchemeImpactNet, an intelligent, data-driven machine learning and mathematical optimization framework engineered to automate public policy intelligence. By framing public scheme management as a multi-stage data science pipeline, the system ingests, cleans, and structures authentic government ad-ministrative data into a leak-free panel dataset spanning more than a decade. It systematically benchmarks spatial operational performance and generates highly accurate predictive trajec-tories for future localized labor requirements. Utilizing these empirical predictive baselines, SchemeImpactNet hosts a linear programming model designed to perform automated macro-budgetary optimization, recommending precise adjustments to maximize nationwide employment output without increasing public spending. A. Contributions

The specific academic and operational contributions of this research are defined as follows:

• We compile, validate, and process a comprehensive district-year panel dataset comprising 7,758 adminis-trative observations across 759 districts and 34 In-dian states/UTs from 2014–15 to 2024–25, incorporating multi-scheme data linkages and climate indicators.

• We execute a rigorous comparative study of both paramet-ric linear regressors and non-parametric ensemble tree-based models, utilizing a robust temporal validation split to measure predictive reliability across normal and macro-economic anomaly periods.

• We design a dual-stage mathematical optimization engine utilizing linear programming via the PuLP framework that maximizes total generated employment (person-days) subject to strict state-level fiscal boundaries.

• We deploy a containerized production-ready architecture featuring a high-throughput FastAPI REST backend and an interactive Streamlit presentation layer, augmented with explainable Large Language Model (LLM) inter-faces for non-technical policy planners.

II.RELATED WORK

The incorporation of empirical computation into public welfare systems is a developing domain situated at the con-vergence of statistical modeling, resource allocation, and e-governance. Prior literature exhibits a clear bifurcation be-tween operational tracking platforms and theoretical academic prototypes.

A.Government Analytics and Welfare Dashboards

Early technological interventions in public administration focused primarily on migrating paper records into relational environments. Portals such as the early Government Welfare Portal (2018–19) or the Digital India Portal (2019) introduced baseline digitizations, enabling online application intakes and static data displays via web and mobile interfaces. While these implementations improved data accessibility, they functioned exclusively as descriptive repositories. These portals lack predictive analytics or optimization layers, requiring adminis-trative staff to execute subjective manual evaluations to extract programmatic insights.

B.Machine Learning for Public Policy

Recent research has integrated supervised algorithms into state welfare logistics. Rao et al.

[1] leveraged decision trees, random forests, and k-nearest neighbors to classify agrarian land data, attaining 97.25% accuracy in evaluating household eligibility for farming subsidies. Similarly, Kannan et al.

[2] applied multilayer perceptrons and gradient-boosted trees to predict financial vulnerability within low-income house-holds utilizing post-disaster survey indicators. In a separate approach, Deshpande and Pandurangi [13] employed Naive Bayes and maximum entropy configurations on social media text strings to deduce public sentiment regarding large-scale policy moves such as Digital India and Demonetization.

C.Research Gap

Although predictive classification has been successfully integrated into individual eligibility tracking and sentiment extraction, current systems display substantial research gaps when evaluated against macro-level operational metrics:

1)Absence of Continuous Impact Forecasting:

Existing models focus heavily on binary eligibility classification rather than continuous, population-level regression to estimate specific operational performance milestones.

2)Isolationist Analytical Paradigms:

Contemporary sys-tems evaluate public assistance schemes independently, omitting cross-scheme features and interdependencies.

3)Disconnection Between Prediction and Allocation:

Prior predictive frameworks stop at trajectory forecasting, providing no mathematical mechanism to link projected demand to optimal budgetary distribution models.

4)Deployment Barriers:

Most public policy machine learn-ing research remains confined to academic scripting or local experimental pipelines, lacking production-ready, open-access cloud deployments.

SchemeImpactNet addresses this critical gap by unifying continuous time-series predictive modeling, explicit inter-scheme feature engineering, and linear optimization within a single deployed framework.

III.DATASET AND FEATURE ENGINEERING

To evaluate the framework under authentic settings, we build a multi-source panel dataset mapping regional socio-economic variables across India.

A.Data Sources

Data ingestion extracts administrative metrics from four primary repositories [9]–[12]:

• Ministry of Rural Development (MoRD) MIS Portal: Ex-tracting targeted district-level operational logs for MGN-REGA, detailing realized person-days, capital expendi-ture, and wage structures.

• NITI Aayog National Multidimensional Poverty Index (MPI): Providing structural data regarding regional poverty intensities and baseline deprivation indices.

• India Meteorological Department (IMD): Contributing regional time-series precipitation logs to construct explicit climate vectors.

• Census of India: Supplying core demographic variables, including baseline rural population data.

B.Dataset Characteristics

The uncompressed database holds 7,758 initial longitudi-nal records encompassing 759 districts and 34 states/union territories over an 11-year operational window (2014–15 to 2024–25). The primary target variable is defined as continuous person_days_lakhs, quantifying total manual employ-ment generated per district-year.

C.Data Preprocessing and Feature Engineering

The preprocessing layer executes automated alignment, missing data attribution, and variable transformations. Initial raw data structures are aggregated to annual district-year levels. Rural demographic segments are projected dynamically across inter-censal periods utilizing a localized annual popu-lation growth factor of 1.2%.

To prevent target leakage and isolate causal variables, we generate a high-dimensional feature matrix consisting of 17 key model inputs. The engineering configuration

computes lag metrics, rolling temporal windows, and inter-scheme interac-tion vectors as detailed below:

• Temporal Autoregressive Regressors: Tracking historical demand patterns through localized temporal steps, defined as lag1_pd, lag2_pd, and lag3_pd.

• Moving Average Filters: Quantifying regional baseline momentum through rolling mathematical windows, in-cluding a 2-year mean (roll2_mean), 3-year mean (roll3_mean), and a 3-year standard deviation filter (roll3_std).

• Cross-Scheme Integration Vectors: Integrating household uptake variables from parallel welfare initiatives (PM-KISAN and PMAY-G) computed based on execution dates and national deployment indicators to model struc-tural employment substitution.

• Macro-Economic and Climate Controls: Incorporating Year-over-Year wage volatility (wage_yoy), local base wage rates (avg_wage_rate), and localized precipita-tion anomalies derived from IMD records.

• Anomalous Shock Vectors: Formulating categorical indi-cators (is_covid, lag1_is_covid) to isolate struc-tural economic anomalies during the 2020–2021 global health crisis.

To construct leak-free rolling features, records with insuf-ficient historical baselines are removed via systematic trunca-tion, dropping 1,528 early rows. This outputs a final high-fidelity matrix consisting of 6,230 operational district-year observations covering 736 districts from 2016 to 2024

TABLE I
DATASET STRUCTURAL CHARACTERISTICS

Dataset Parameter	Value/Specification
Temporal Window	Financial Years 2016 – 2024
Geospatial Entities	34 States/UTs — 736 Districts
Total Panel Records	6,230 Observations
Cumulative Realized Target Value	257,636.2 Lakh Person-Days
Engineered Modeling Features	17 Operational Independent Vect
Primary Output Target Metric	Localized Annual Employment

Exploratory analysis reveals extreme spatial distribution variations across regional administrative entities. The top 5 employment-generating subdivisions contribute massive scale, led by 24 Parganas South (296.39 lakh person-days) and Barmer (266.89 lakh person-days). Conversely, the lowest 5 subdivisions exhibit near-zero manual demand, typified by South Goa (0.36 lakh person-days) and Ghaziabad (0.00 lakh person-days).

IV.PROPOSED FRAMEWORK

SchemeImpactNet relies on an integrated, multi-tier compu-tational architecture that coordinates data ingestion, predictive inference, optimization routing, and client presentation.

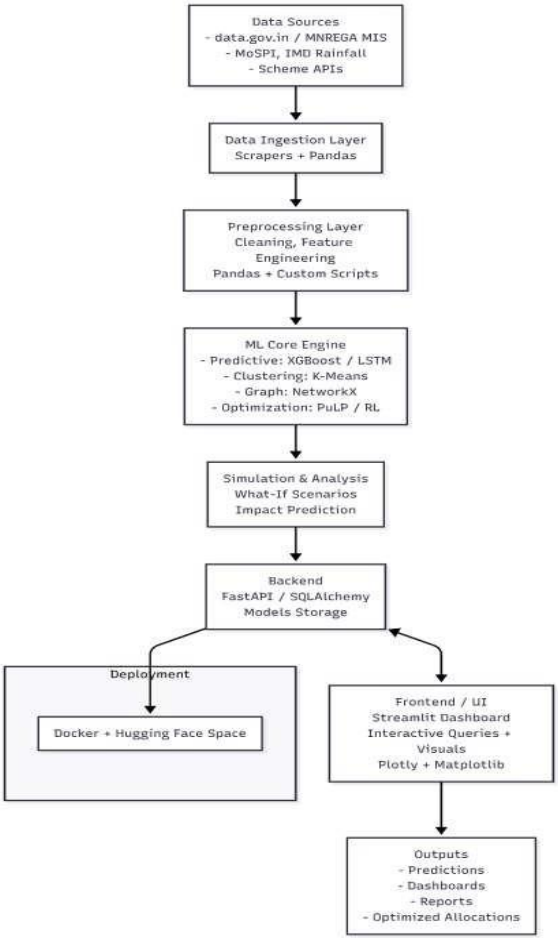

Fig. 1. System Architecture of SchemeImpactNet, detailing the data extraction pipeline, supervised learning configurations, optimization solvers, and client delivery layers.

As shown in Fig. 1, raw public records are extracted from source APIs and flat CSV modules into an offline staging database. These sources pass sequentially through automated feature engineering transformations wrapped in a scikit-learn pipeline object to prevent target leakage during runtime. The downstream core engine branches into parallel analytics pipelines: a predictive regression layer for demand forecasting, an unsupervised clustering layer for efficiency tracking, and an analytical budget optimizer managed via linear programming routines.

A.Predictive Modeling Layer

The supervised learning pipeline runs multiple ensemble implementations to model the targeted continuous labor de-mand. Non-parametric tree models utilize localized structural partitions to map non-linear climate anomalies and cross-scheme relationships. Given input feature vector Xi,t for district i in year t, the models are optimized to estimate the future target value yi,t+1.

B.Mathematical Optimization Engine

The framework translates localized predictive outputs into macroeconomic resource adjustments via a constrained lin-ear programming optimization structure modeled using PuLP. Rather than applying unoptimized linear adjustments based on legacy allocations, the optimization model systematically evaluates the marginal employment returns across districts to maximize state-level execution metrics.

Let i ∈ {1, 2, . . . , N } denote individual district entities within a specific regional state boundary. Let Bi represent the historical status-quo budget allocated to district i, and let xi define the optimized budget allocation to be computed by the LP model. Let yˆi(xi) denote the predicted employment generation function for district i modeled as a function of the local resource allocation.

The objective function maximizes total simulated labor output across all districts within the state boundary:

$$\max_{x} \sum_{i=1|}^{N} \hat{y}_i(x_i) \quad (1)$$

Subject to the following operational and financial con-straints:

1)Total Budget Conservation Constraint:

The cumulative optimized fiscal allocation cannot exceed the fixed his-toric total state budget allocation:

$$\sum_{i=1}^{N} x_i \leq \sum_{i=1}^{N} B_i \quad (2)$$

2)Proportional Allocation Boundaries:

To prevent severe regional destabilization and maintain operational com-pliance, the optimization engine restrains budget adjust-ments within a dynamic scaling factor relative to the baseline historical values:

$$(1 - \alpha)B_i \leq x_i \leq (1 + \beta)B_i \quad (3)$$

Where α and β define operational parameters set dynamically during the first stage of the optimizer execution to mirror localized infrastructure capacities.

V.EXPERIMENTAL SETUP

To ensure statistical reproducibility, the experimentation framework is deployed under an open-source Docker config-uration running identical python dependencies.

A.Execution Environment

The software pipeline is developed using Python 3.10+. Ad-vanced supervised implementations rely on XGBoost 3.2.0 and scikit-learn 1.8.0. Data manipulation and numerical operations run over Pandas 3.0.1 and NumPy 2.4.2, while the linear pro-gramming solver relies on the PuLP optimization ecosystem. Hardware requirements map to baseline commodity hardware, specifying a minimum of 4 GB RAM and standard multi-core processing configurations.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

B.Temporal Validation Strategy

Traditional randomized cross-validation introduces severe target leakage when applied to longitudinal panel data due to the propagation of forward temporal information into past predictions. To address this, SchemeImpactNet uses a rigorous time-forward walk-forward temporal cross-validation protocol. The dataset is partitioned sequentially by financial years, training models exclusively on historical chronological subsets (e.g., $t \in [2016, T]$) to predict adjacent out-of-sample forward years ($T + 1$). Testing runs across consecutive out-of-sample periods from 2018 to 2024 to verify model performance across distinct socio-economic conditions.

C.Evaluation Metrics

Model reliability is quantified across three independent statistical validation metrics:

• Coefficient of Determination (R2 Score): Measures the proportion of target variance captured by the model relative to a naive historical baseline.

• Mean Absolute Error (MAE): Quantifies the average absolute predictive error magnitude, expressed in lakhs of person-days.

• Root Mean Squared Error (RMSE): Tracks variance in prediction deviations, applying heavier penalties to large localized outlier errors.

VI.RESULTS AND DISCUSSION

The experimental results demonstrate substantial variations in predictive accuracy across linear baselines and advanced non-parametric configurations.

A.Predictive Model Comparison

Table II details the comprehensive cross-validated performance breakdown across seven continuous test years.

The comparative analysis demonstrates that ensemble tree-based architectures outperform parametric linear alternatives across almost all out-of-sample periods. Under stable eco-nomic conditions, the primary models attain high performance, with Gradient Boosting reaching an R2 score of 0.9345 and XGBoost attaining 0.9334 in the 2024 evaluation window.

B.Macroeconomic Shock and Error Analysis

Evaluating model performance across the complete time series highlights a significant drop in accuracy during the 2022 financial year, where Gradient Boosting's localized R2 falls to

TABLE II

COMPREHENSIVE CROSS-VALIDATED ALGORITHM PERFORMANCE BREAKDOWN (2018 – 2024)

Algorithm Configuration	2018	2019	2020	2021	2022	2023	2024	Mean R^2 (Full)	Mean R^2 (Excl. 2022)
GradientBoosting	0.9160	0.9262	0.8354	0.9261	0.5101	0.9089	0.9345	0.8510	**0.9078**
RandomForest	0.9146	0.9299	0.8310	0.9194	0.4541	0.9147	0.9280	0.8417	0.9063
XGBoost	0.9121	0.9396	0.8054	0.9259	0.5528	0.9042	0.9334	**0.8533**	0.9034
Ridge Linear Regression	0.8459	0.9141	0.8117	0.9226	0.3177	0.9067	0.8936	0.8018	0.8824
ElasticNet Regularization	0.8534	0.9184	0.8088	0.9265	0.3012	0.8809	0.8985	0.7982	0.8811

0.5101 and RandomForest degrades to 0.4541. This system-wide drop stems from post-COVID macroeconomic correc-tions. During the 2020–2021 pandemic years, administrative entities registered massive, unprecedented spikes in rural labor request volumes due to widespread urban displacement, caus-ing total person-days to surge to a mean of 55.01 lakh per subdivision. The subsequent 2022 period encountered highly volatile structural contraction trends as labor forces returned to urban environments, generating an unpredictable anomaly year that diverged from historical autoregressive lags. When excluding this extreme anomaly period, Gradient Boosting achieves the highest predictive metric with a cross-validated mean R2 of 0.9078 and a Mean Absolute Error of 8.554 lakh person-days.

C.Feature Importance Analysis

Plotting the structural decision paths of the champion Gra-dient Boosting configuration reveals the underlying feature weights driving localized forecasting outputs:

• lag1_pd (Relative Importance Score: 0.5270): Operat-ing as the dominant predictive vector, confirming that short-term historical labor demand remains the primary structural anchor for future projections.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

• lag1_adj (Relative Importance Score: 0.2512): Providing substantial predictive value by tracking immediate localized target volatility.

• state_lag1_zscore (Relative Importance Score: 0.0837): Quantifying relative district performance vari-ance compared to state-level baseline trends.

• roll2_mean (Relative Importance Score: 0.0612): Smoothing localized short-term variance to capture mid-term structural momentum.

• blended_capacity (Relative Importance Score: 0.0199): Establishing the underlying operational capacity limits of regional administrative bodies.

D.Budget Optimization Evaluation

The automated resource allocation engine was evaluated using out-of-sample records for 706 active districts in the 2024 financial year. The baseline status-quo allocation model rep-resents the real-world administrative budget tracking schema, which matches actual expenditures of Rs. 9,802,710 lakh.

As mapped out in the DFD shown in Fig. 2, the primary inputs run through the system pipeline to link predictive inference with the optimization engine. Under the status-quo

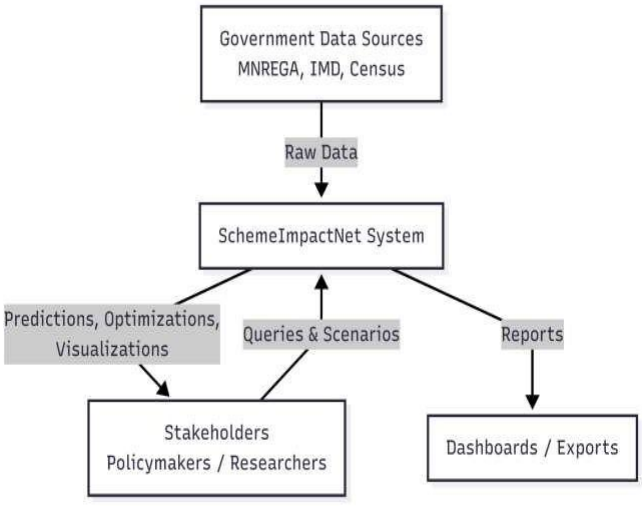

Fig. 2. Data Flow Diagram mapping data pipelines, prediction routines, and output layers within SchemeImpactNet

budgetary distribution model, the 706 districts generated a total labor volume of 28,829.38 lakh person-days. Without changing the total nominal expenditure of Rs. 9,802,710 lakh, the PuLP linear optimizer identifies and leverages regional efficiency variations, redirecting funds toward high-marginal-impact sub-divisions. This mathematical reallocation increases simulated performance to 30,669.79 lakh person-days. This represents a net simulated employment gain of 1,840.41 lakh person-days, demonstrating a 6.38% increase in delivery efficiency achieved solely through algorithmic resource allocation.

Table III shows the targeted adjustments computed by the optimizer. The engine diverts funding from regions with declining marginal returns or historical over-allocation to high-efficiency centers. For instance, Jhalawar receives a 77.96% budget increase, yielding an employment expansion of 110.755 lakh person-days. Conversely, over-allocated centers like Barmer are scaled back by 26.35%, allowing excess cap-ital to be optimized and deployed elsewhere without dropping below required baseline targets.

VII.DEPLOYMENT ARCHITECTURE

To transition these empirical findings into active policy-making tools, SchemeImpactNet is deployed publicly on Hugging Face Spaces. The web-based application interface provides specific analytical access points tailored to different administrative users. Policymakers interact directly with sce-nario panels to run what-if simulations and evaluate optimized.

TABLE III

STRATEGIC BUDGET REALLOCATION ACTIONS (TOP INCREASES VS. TOP DECREASES)

State	District	Base Budget (Rs. Lakhs)	Opt. Budget (Rs. Lakhs)	Net Gain (PD Lakhs)
Recommended Budget Augmentations (Top 5 Districts)				
Rajasthan	Jhalawar	34,513.79	61,420.08	+110.755
Rajasthan	Bhilwara	63,021.65	96,309.49	+109.616
Rajasthan	Jodhpur	43,157.28	72,509.08	+105.309
Odisha	Ganjam	22,347.28	40,981.25	+96.170
Tamil Nadu	Chengalpattu	22,102.08	40,171.85	+84.506
Recommended Budget Reductions (Top 5 Districts)				
Rajasthan	Udaipur	60,574.80	32,307.89	-66.892
Andhra Pradesh	Alluri Sith.	54,797.30	27,541.70	-59.411
Tamil Nadu	Villupuram	60,813.40	34,873.88	-64.341
Tamil Nadu	Cuddalore	53,050.01	28,436.33	-58.703
Rajasthan	Barmer	92,571.15	68,182.19	-68.106

budget reallocations. Researchers navigate through geospatial dashboards to evaluate error indices, while administrators maintain backend configuration endpoints to update raw data files and export policy briefs.

The system runs on a containerized Docker multi-tier con-figuration. The presentation frontend is built with Streamlit, rendering interactive geospatial trend lines and dashboard scorecards powered by Plotly. Planners can filter execution states, trace historical district performance, and run simulated policy scenarios. The frontend communicates directly with a FastAPI backend REST service, which exposes optimized endpoints (/predict, /efficiency, /optimize) that return low-latency JSON payloads. To make complex data science metrics interpretable for non-technical administrators, the deployment integrates Google Generative AI (Gemini) interfaces. This allows users to evaluate optimization results using natural language syntax (e.g., querying specific impacts if state allocations shift by a fixed percentage), auto-generating natural language explanations alongside the mathematical out-puts

VIII.LIMITATIONS AND FUTURE WORK

While SchemeImpactNet delivers high performance and measurable optimization gains, it exhibits structural constraints that present opportunities for future research:

• Data Dependency Inconsistencies: Certain historical indi-cators, such as regional precipitation figures from 2018–2024 and specific parallel welfare records, were derived using statistical estimations due to primary source report-ing gaps.

• Static Linear Modeling Assumptive Bounds: The LP engine assumes a constant linear relationship between budget changes and generated person-days, omitting non-linear constraints caused by sudden local administrative bottlenecks.

• Lack of Real-Time Pipelines: The current pipeline uses batch-loaded data files, requiring manual updates rather than direct synchronization with live government MIS portals. Future research will focus on replacing the static linear programming configuration with a multi-period Reinforce-ment Learning agent capable of optimization under weather anomalies and financial uncertainty. Additionally, we plan to incorporate satellite imagery and deep learning vegetation indexes to improve predictive accuracy within ecologically vulnerable regions.

IX.CONCLUSION

This paper presented SchemeImpactNet, a reproducible data science framework developed to transform public welfare administration from descriptive historical management into proactive, evidence-based decision support. By organizing and evaluating an extensive panel dataset of 7,758 district-year observations within a strict walk-forward temporal cross-validation protocol, we demonstrate that non-parametric Gra-dient Boosting configurations achieve a robust mean R2 of 0.9078, outperforming classic parametric linear formulations. By connecting these continuous localized forecasts with a linear programming optimization engine, the framework estab-lishes that mathematically optimized resource reallocation can expand nationwide welfare delivery by 1,840.41 lakh person-days—a 6.38% net efficiency gain achieved without increasing nominal public spending. The containerized web deployment, augmented with Large Language Model interpretation layers, confirms that complex data analytics can be effectively trans-lated into accessible tools for public administrators, advancing the objectives of Digital India and transparent, data-driven governance.

REFERENCES

[1]D. Srinivasa Rao, A. Krishna Teja, V. Manoj Kumar, A. Surya Teja, Ch. Ramesh Babu, and A. Sai Karthik, “Machine Learning Driven Classification of Farmer Land Data for Government Scheme Eligibility,” 6th IEEE International Conference on Recent Advances in Information Technology (RAIT), 2025, DOI: 10.1109/RAIT65068.2025.11088933.

[2]R. Kannan, K. W. Shing, K. Ramakrishnan, H. B. Ong, and A. Alam-syah, “Machine Learning Models for Predicting Financially Vigilant Low-Income Households,” IEEE Access, vol. 10, pp. 70418–70427, 2022, DOI: 10.1109/ACCESS.2022.3187564. [3]T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.

[4]J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[5]L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[6]F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[7]S. Mitchell, "PuLP: A Linear Programming Toolkit in Python," Techni-cal Report, University of Auckland, Auckland, New Zealand, 2011.

[8]Savitribai Phule Pune University, BE Computer Engineering Project Guidelines 2023-24, Department of Computer Engineering, Siddhant College of Engineering, Sudumbare.

[9]Ministry of Rural Development, Government of India, "MGNREGA Management Information System (MIS)," Available: https://nreganarep. nic.in.

[10]NITI Aayog, Government of India, "National Multidimensional Poverty Index: Progress Review Report," New Delhi, 2023.

[11]India Meteorological Department, "Historical District-wise Rainfall Gridded Datasets (2014-2024)," Ministry of Earth Sciences, Government of India.

[12]Office of the Registrar General & Census Commissioner, Ministry of Home Affairs, Government of India, "Census of India 2011: Primary Census Abstract," New Delhi, 2011.

[13]M. Deshpande and V. Pandurangi, "Measuring Effectiveness of e-Governance Initiatives via Machine Learning Approaches on Sentiment Expression," International Conference on Advances in Computing and Communications, 2018.

[14]G. B. Dantzig, Linear Programming and Extensions, Princeton Univer-sity Press, 1963.

[15]SchemeImpactNet GitHub Repository, https://github.com/sameerdev7/ SchemeImpactNet.

[16]SchemeImpactNet Deployed Application, https://huggingface.co/spaces/ sammeeer/SchemeImpactNet.