

## Integration of Individual Data and Family Cards in Optimizing Tax Revenue: a Public Accounting and Predictive Analytics Approach to Regional Development Planning

Sarsiti Sarsiti<sup>1\*</sup>, Tamam Rosid<sup>2</sup>, Juli Prastyorini<sup>3</sup>, Putri Nilam Aisyah<sup>4</sup>

<sup>1,4</sup> Faculty of Economics, University of Surakarta

<sup>2</sup> Muhammadiyah University of Berau, East Kalimantan,

<sup>3</sup> STIAMAK Barunawati Surabaya,

**Corresponding Author:** Author name: sitiunsa76@gmail.com

---

### ARTICLE INFO

*Keywords:* Tax Revenue Optimization, Data Integration, Predictive Analytics, Public Accounting, Regional Development Planning, Family Registration System

*Received :* 24 October 2025

*Revised :* 24 December 2025

*Accepted:* 25 January 2026

©2026 Sarsiti, Rosid, Prastyorini, Aisah: This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



### ABSTRACT

This study examines the integration of individual population data with family registration records to optimize regional tax revenue in Indonesia. Using a mixed-methods approach across three districts (2021–2024), the findings show that integrated data systems increase tax compliance by 23–31% and expand the tax base by 18–26%. Predictive analytics using machine learning achieved 84% accuracy in identifying potential taxpayers and 78% precision in predicting payment behavior. Data integration also reduced administrative costs by 35–42% and improved revenue forecasting accuracy from 68% to 89%. Despite challenges related to legal frameworks, interoperability, and institutional capacity, integrated data systems enhance tax administration efficiency and support more equitable regional development planning.

## **INTRODUCTION**

Regional tax revenue constitutes a fundamental pillar of fiscal decentralization and autonomous development planning in emerging economies (Bird & Slack, 2023; Martinez-Vazquez et al., 2021). However, inefficiencies in tax administration systems, characterized by fragmented data infrastructure and limited analytical capabilities, significantly constrain revenue mobilization efforts (Prichard et al., 2019). The persistent gap between potential and actual tax collection estimated at 35-50% in many developing regions reflects systemic weaknesses in taxpayer identification, assessment accuracy, and compliance monitoring (Mascagni et al., 2023).

Contemporary public finance literature increasingly emphasizes the transformative potential of digital technologies and data analytics in modernizing tax administration (Okunogbe & Santoro, 2023). The integration of population registration systems with tax databases represents a particularly promising avenue for enhancing revenue performance (Chen et al., 2022). Family card (Kartu Keluarga) systems, which maintain comprehensive household demographic and socioeconomic information, offer untapped potential for improving taxpayer profiling and expanding the tax net (Purwanto & Kusumawardani, 2024).

Recent advances in predictive analytics and machine learning provide sophisticated tools for analyzing integrated datasets to identify tax compliance patterns, predict revenue streams, and detect evasion risks (Alm & Soled, 2022). When combined with public accounting principles emphasizing transparency, accountability, and fiscal sustainability, these technological capabilities can fundamentally transform subnational tax administration (Khanh & Khanh, 2023).

Despite growing recognition of data integration benefits, empirical evidence on implementation strategies, measurable outcomes, and contextual adaptation remains limited (Hassan & Rahman, 2023). This research addresses these gaps by examining how integrated individual-household data systems enhance tax revenue optimization and inform regional development planning. The study focuses on Indonesian local governments, where administrative decentralization has created both opportunities and challenges for tax mobilization (Sjahrir et al., 2023).

The Indonesian context provides valuable insights given its complex multi-tiered governance structure, diverse socioeconomic conditions, and ongoing digitalization initiatives (Fadlurrohim et al., 2022). Local governments manage various tax instruments, including property tax (Pajak Bumi dan Bangunan), hotel and restaurant tax, vehicle tax, and other local levies, while grappling with informal sector prevalence and limited enforcement capacity (Gunawan & Salomo, 2023).

This study employs a mixed-methods research design integrating quantitative analysis of administrative data, predictive modeling experiments, and qualitative policy assessment. We analyze tax collection data, demographic databases, and family registration records from three Indonesian districts representing diverse urban-rural characteristics over the 2021-2024 period. Advanced analytics techniques, including machine learning classification,

regression modeling, and network analysis, are applied to assess integration impacts on revenue performance, compliance behavior, and planning effectiveness.

## **LITERATURE REVIEW**

### **2.1 Tax Administration Modernization and Digital Integration**

The digitalization of tax administration has emerged as a critical reform strategy for improving revenue mobilization in developing countries (Gupta et al., 2021). Studies demonstrate that electronic filing systems, digital payment platforms, and automated data matching significantly enhance compliance rates while reducing administrative costs (Dzingirai & Gandiwa, 2023). Okunogbe and Santoro (2023) found that implementing electronic billing systems in Nigeria increased tax collection by 32% through improved transparency and reduced corruption opportunities.

Data integration across government agencies represents an advanced form of digitalization with multiplicative benefits for tax administration (Kleven et al., 2023). By connecting income registries, property databases, business licenses, and population records, tax authorities can construct comprehensive taxpayer profiles enabling accurate assessment and targeted enforcement (Beer et al., 2022). Cross-referencing mechanisms facilitate the identification of discrepancies between declared income and lifestyle indicators, significantly improving audit efficiency (Alstadsæter et al., 2022).

### **2.2 Population Registration Systems and Tax Compliance**

Family registration systems serve multiple governance functions, including identity verification, civil documentation, and social protection targeting (Gelb & Metz, 2023). Research indicates that robust population databases correlate with higher tax compliance through enhanced government capacity for taxpayer identification and monitoring (Besley & Persson, 2023). In Thailand, the integration of national ID cards with tax registration expanded the formal taxpayer base by 28% within three years (Thongmee & Srivannaboon, 2022).

The household-level information captured in family card systems provides valuable socioeconomic indicators for tax potential assessment (Nkurunziza et al., 2023). Demographic composition, education levels, employment status, asset ownership, and residential characteristics contained in these records enable sophisticated taxpayer segmentation and risk profiling (Manolo et al., 2024). However, data quality challenges, including incomplete registration, delayed updates, and inconsistent standards, limit analytical utility in many contexts (Philbrick et al., 2022).

### **2.3 Predictive Analytics in Public Finance**

Machine learning applications in public sector financial management have demonstrated substantial value for revenue forecasting, expenditure optimization, and fraud detection (Issa et al., 2022). Supervised learning algorithms, including decision trees, random forests, and neural networks, achieve high predictive accuracy for tax compliance behavior when trained on

historical taxpayer data (Albuquerque & Garzón, 2024). Mascagni et al. (2023) employed gradient boosting models to predict property tax payment likelihood in Rwanda with 82% accuracy, enabling more effective enforcement resource allocation.

Natural language processing and text mining techniques extract valuable insights from unstructured administrative documents, taxpayer correspondence, and audit reports (Chen & Tsai, 2023). Network analysis reveals patterns of tax evasion schemes and identifies high-risk taxpayer connections (Nicolò et al., 2024). Unsupervised clustering methods segment taxpayers based on behavioral patterns, facilitating customized communication strategies and service delivery (Kumar & Srivastava, 2023).

#### **2.4 Public Accounting Principles in Local Government**

Contemporary public accounting frameworks emphasize accrual-based systems, performance measurement, and integrated financial reporting to enhance fiscal transparency and accountability (Jones & Caruana, 2022). The shift from cash to accrual accounting enables a more accurate assessment of the government's financial position and long-term sustainability (Van der Hoek, 2023). International Public Sector Accounting Standards (IPSAS) adoption facilitates cross-jurisdictional comparability and strengthens financial management capacity (Bellanca & Vandernoot, 2021).

Revenue recognition principles under accrual systems require sophisticated information systems linking tax assessments, billing, collection, and enforcement activities (Carnegie et al., 2023). Integrated data infrastructure supporting comprehensive revenue cycle management improves both accounting accuracy and operational efficiency (Adhikari et al., 2023). Moreover, granular revenue data enable performance-based budgeting and evidence-informed resource allocation aligned with development priorities (Anessi-Pessina et al., 2022).

#### **2.5 Regional Development Planning and Fiscal Capacity**

Subnational fiscal autonomy critically depends on local revenue mobilization capacity, which directly constrains development planning ambitions (Smoke, 2022). Revenue volatility and unpredictability complicate multi-year planning and infrastructure investment decisions (Martinez-Vazquez & Vulovic, 2023). Enhanced tax collection systems not only increase available resources but also improve revenue forecasting reliability, enabling more strategic development planning (Nath, 2024).

Spatially disaggregated revenue data facilitate evidence-based priority setting and equitable resource distribution across localities (Fajri et al., 2023). Integration of demographic and socioeconomic information with fiscal data enables needs-based allocation formulas and targeted intervention design (Boex & Edwards, 2022). Furthermore, transparent revenue reporting strengthens accountability mechanisms and citizen engagement in planning processes (Khagram et al., 2023).

## METHODOLOGY

### 3.1 Research Design and Data Sources

This study employs a convergent mixed-methods design integrating quantitative analysis of administrative datasets with qualitative assessment of policy implementation. Three Indonesian districts were selected, representing diverse characteristics: one urban district (District A, population 1.2 million), one peri-urban district (District B, population 650,000), and one rural district (District C, population 380,000). The selection ensures variation in economic development levels, administrative capacity, and tax base composition.

Primary data sources include: (1) local tax administration databases containing assessment, billing, collection, and enforcement records for the 2021-2024 period; (2) population and civil registration (PRCR) databases with individual demographic information; (3) family card (Kartu Keluarga) system records capturing household composition and characteristics; and (4) regional development planning documents and budget reports. Secondary data comprises national statistics, policy documents, and relevant administrative regulations.

The quantitative dataset encompasses 2.23 million individual records, 587,000 family units, and 1.47 million tax accounts across the three districts. Variables include demographic characteristics (age, gender, education, occupation), household attributes (size, structure, assets), geographic identifiers, tax types, assessment amounts, payment histories, and compliance indicators. Data quality assurance involved validation protocols, outlier analysis, and consistency checks across integrated sources.

### 3.2 Data Integration Framework

The integration process followed a staged approach: (1) standardization of identification numbers across systems using deterministic and probabilistic matching algorithms; (2) consolidation of individual demographic records with family card household information; (3) linkage of integrated population-household data with tax administration records; and (4) enrichment with geospatial and socioeconomic contextual variables.

Data matching employed a combination of unique identifiers (national ID numbers, family registration numbers) and fuzzy matching techniques for records lacking consistent identifiers. The integration achieved a 94.3% match rate for individual-to-household linkage and 87.6% for household-to-tax account linkage. Unmatched records were analyzed to assess bias implications.

Privacy protection measures included data anonymization, role-based access controls, and compliance with national data protection regulations. The research protocol received ethics approval from the institutional review board and data access permissions from participating district governments.

### 3.3 Analytical Methods

**Descriptive and Comparative Analysis:** Statistical summaries characterized taxpayer profiles, revenue composition, and compliance patterns before and after data integration implementation. Difference-in-differences

estimation assessed causal impacts on collection efficiency, comparing treatment districts implementing integration with control areas maintaining separate systems.

**Predictive Modeling:** Machine learning classification models were developed to predict tax compliance likelihood and identify potential taxpayers. The dataset was partitioned into training (70%), validation (15%), and testing (15%) subsets. Five algorithms were compared: Logistic Regression, Decision Tree, Random Forest, Gradient Boosting Machine (GBM), and Extreme Gradient Boosting (XGBoost). Hyperparameter optimization employed grid search with cross-validation.

Model performance evaluation utilized multiple metrics: accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). Feature importance analysis identified key predictors of compliance behavior. Revenue forecasting employed time series models (ARIMA, Prophet) and regression techniques incorporating integrated demographic-fiscal data.

**Network Analysis:** Graph-based methods mapped relationships between family members, business entities, and property ownership to detect potential evasion schemes and identify high-value taxpayer networks. Community detection algorithms segmented taxpayers into clusters with similar characteristics.

**Qualitative Assessment:** Semi-structured interviews with 47 tax administrators, IT specialists, and planning officials explored implementation challenges, perceived benefits, and organizational change dynamics. Thematic analysis identified recurring patterns and critical success factors.

### 3.4 Variables and Measurement

#### Dependent Variables:

- a. Tax collection efficiency: Ratio of actual to potential revenue
- b. Compliance rate: Percentage of assessed taxpayers making timely payments. Tax base expansion: Net increase in registered taxpayers
- c. Revenue forecast accuracy: Mean absolute percentage error
- d. Administrative cost ratio: Collection costs relative to revenue

#### Independent Variables:

- a. Data integration status: Binary indicator of system implementation
- b. Household socioeconomic indicators: Derived from family card data
- c. Demographic characteristics: Age, education, occupation, household  
Geographic factors: Urban/rural classification, accessibility indices
- d. Prior compliance history: Payment regularity, outstanding liabilities

#### Control Variables:

- a. Economic conditions: Local GDP growth, inflation, unemployment
- b. Policy changes: Tax rate adjustments, amnesty programs, etc,  
Administrative capacity: Staff numbers, IT infrastructure, training investments

## 4. RESULTS

### 4.1 Descriptive Statistics and Data Integration Outcomes

Table 1 presents summary statistics for key variables across the three study districts before and after data integration implementation. The integrated system successfully linked 94.3% of individual population records with family card household data and 87.6% of households with tax administration accounts.

**Table 1. Descriptive Statistics Before and After Data Integration (2021-2024)**

Variable	District A (Urban)	District B (Peri-urban)	District C (Rural)
<b>Population Coverage</b>			
Total population	1,203,450	648,230	382,140
Registered in family cards (%)	96.8 → 98.4	94.2 → 97.1	91.5 → 95.8
Linked to tax system (%)	42.3 → 68.7	38.6 → 61.2	31.4 → 53.8
<b>Tax Administration Performance</b>			
Registered taxpayers	508,460 → 826,770	250,120 → 396,490	120,180 → 205,660
Active taxpayers (%)	67.2 → 82.8	63.5 → 78.4	58.3 → 72.1
Compliance rate (%)	61.4 → 79.6	58.7 → 75.3	54.2 → 69.8
Average collection ratio (%)	68.3 → 84.7	64.9 → 81.2	59.6 → 76.4
<b>Revenue Outcomes</b>			
Total revenue (billion IDR)	487.3 → 722.8	213.6 → 304.9	98.4 → 138.7
Revenue growth rate (%)	12.4 → 48.3	10.7 → 42.8	8.9 → 41.0
Administrative cost ratio (%)	18.6 → 10.8	21.3 → 13.5	24.7 → 16.2
Forecast accuracy (MAPE)	32.1 → 11.3	35.8 → 13.7	38.4 → 15.9

*Note: Values shown as Before → After integration implementation. MAPE = Mean Absolute Percentage Error.*

The integration resulted in a substantial expansion of the registered taxpayer base, increasing by 62.6% in District A, 58.5% in District B, and 71.1% in District C. This expansion reflects improved identification of previously unregistered individuals and businesses through cross-referencing population and household data. Active taxpayer rates improved by 15.6-13.8 percentage points across districts, indicating enhanced compliance.

Revenue collection demonstrated significant improvement, with total local tax revenue increasing by 48.3%, 42.8%, and 41.0% in the three districts, respectively, over the study period. While economic growth contributed to baseline revenue increases, difference-in-differences analysis (controlling for

non-integrated comparison districts) attributes 31.7-37.4% of the revenue growth directly to data integration effects.

Administrative efficiency gains were substantial, with collection cost ratios declining by 7.8-8.5 percentage points. Automated data matching and risk-based enforcement reduced manual processing requirements and improved resource allocation. Revenue forecasting accuracy improved dramatically, with mean absolute percentage errors decreasing from 32.1-38.4% to 11.3-15.9%, enabling more reliable budget planning.

#### 4.2 Predictive Model Performance

Table 2 summarizes the performance of five machine learning models developed to predict tax compliance behavior using integrated individual-household data. The models were trained on 70% of the dataset (1.56 million records) and evaluated on the remaining 30% test set.

**Table 2. Predictive Model Performance Comparison**

Model	Accuracy	Precision	Recall	F1-Score	AUC	Training Time
<b>Logistic Regression</b>	0.742	0.716	0.689	0.702	0.811	4.3 min
<b>Decision Tree</b>	0.768	0.753	0.721	0.736	0.829	8.7 min
<b>Random Forest</b>	0.843	0.827	0.806	0.816	0.914	47.2 min
<b>Gradient Boosting (GBM)</b>	0.836	0.819	0.798	0.808	0.907	52.8 min
<b>XGBoost</b>	0.851	0.834	0.815	0.824	0.921	38.6 min

*Note: Models trained on the integrated demographic-household-tax dataset. Metrics calculated on 30% holdout test set.*

XGBoost achieved the highest performance across most metrics, with 85.1% overall accuracy and 0.921 AUC. The model correctly identified 83.4% of compliant taxpayers (precision) and captured 81.5% of actual compliant cases (recall). Random Forest demonstrated competitive performance at 84.3% accuracy with substantially lower computational requirements.

Feature importance analysis revealed that the top predictors of tax compliance were: (1) household income proxy (derived from asset ownership and employment data in family cards) - 18.4% importance; (2) prior payment history - 16.7%; (3) education level - 12.3%; (4) property ownership status - 11.8%; (5) household head occupation - 9.6%; (6) number of economically active household members - 8.2%; (7) geographic accessibility index - 7.4%; (8) business registration status - 6.9%; and (9) age - 5.3%.

Household-level variables derived from integrated family card data contributed 42.7% of total predictive power, demonstrating the value of incorporating demographic and socioeconomic information beyond standard tax administration records. The model's ability to leverage family structure and

household composition data proved particularly valuable for identifying informal sector participants and assessing tax capacity.

**Table 3. Taxpayer Segmentation and Risk Profiling Results**

Segment	Share (%)	Avg. Compliance	Avg. Tax Liability	Enforcement Priority
High-value compliant	12.4	94.6%	IDR 14.3 million	Low - maintain
Medium-value compliant	28.7	86.2%	IDR 4.8 million	Low - maintain
High-value non-compliant	8.3	43.7%	IDR 16.9 million	Very High
Medium-value inconsistent	19.6	68.4%	IDR 5.7 million	Medium-High
Low-value compliant	23.8	79.3%	IDR 1.2 million	Low
Low-value non-compliant	7.2	38.1%	IDR 1.4 million	Medium

*Note: Segmentation based on k-means clustering using compliance history, tax liability, and socioeconomic indicators.*

Clustering analysis segmented taxpayers into six distinct groups with varying compliance patterns and revenue potential. The high-value non-compliant segment, representing 8.3% of taxpayers but 23.7% of potential revenue, became the primary target for intensive enforcement. Predictive models enabled proactive identification of taxpayers likely to transition from compliant to non-compliant status, facilitating preventive interventions.

#### 4.3 Impact on Regional Development Planning

The integration of demographic-household-fiscal data fundamentally enhanced evidence-based development planning capacity. Table 4 presents key planning indicators before and after data integration across the study districts.

**Table 4. Regional Development Planning Enhancement Indicators**

Planning Dimension	District A	District B	District C
<b>Revenue Forecasting</b>			
Budget variance (%)	28.4 → 9.7	31.6 → 11.2	35.2 → 13.8
Multi-year plan reliability (%)	64.3 → 87.6	61.8 → 84.3	58.4 → 79.7
<b>Resource Allocation</b>			
Spatial equity index (0-1)	0.63 → 0.82	0.59 → 0.78	0.56 → 0.74
Needs-based targeting accuracy (%)	67.2 → 88.4	64.8 → 85.7	62.1 → 82.3
<b>Infrastructure Investment</b>			
Project completion rate (%)	71.3 → 89.7	68.9 → 86.2	66.4 → 83.8
Cost overrun reduction (%)	-	18.4	21.7

<b>Social Protection</b>			
<b>Targeting leakage rate (%)</b>	34.7 → 12.8	37.2 → 14.6	39.8 → 16.9
<b>Coverage of eligible population (%)</b>	68.4 → 91.3	65.7 → 88.6	62.9 → 85.4

*Note: Values shown as Before → After integration. Equity index measures resource distribution relative to need.*

Revenue forecasting reliability improved dramatically, with actual collections deviating from budget projections by only 9.7-13.8% post-integration compared to 28.4-35.2% previously. This enhanced predictability enabled more ambitious multi-year development plans and infrastructure investments. The proportion of development programs successfully completed according to plan increased by 18.4-17.4 percentage points.

Spatially disaggregated revenue and demographic data facilitated more equitable resource allocation across sub-districts and villages. The spatial equity index, measuring the correspondence between resource allocation and need indicators, improved by 0.18-0.19 points on a 0-1 scale. Needs-based targeting accuracy for development programs increased by 20-21 percentage points, ensuring resources reached the most underserved communities.

Social protection program efficiency benefited substantially from integrated data, enabling precise identification of eligible households. Targeting leakage rates (benefits reaching ineligible households) declined by 21.9-22.9 percentage points, while coverage of truly eligible populations increased by 22.9-22.5 percentage points. These improvements generated fiscal savings while enhancing program effectiveness.

Qualitative interviews revealed that integrated data systems transformed planning processes by providing granular insights into local socioeconomic conditions, enabling evidence-based priority setting, and facilitating performance monitoring. Planning officials reported greater confidence in budget projections and improved ability to justify resource allocation decisions to legislative bodies and citizens.

#### **4.4 Implementation Challenges and Success Factors**

Qualitative analysis identified several critical implementation challenges. Legal and regulatory frameworks for inter-agency data sharing required clarification, with privacy concerns necessitating robust governance mechanisms. Technical interoperability posed significant hurdles, as legacy systems employed inconsistent data standards and formats. Organizational resistance emerged from concerns about workload changes and potential accountability shifts.

Capacity building requirements proved substantial, encompassing technical skills for data analytics, change management competencies, and new operational procedures. Initial implementation costs including system development, data migration, and staff training averaged IDR 12.4-16.7 billion per district, equivalent to 3.8-4.2% of annual local budgets. However, efficiency gains generated payback periods of 2.1-2.7 years.

Key success factors included: (1) strong political commitment from district leaders; (2) dedicated implementation teams with cross-departmental

representation; (3) phased rollout allowing iterative learning and adjustment; (4) investment in user-friendly interfaces reducing technical barriers; (5) transparent communication about benefits and privacy protections; and (6) performance incentives linking personnel evaluations to system utilization and outcomes.

Districts that invested heavily in change management and stakeholder engagement achieved substantially faster adoption and better results. Technical solutions alone proved insufficient without addressing organizational culture and process redesign.

## **5. DISCUSSION**

### **5.1 Theoretical and Practical Implications**

The findings demonstrate that integrated individual-household-fiscal data systems can fundamentally transform subnational tax administration and development planning capacity. The 31.7-48.3% revenue increases observed across study districts substantially exceed typical outcomes from conventional tax administration reforms, which generally yield 8-15% improvements (Prichard et al., 2019). This superior performance reflects the multiplicative benefits of data integration, enabling simultaneous enhancement of taxpayer identification, compliance monitoring, risk assessment, and enforcement targeting.

The predictive modeling results validate the application of advanced analytics to public finance challenges in developing country contexts. The 84-85% accuracy achieved by ensemble machine learning methods compares favorably with international benchmarks, despite data quality challenges inherent in emerging economy settings (Mascagni et al., 2023). Feature importance analysis confirms that household-level socioeconomic information substantially enhances predictive power beyond traditional tax administration data, justifying integration investments.

From a public accounting perspective, the integration enables more accurate revenue recognition, comprehensive asset-liability assessment, and performance-based reporting aligned with accrual principles (Carnegie et al., 2023). The dramatic improvement in revenue forecasting accuracy from 32-38% error rates to 11-16% facilitates the adoption of medium-term expenditure frameworks and evidence-based budgeting practices recommended by international public financial management standards (Van der Hoek, 2023).

The regional development planning implications prove equally significant. Spatially disaggregated demographic-fiscal data enable precision in resource allocation, addressing long-standing equity concerns in decentralized governance systems (Smoke, 2022). The 0.18-0.19 point improvements in spatial equity indices represent substantial progress toward inclusive development, particularly benefiting previously underserved rural and peri-urban populations.

### **5.2 Comparative Analysis with International Evidence**

The observed compliance rate improvements of 15.6-18.3 percentage points align with evidence from Rwanda (14.7%), Kenya (17.2%), and Ghana (16.8%)

following tax administration digitalization reforms (Mascagni et al., 2023; Okunogbe & Santoro, 2023). However, the taxpayer base expansion of 58.5-71.1% exceeds typical outcomes, suggesting that family card integration provides a particularly effective means for identifying informal sector participants common in developing Asian economies.

Administrative efficiency gains with cost ratios declining from 18.6-24.7% to 10.8-16.2% – compare favorably to international benchmarks. OECD countries typically achieve 0.5-2.0% collection cost ratios, but these reflect long-established systems and higher tax capacity (Beer et al., 2022). Among developing countries, the study districts' post-integration ratios approach best-practice levels observed in Malaysia (9.4%), Thailand (11.8%), and Chile (13.2%) (Martinez-Vazquez et al., 2021).

The predictive model performance metrics align with recent literature on machine learning applications in tax administration. Albuquerque and Garzón (2024) reported 81-83% accuracy for property tax compliance prediction in Latin America, while Chen and Tsai (2023) achieved 86% accuracy for value-added tax fraud detection in Taiwan using similar ensemble methods.

### 5.3 Policy Recommendations

Based on the research findings, we propose several policy recommendations for governments pursuing data integration strategies:

**Legal and Institutional Framework:** Establish clear legal authorization for inter-agency data sharing while incorporating privacy safeguards. Indonesia's recent Personal Data Protection Law (2022) provides a foundation but requires implementing regulations specifying permissible uses, consent requirements, and security standards. Independent oversight mechanisms should monitor compliance and address concerns.

**Technical Architecture:** Adopt interoperability standards (e.g., HL7, FHIR), facilitating system integration without requiring wholesale replacement of legacy infrastructure. Develop API-based architectures enabling modular upgrades and cross-platform data exchange. Prioritize cloud-based solutions offering scalability and cost-effectiveness for resource-constrained local governments.

**Implementation Approach:** Pursue phased rollout beginning with pilot districts demonstrating proof-of-concept before scaling nationally. This allows iterative learning, problem-solving, and local adaptation. Establish implementation support units providing technical assistance and facilitating knowledge sharing across jurisdictions.

**Capacity Building:** Design comprehensive training programs addressing technical skills (data analytics, system operation), process knowledge (new workflows and procedures), and change management competencies. Create communities of practice enabling peer learning among tax administrators and planning officials.

**Performance Management:** Implement monitoring frameworks tracking integration progress, utilization rates, and outcome indicators. Link personnel performance evaluations and budget allocations to system adoption and results

achievement. Publish transparency reports demonstrating benefits and maintaining public trust.

**Equity Safeguards:** Ensure data integration benefits extend to all populations, including rural residents, informal sector workers, and marginalized groups. Design user interfaces accommodating limited digital literacy. Provide alternative access channels for those lacking connectivity or devices.

#### 5.4 Limitations and Future Research

Several limitations warrant acknowledgment. The study's geographic scope, which includes three Indonesian districts, may limit generalizability to other contexts with different institutional arrangements, technical capacities, or socioeconomic conditions. The 2021-2024 timeframe captures initial implementation effects but not long-term sustainability and evolution. The selection of districts implementing integration introduces potential selection bias, though the difference-in-differences approach with comparison areas partially mitigates this concern.

Data quality challenges, including incomplete registration, delayed updates, and measurement errors, likely attenuate observed effects, suggesting potential underestimation of integration benefits under ideal conditions. The predictive models' performance may degrade over time as taxpayer behavior adapts to enhanced enforcement, requiring continuous model retraining and refinement.

Future research should examine long-term integration impacts beyond the initial implementation period, including the sustainability of compliance improvements and the evolution of evasion strategies. Cross-country comparative studies would illuminate how institutional contexts shape integration outcomes and identify transferable best practices. Investigation of equity dimensions, whether benefits accrue proportionally across income groups and geographic areas, merits focused attention.

Advanced analytical techniques, including causal inference methods (synthetic control, regression discontinuity), dynamic modeling of behavioral responses, and cost-benefit analyses, would strengthen evidence base. Research on optimal data governance frameworks balancing efficiency gains against privacy protection remains critical. Finally, exploration of integration applications beyond taxation, including social protection, public service delivery, and disaster response, would demonstrate broader digital government transformation potential.

## 6. CONCLUSION

This research demonstrates that integrating individual demographic data with family card household information fundamentally enhances subnational tax administration effectiveness and regional development planning capacity. Across three Indonesian districts representing diverse urban-rural characteristics, data integration yielded revenue increases of 31.7-48.3%, compliance rate improvements of 15.6-18.3 percentage points, and taxpayer base

expansion of 58.5-71.1%. Administrative efficiency gains reduced collection cost ratios by 35-42% while improving revenue forecasting accuracy from 68% to 89%. Predictive analytics leveraging integrated datasets achieved 84-85% accuracy in compliance behavior prediction, enabling sophisticated risk profiling and targeted enforcement. Machine learning models identified previously unregistered taxpayers, predicted payment likelihood, and segmented populations for customized interventions. The household-level socioeconomic information from family cards contributed 42.7% of predictive power, validating integration investments.

Regional development planning benefited from enhanced revenue predictability, spatially disaggregated demographic-fiscal insights, and improved social protection targeting. Budget variance decreased from 28-35% to 10-14%, spatial equity indices improved by 0.18-0.19 points, and program targeting leakage declined by 22-23 percentage points. These improvements enable more ambitious, evidence-based development strategies aligned with local needs.

Implementation challenges legal frameworks, technical interoperability, organizational resistance, capacity constraints required systematic attention through phased approaches, stakeholder engagement, and sustained capacity building. Success factors included political commitment, cross-departmental coordination, user-friendly system design, transparent communication, and performance incentives linking outcomes to organizational objectives.

The research contributes methodologically by demonstrating the effective application of advanced analytics in data-constrained developing country contexts. Ensemble machine learning methods proved robust despite data quality imperfections, while mixed-methods approaches captured both quantitative impacts and qualitative implementation dynamics. The findings validate theoretical frameworks linking digital governance infrastructure to fiscal capacity and development outcomes.

Policy implications extend beyond taxation to broader digital government transformation. Integrated population-service delivery systems can enhance efficiency, equity, and accountability across multiple public functions. However, realizing these benefits requires addressing governance challenges around data privacy, interoperability standards, and institutional coordination. Legal frameworks must balance administrative efficiency against individual rights, while technical architectures must enable integration without creating dependencies or vulnerabilities.

The Indonesian experience offers valuable lessons for similarly positioned developing countries pursuing fiscal decentralization and digital modernization. Family registration systems common across Asia, Africa, and Latin America represent underutilized assets for enhancing government capacity. When combined with public accounting principles and predictive analytics capabilities, these population databases can transform tax administration from compliance enforcement to strategic revenue optimization, supporting equitable development.

Future trajectories should emphasize continuous system evolution, incorporating emerging technologies (artificial intelligence, blockchain, mobile platforms), expanding analytical sophistication (real-time monitoring, prescriptive recommendations), and deepening citizen engagement (transparency portals, participatory planning). The integration approach demonstrated here provides a foundation for comprehensive digital governance ecosystems, enabling responsive, effective, and accountable public institutions.

Ultimately, this research affirms that data-driven approaches to public financial management can substantially enhance fiscal capacity while promoting more equitable development outcomes. As governments worldwide grapple with resource mobilization challenges amid rising service expectations, integrated information systems leveraging demographic, socioeconomic, and administrative data offer promising pathways toward sustainable, inclusive prosperity.

### ACKNOWLEDGMENT

This section gives you the opportunity to thank your colleagues who provided suggestions for your paper. You can also express your appreciation for the financial assistance you received, in completing this research.

### REFERENCES

- Adhikari, P., Kuruppu, C., & Matilal, S. (2023). Public sector accounting reforms in emerging economies: A structuration perspective on Nepalese central government accounting. *Financial Accountability & Management*, 39(1), 45–68. <https://doi.org/10.1111/faam.12315>
- Albuquerque, P. H., & Garzón, J. P. (2024). Machine learning models for property tax compliance prediction in Latin American cities. *Government Information Quarterly*, 41(1), 101876. <https://doi.org/10.1016/j.giq.2023.101876>
- Alm, J., & Soled, J. A. (2022). W(h)ither the tax gap? *National Tax Journal*, 75(3), 489–519. <https://doi.org/10.1086/720614>
- Alstadsæter, A., Johannesen, N., & Zucman, G. (2022). Tax evasion and inequality. *American Economic Review*, 112(6), 2073–2103. <https://doi.org/10.1257/aer.20191854>
- Anessi-Pessina, E., Sicilia, M., & Steccolini, I. (2022). Budgeting and performance management in the public sector: Quo vadis? *Journal of Public Budgeting, Accounting & Financial Management*, 34(6), 1–18. <https://doi.org/10.1108/JPBAFM-02-2022-0021>
- Beer, S., Coelho, M., & Leduc, S. (2022). Hidden debt: Solutions to avert the next financial crisis in South Asia. *World Bank Economic Review*, 36(2), 382–408. <https://doi.org/10.1093/wber/lhab023>
- Bellanca, S., & Vandernoot, J. (2021). International public sector accounting standards (IPSAS) implementation: A systematic literature review and future research agenda. *Public Money & Management*, 41(7), 539–551. <https://doi.org/10.1080/09540962.2020.1835495>

- Besley, T., & Persson, T. (2023). The political economics of development: A political economy approach. *Annual Review of Economics*, 15, 1–24. <https://doi.org/10.1146/annurev-economics-082222-065846>
- Bird, R. M., & Slack, E. (2023). Local taxes and local expenditures in developing countries: Strengthening the wicksellian connection. *Public Finance Review*, 51(2), 187–215. <https://doi.org/10.1177/10911421221138421>
- Boex, J., & Edwards, B. (2022). The political economy of local government finance reforms in developing countries. *Public Administration and Development*, 42(3), 156–169. <https://doi.org/10.1002/pad.1982>
- Carnegie, G. D., Ballas, A., Bloomfield, B. P., Caperchione, E., Christiaens, J., Dimnik, T., Ellwood, S., Goddard, A., Guerin, B., Manes Rossi, F., Reichard, C., Ridder, H.-G., Ruggiero, P., Sicilia, M., & Steccolini, I. (2023). A framework for twenty-first-century public sector accounting research. *Financial Accountability & Management*, 39(1), 3–26. <https://doi.org/10.1111/faam.12339>
- Chen, H. C., & Tsai, W. H. (2023). Applying machine learning techniques for value-added tax fraud detection. *Expert Systems with Applications*, 213, 119012. <https://doi.org/10.1016/j.eswa.2022.119012>
- Chen, S., Chen, D. L., & Liang, P. (2022). Digital tax administration, third-party information, and tax compliance: Evidence from a natural experiment. *Journal of Public Economics*, 214, 104725. <https://doi.org/10.1016/j.jpubeco.2022.104725>
- Dzingirai, C., & Gandiwa, E. (2023). Digital transformation and tax revenue performance in developing countries: Evidence from sub-Saharan Africa. *Development Policy Review*, 41(3), e12665. <https://doi.org/10.1111/dpr.12665>
- Fadlurrohman, I., Husein, A. M., Winarno, W. A., & Rasywir, E. (2022). Evaluation of digital transformation in the Indonesian local government. *Digital Policy, Regulation and Governance*, 24(5), 456–473. <https://doi.org/10.1108/DPRG-01-2022-0008>
- Fajri, R. N., Setiawan, E. B., & Kurniawan, R. (2023). Spatial analysis of fiscal capacity and development outcomes in Indonesian districts. *Regional Science Policy & Practice*, 15(4), 847–866. <https://doi.org/10.1111/rsp3.12582>
- Gelb, A., & Metz, A. D. (2023). Identification revolution: Can digital ID be harnessed for development? *World Development*, 161, 106073. <https://doi.org/10.1016/j.worlddev.2022.106073>
- Gunawan, H., & Salomo, R. V. (2023). Local tax administration reform and revenue mobilization in Indonesia: Evidence from panel data analysis. *Asia-Pacific Journal of Regional Science*, 7(1), 129–156. <https://doi.org/10.1007/s41685-022-00268-4>
- Gupta, S., Keen, M., Shah, A., & Verdier, G. (2021). *Digital revolutions in public finance*. International Monetary Fund. <https://doi.org/10.5089/9781513511818.071>
- Hassan, I., & Rahman, M. M. (2023). Data integration and public sector performance: A systematic review and conceptual framework. *International*

- Journal of Public Administration*, 46(8), 573–589.  
<https://doi.org/10.1080/01900692.2022.2038574>
- Issa, H., Jabbouri, R., & Palmer, M. (2022). An artificial intelligence (AI)-readiness and adoption framework for AgriTech firms. *Technological Forecasting and Social Change*, 182, 121874. <https://doi.org/10.1016/j.techfore.2022.121874>
- Jones, R., & Caruana, J. (2022). A perspective on the development and adoption of accrual accounting in the public sector. *Public Money & Management*, 42(1), 17–26. <https://doi.org/10.1080/09540962.2021.1990313>
- Khagram, S., Fung, A., & De Renzio, P. (2023). *Open budgets: The political economy of transparency, participation, and accountability*. Brookings Institution Press.
- Khanh, C. N., & Khanh, H. T. (2023). Digital transformation in public financial management: Evidence from Vietnamese local governments. *Journal of Asian Public Policy*, 16(2), 234–252. <https://doi.org/10.1080/17516234.2021.1996682>
- Kleven, H., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2023). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 91(3), 1119–1153. <https://doi.org/10.3982/ECTA18114>
- Kumar, S., & Srivastava, A. (2023). Taxpayer segmentation using machine learning: Behavioral insights for tax administration. *International Journal of Public Sector Management*, 36(3), 287–305. <https://doi.org/10.1108/IJPSM-08-2022-0183>
- Manolo, R. A., Santos, E. M., & Reyes, C. M. (2024). Household registration systems and social protection targeting in Southeast Asia. *Asian Development Review*, 41(1), 94–118. <https://doi.org/10.1142/S0116110524500043>
- Martínez-Vázquez, J., & Vulovic, V. (2023). *How well do subnational borrowing regulations work?* Asian Development Bank Economics Working Paper Series No. 678.
- Martínez-Vázquez, J., Lago-Peñas, S., & Sacchi, A. (2021). The impact of fiscal decentralization: A survey. *Journal of Economic Surveys*, 35(4), 1106–1157. <https://doi.org/10.1111/joes.12475>
- Mascagni, G., Nell, C., & Monkam, N. (2023). One size does not fit all: A field experiment on the drivers of tax compliance and delivery of public services in Rwanda. *Journal of Development Economics*, 161, 103028. <https://doi.org/10.1016/j.jdeveco.2022.103028>
- Nath, S. (2024). Fiscal decentralization and subnational revenue mobilization: New evidence from panel data. *World Development*, 173, 106412. <https://doi.org/10.1016/j.worlddev.2023.106412>
- Nicolò, G., Sannino, G., & Rossi, P. (2024). Network analysis for tax fraud detection: Methodological advances and empirical evidence. *Government Information Quarterly*, 41(2), 101898. <https://doi.org/10.1016/j.giq.2024.101898>
- Nkurunziza, M., Brockmeyer, A., Russ, J., & Slemrod, J. (2023). Field experiments on tax compliance in developing countries: A review. *Journal of Economic Literature*, 61(2), 512–561. <https://doi.org/10.1257/jel.20211652>

- Okunogbe, O., & Santoro, F. (2023). The promise and limitations of information technology for tax mobilization. *World Bank Research Observer*, 38(2), 295–324. <https://doi.org/10.1093/wbro/lkad001>
- Philbrick, W., Sandholtz, W., & Wasson, E. (2022). Civil registration and vital statistics systems: Institutional foundations of governance. *World Development*, 157, 105925. <https://doi.org/10.1016/j.worlddev.2022.105925>
- Prichard, W., Custers, A., Dom, R., & Davenport, S. (2019). Innovations in tax compliance: Building trust, navigating politics. *World Bank Policy Research Working Paper*, 8923.
- Purwanto, A., & Kusumawardani, A. (2024). Family registration data utilization for local government service delivery in Indonesia. *Asian Journal of Political Science*, 32(1), 78-97. <https://doi.org/10.1080/02185377.2023.2267894>
- Sjahrir, B. S., Kis-Katos, K., & Schulze, G. G. (2023). Political budget cycles in Indonesia at the district level. *Economics of Governance*, 24(1), 37–65. <https://doi.org/10.1007/s10101-022-00285-7>
- Smoke, P. (2022). Improving subnational government development finance in emerging and developing economies: Toward a strategic approach. *Public Administration and Development*, 42(4), 201–213. <https://doi.org/10.1002/pad.2001>
- Thongmee, S., & Srivannaboon, S. (2022). Digital identity infrastructure and tax compliance: Evidence from Thailand's national ID system. *Information Technology for Development*, 28(4), 782–804. <https://doi.org/10.1080/02681102.2021.2002023>
- Van der Hoek, M. P. (2023). From cash to accrual accounting: Perspectives on public sector accounting. *Public Money & Management*, 43(2), 145–153. <https://doi.org/10.1080/09540962.2022.2098390>