

# AI-DRIVEN DATA QUALITY MANAGEMENT AND ANOMALY DETECTION IN LARGE-SCALE E-COMMERCE DATABASES: A COMPREHENSIVE ANALYSIS WITH FOCUS ON UZBEKISTAN'S DIGITAL MARKETPLACE ECOSYSTEM

*E.A. Majidov*

[mazhidoov@gmail.com](mailto:mazhidoov@gmail.com)

*Assistant Professor of ISFT institute Samarkand Branch*

**Abstract:** Contemporary e-commerce environments face unprecedented challenges in maintaining data integrity across vast transnational databases. This investigation examines artificial intelligence applications for automated data quality control and anomaly identification within high-volume digital commerce platforms. Our empirical study, conducted across multiple e-commerce ecosystems including Uzbekistan's rapidly expanding digital market, processed 62.7 million transactions over eight months. Machine learning implementations achieved 91.8% precision in detecting data inconsistencies while reducing manual oversight requirements by 72%. Ensemble-based detection systems demonstrated 38% superior performance compared to conventional rule-based approaches, particularly in identifying fraudulent patterns and database anomalies. The Uzbekistan market analysis revealed unique data challenges including multi-currency processing, diverse payment integration, and multilingual content management, providing valuable insights for emerging digital economies.

**Keywords:** Data Quality Assurance, Machine Learning Detection, Digital Commerce, Database Optimization, Uzbekistan E-commerce

## 1. Introduction

Digital commerce platforms worldwide experience exponential transaction volume growth, creating complex data management challenges that traditional approaches struggle to address effectively.

Uzbekistan's e-commerce sector exemplifies these challenges perfectly, with over 50 marketplaces

generating \$300 million annually, projected to reach \$1 billion by 2027. Uzum, the leading platform, serves nearly a third of the nation's population monthly, processing 14 million orders in the first nine months of 2024 with daily GMV capacity of \$4.5 million.

Data quality issues manifest through duplicate customer records, inconsistent product categorization, pricing anomalies, and fraudulent transaction patterns. With Uzbekistan's eCommerce market projected to grow by 11.12% annually through 2029, reaching \$2.6 billion, addressing these challenges becomes critical for sustainable digital commerce growth. Traditional rule-based systems prove inadequate when

confronting the scale and complexity of modern e-commerce operations, necessitating sophisticated AI- driven solutions.

## 2. Research Methodology

This investigation employed a comprehensive mixed-methods approach, combining quantitative performance analysis with qualitative assessment of implementation challenges across diverse e- commerce environments over eight months (January-August 2024).

### 2.1 Data Sources and Collection Framework

The research dataset comprised 62.7 million transactions, 4.2 million customer profiles, and 1.8 million product entries from three platforms:

- **Platform Alpha (International Electronics):** 22.4 million transactions, \$287 average order value, 45 countries
- **Platform Beta (Regional Fashion):** 18.6 million transactions, \$73 average order value, Central Asia focus
- **Uzum Market (Uzbekistan):** 21.7 million transactions, \$52 average order value, 600,000+ SKUs, 9,000+ merchants, multilingual support (Uzbek, Russian, English)

### 2.2 Algorithm Implementation

Four machine learning approaches were implemented: Random Forest Classifier, Gradient Boosting (XGBoost), Support Vector Machines, and Isolation Forest for unsupervised detection. An ensemble methodology combining weighted voting and stacking approaches was developed for optimal performance across diverse data types and cultural contexts.

## 3. Results and Analysis

### 3.1 Data Quality Issue Distribution

Analysis revealed distinct patterns across platforms, with Uzbekistan showing unique characteristics:

Issue Category	Platform Alpha	Platform Beta	Uzum Market	Total %
Missing Data Fields	31,200	42,150	38,920	25.4%
Format Inconsistencies	22,480	28,340	41,680	20.9%
Duplicate Records	18,420	24,680	31,240	16.8%
Product Classification	14,720	21,450	19,280	12.5%
Address Standardization	12,680	16,820	24,360	12.1%
Currency Conversion Errors	3,240	5,670	18,440	6.2%
Payment Validation	4,180	6,240	15,680	5.9%

exhibited significantly higher currency conversion errors and payment validation issues, reflecting complexity in handling UZS-USD transactions and integrating multiple local payment systems  
(Click, Payme, Uzcard, Humo).

### 3.2 Machine Learning Performance

Performance evaluation revealed ensemble approaches yielding optimal results:

Algorithm	Precision	Recall	F1-Score	AUC-ROC
Ensemble Model	0.918	0.921	0.920	0.952
Gradient Boosting	0.916	0.902	0.909	0.938
Random Forest	0.893	0.876	0.884	0.921
Support Vector Machine	0.871	0.859	0.865	0.904
Isolation Forest	0.858	0.834	0.846	0.887

The ensemble model's superior performance was particularly pronounced in the Uzbekistan data set, attributed to its capability to simultaneously handle diverse data types, multiple languages, and varied payment methods.

### 3.3 Uzbekistan Market-Specific Findings

**Geographic Complexity:** Address standardization challenges due to ongoing administrative reforms and coexistence of traditional MAHALLA names with modern postal codes resulted in 28% of addresses flagged as potentially inconsistent.

**Multi-Payment Integration:** Local payment system integration created distinctive anomaly patterns, requiring specialized training achieving 73% accuracy improvement in fraud detection after incorporating Uzbek-specific transaction patterns.

**Trilingual Processing:** Product descriptions in three languages required specialized NLP models, ultimately achieving 86% accuracy in identifying inconsistent product information across languages.

**Cultural Seasonality:** Traditional shopping behaviors during Ramadan and Nawruz created distinct patterns requiring culturally-aware training data for accurate anomaly detection.

## 4. Implementation Impact and Discussion

### 4.1 Economic Benefits

Implementation demonstrated substantial economic benefits:

Metric	Pre-Implementation	Post-Implementation	Improvement
Manual QA Hours/Month	3,240	742	77% Reduction
Data Processing Time	5.8 hours	1.2 hours	79% Faster
Processing Cost/Transaction	\$0.052	\$0.014	73% Lower
Customer Complaint Resolution	14.2 hours	3.8 hours	73% Faster

Uzbek-specific benefits included 82% reduction in multilingual verification costs, 69% decrease in currency conversion error handling, and 23% increase in customer retention through improved data accuracy.

#### 4.2 Technical Challenges and Solutions

**Infrastructure Development:** Limited local cloud infrastructure necessitated hybrid deployment strategies combining domestic data centers with international services, implementing edge computing nodes reducing response times from 340ms to 87ms.

**Cultural Adaptation:** Local business practices required specialized algorithm training incorporating traditional bazaar-style negotiations, seasonal patterns, and extended family purchasing behaviors.

**Regulatory Compliance:** Uzbekistan's evolving data protection regulations required privacy-preserving techniques while maintaining system performance and benefiting from 50% tax rate reductions for compliant platforms.

#### 4.3 Implications and Future Directions

**Theoretical Contributions:** The ensemble approach established a framework for cross-cultural AI implementation in emerging markets, demonstrating superior performance in culturally diverse environments.

**Practical Applications:** Organizations in emerging markets should prioritize ensemble machine learning approaches, invest in cultural context training data, and develop hybrid infrastructure solutions balancing performance with regulatory compliance.

**Limitations:** The eight-month observation period and focus on three platforms may not represent entire market diversity. Cultural adaptation findings require validation across different emerging markets.

### Conclusion

This comprehensive investigation demonstrates the trans-formative potential of AI applications in data quality management for large-scale e-commerce systems. Processing 62.7 million transactions across diverse platforms, with particular emphasis on Uzbekistan's digital marketplace, the research establishes that machine learning algorithms achieve 91.8% precision while reducing operational costs by 73%.

Key findings include successful ensemble model performance achieving superior metrics across all evaluations, effective multilingual content handling with 86% accuracy, and significant economic impact through cost reduction and revenue enhancement. The Uzbekistan implementation highlighted critical considerations for AI deployment in emerging markets, where cultural context and regulatory frameworks significantly influence system effectiveness.

With Uzbekistan's e-commerce market projected to reach \$2.6 billion by 2029, the demonstrated economic benefits justify substantial investment in AI-driven data quality systems. The framework provides a road-map for other emerging markets undergoing digital transformation, particularly in Central Asia facing similar cultural and economic challenges.

Future research should focus on real-time system adaptation, expanded cultural context integration, and standardized frameworks for cross-cultural AI deployment in emerging digital economies. The success of this implementation establishes a foundation for broader adoption of AI-driven data quality management across developing e-commerce markets.

### References:

1. Abdullayev, R., & Karimov, F. (2024). Digital transformation strategies in Central Asian e-commerce markets. *Journal of Emerging Market Technologies*, 15(3), 45-62.
2. Chen, L., Martinez, P., & Rodriguez, A. (2020). Unsupervised anomaly detection methodologies for online retail transaction analysis. *International Conference on Data Mining Applications*, 267-284.
3. Digital Commerce Analytics Institute. (2024). *Uzbekistan E-commerce Sector Performance Report 2024*. DCAI Publications, Tashkent.
4. Henderson, M., & Kumar, S. (2019). Supervised learning applications in financial database consistency verification. *Database Systems Quarterly*, 28(4), 178-195.
5. Johnson, R., Williams, C., & Brown, D. (2023). Economic impact assessment of artificial intelligence implementation in retail technology sectors. *Business Intelligence Analytics*, 42(7), 234-251.
6. KPMG Uzbekistan. (2023). *E-commerce Market Analysis: Growth Trajectories and Investment Opportunities*. KPMG Professional Services, Tashkent.
7. Lee, K., Patel, N., & Singh, R. (2023). Cross-cultural adaptation challenges in AI system deployment for emerging markets. *International Journal of Artificial Intelligence Applications*, 29(5), 312-329.
8. National Agency for Project Management of Uzbekistan. (2024). *Digital Economy Development Statistics and Projections*. NAPM Official Publications, Tashkent.