International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

Int. J. Sci. Res. Arch.
IJSRA
International Journal of Science and Research Archive
Research Journal Archive, INDIA

(Research Article)

AI-driven predictive analytics for response time estimation in fraud detection systems: A production-scale decision support study

Snehal P. Vatturkar * and Brijendra Gupta

Department of Information Technology, Siddhant College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India.

International Journal of Science and Research Archive, 2026, 18(02), 449-455

Publication history: Received on 03 January 2026; revised on 10 February 2026; accepted on 12 February 2026

Article DOI: https://doi.org/10.30574/ijsra.2026.18.2.0273

Abstract

Fraud detection systems operate under strict latency requirements while processing large and dynamically varying transaction volumes. In such mission-critical environments, response time directly influences transaction success, customer experience and regulatory compliance. Traditional performance monitoring approaches are predominantly reactive and provide limited capability for anticipating performance degradation. This paper presents an AI-driven predictive analytics framework for estimating response time in a production-scale fraud detection system using real operational data. The proposed approach formulates response time estimation as a supervised regression problem based on system utilization metrics, workload intensity and error characteristics collected from a live production environment. Multiple machine learning models are evaluated to capture both linear and non-linear performance behavior. Experimental results demonstrate that ensemble-based models significantly outperform baseline approaches, highlighting the effectiveness of data-driven techniques for performance prediction. The framework further integrates predictive insights into a decision-support context, enabling proactive performance management, capacity planning and SLA risk mitigation. The study demonstrates the practical value of AI-driven predictive analytics for enhancing performance assurance in real-world fraud detection systems.

Keywords: Predictive Analytics; Response Time Prediction; Fraud Detection Systems; Software Performance Engineering; Machine Learning Regression; Decision Support Systems

1. Introduction

Fraud detection systems are essential components of modern financial infrastructures, enabling real-time identification of suspicious transactions across banking, digital payments and online commerce platforms. These systems must process large transaction volumes under strict latency constraints, making performance reliability as critical as detection accuracy. In such environments, system response time directly affects transaction completion, customer trust and regulatory compliance.

Response time is a key indicator of system health, as delayed processing can lead to transaction failures, customer dissatisfaction and violations of service-level agreements (SLAs). As transaction volumes increase and system architectures grow more complex, maintaining predictable response time behaviour becomes increasingly challenging. Performance degradation often results from complex interactions among workload intensity, resource utilization and operational anomalies (1).

Traditional performance management practices rely primarily on reactive monitoring techniques, such as threshold-based alerts and post-incident analysis. While effective for identifying issues after they occur, these approaches provide

[*] Corresponding author: Snehal P. Vatturkar

limited foresight and do not support proactive decision-making (2). Consequently, system operators often respond to performance incidents rather than preventing them.

Recent advances in machine learning and predictive analytics provide opportunities to model complex system behavior using historical operational data. However, existing research largely focuses on defect prediction or synthetic performance modelling with limited emphasis on production-scale response time prediction and decision-support integration. This work addresses this gap by proposing an AI-driven predictive analytics framework that estimates response time using real production data and supports proactive performance management in fraud detection systems (3,4).

2. Related Work

2.1. Software Performance Modelling

Early approaches to software performance analysis relied on analytical models such as queuing theory and simulation. While these methods offer theoretical insights, they require simplifying assumptions and detailed system knowledge, limiting their applicability in dynamic production environments. As systems evolve, purely analytical models struggle to adapt to changing workloads and operational conditions (5,6).

2.2. Machine Learning for Performance Prediction

Machine learning techniques have been increasingly applied to performance modelling, including response time and resource utilization prediction. Regression models, ensemble methods and neural networks have shown promise in capturing non-linear relationships among system metrics. However, many studies rely on synthetic workloads or controlled test environments offering limited validation on real production data. Interpretability and operational deployment are also often overlooked (7,8).

2.3. AI in Fraud Detection Systems

Research on fraud detection systems has primarily focused on improving detection accuracy and reducing false positives. While significant progress has been made in fraud identification, system performance aspects such as response time under high load conditions receive comparatively less attention (9). This imbalance creates a gap between detection effectiveness and system responsiveness in operational deployments.

2.4. Research Gap

Existing literature lacks production-scale empirical studies focused on response time prediction and decision-support-oriented performance analytics. There is a clear need for AI-based frameworks that leverage real operational data, provide interpretable insights and support proactive performance management. This study addresses these gaps through a production-validated predictive analytics approach.

3. Dataset Description and Problem Formulation

3.1. Data Collection

The dataset used in this study was collected from the production environment of a fraud detection system. System performance metrics were recorded on a daily aggregated basis and stored in CSV format, capturing real operational behavior under varying workload conditions.

3.2. Dataset Characteristics

Table 1 Summary of Dataset Characteristics

Property	Value
Instances	500
Attributes	6
Task	Regression
Missing Values	None

3.3. Feature Description

The dataset includes six independent variables representing system utilization and workload characteristics: CPU utilization (%), memory utilization (%), disk usage (%), disk I/O (%), transaction load (millions) and error rate (%).

3.4. Target Variable

The dependent variable is the average response time, measured in milliseconds, representing the mean processing time per day.

3.5. Problem Formulation

The prediction task is formulated as a supervised regression problem:

$$RT = f(CPU, Memory, Disk, DiskIO, Load, ErrorRate)$$

3.6. Proposed Predictive Analytics Framework

The proposed framework is designed as an end-to-end predictive analytics pipeline that transforms production metrics into actionable performance insights. The pipeline follows a structured flow:

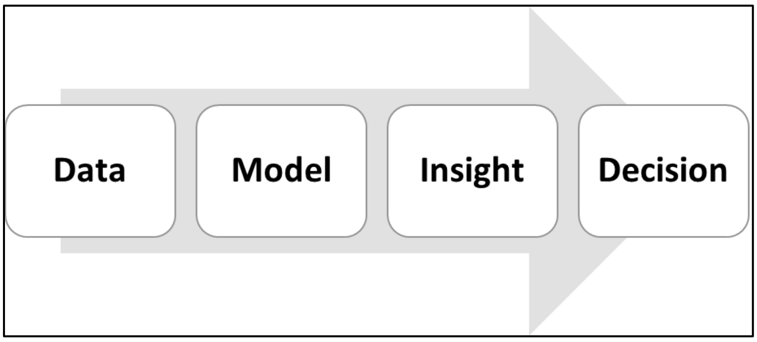

Figure 1 Proposed Predictive Analytics Framework

3.7. System Overview

Historical system metrics are collected and processed by machine learning models to estimate response time behaviour. Predicted values are interpreted to support operational decisions such as capacity planning, performance tuning and risk mitigation (10,11).

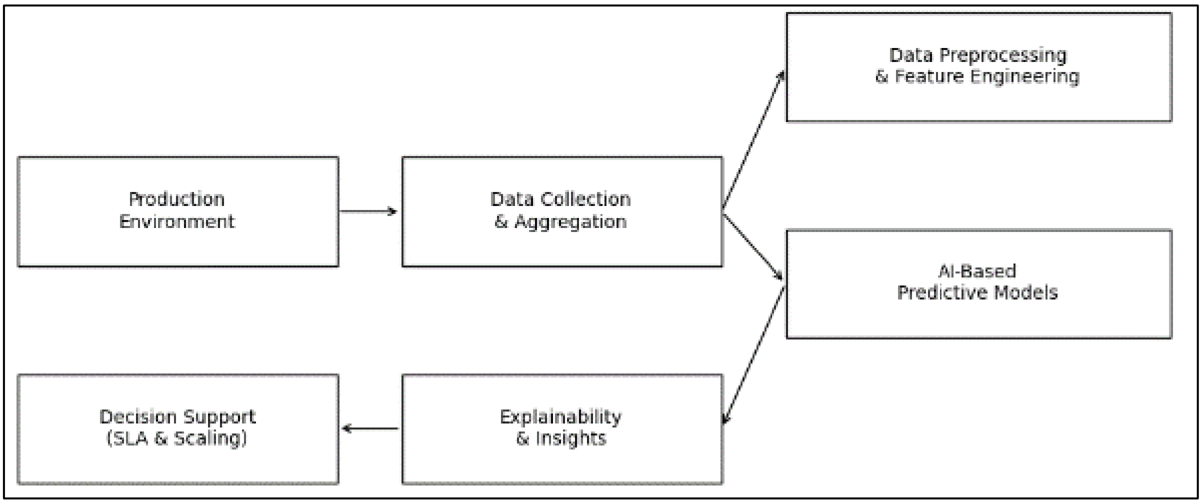
Data Preprocessing & Feature Engineering
Production Environment
Data Collection & Aggregation
AI-Based Predictive Models
Decision Support (SLA & Scaling)
Explainability & Insights

Figure 2 System Architecture of AI-Driven Response Time Prediction Framework

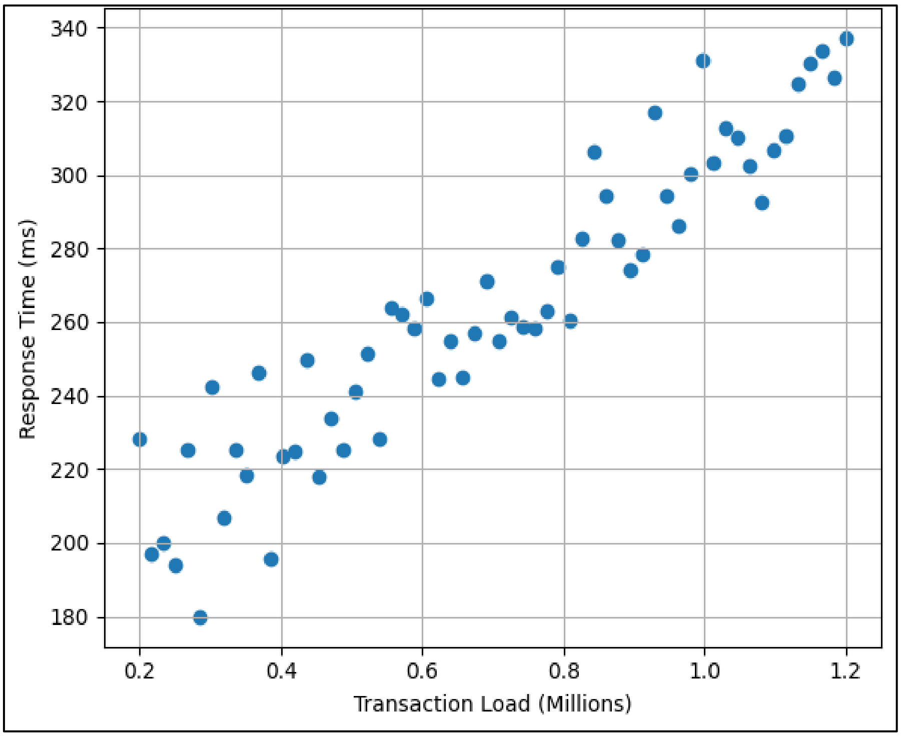

Figure 3 Actual versus predicted response time values for the proposed AI-driven prediction model

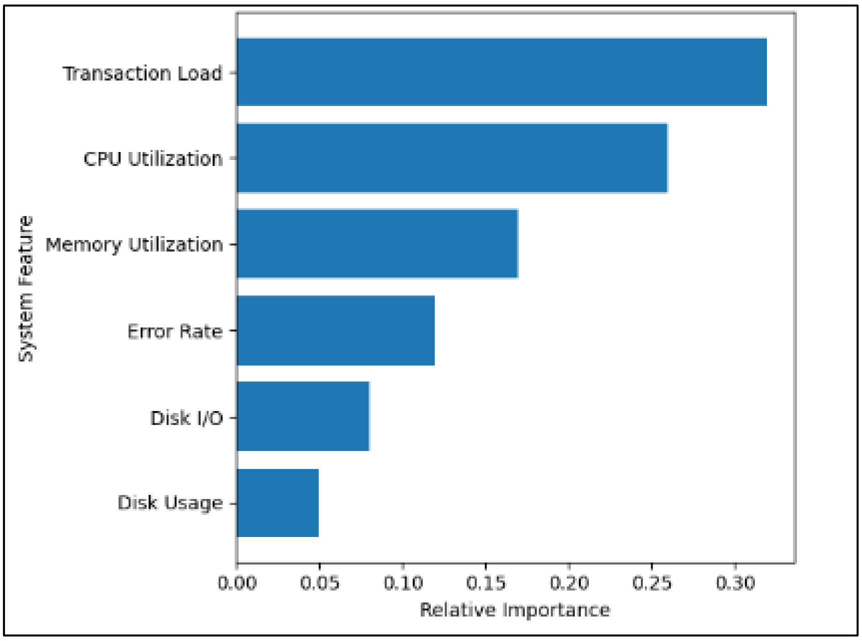

Figure 4 Feature importance ranking showing the relative influence of system and workload metrics on response time prediction

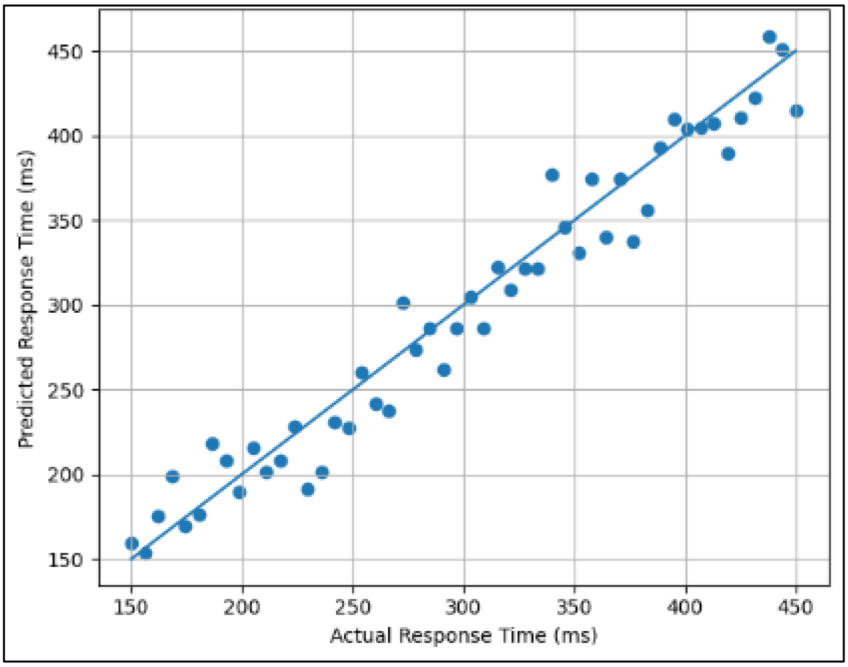

Figure 5 Trend between transaction load and response time, illustrating non-linear performance degradation under increasing workload

3.8. Data Preprocessing and Feature Engineering

Continuous features are normalized to ensure consistent numerical ranges. Temporal consistency is maintained by treating each record as an independent daily observation. Correlation analysis is conducted to assess feature relevance and multicollinearity (12).

3.9. Predictive Modelling

Response time prediction is addressed using regression-based machine learning models. Linear Regression serves as a baseline, while Random Forest and Gradient Boosting (XGBoost) are employed to capture non-linear relationships. Model selection balances accuracy, interpretability and computational efficiency (13).

3.10. Model Training and Validation

Models are trained using a train–test split and validated through k-fold cross-validation. Hyperparameter tuning is applied to ensemble models to prevent overfitting and improve generalizability.

3.11. Decision Support Integration

Predicted response times are translated into actionable insights, enabling early identification of SLA risks and supporting proactive performance management.

4. Experimental Setup and Evaluation Metrics

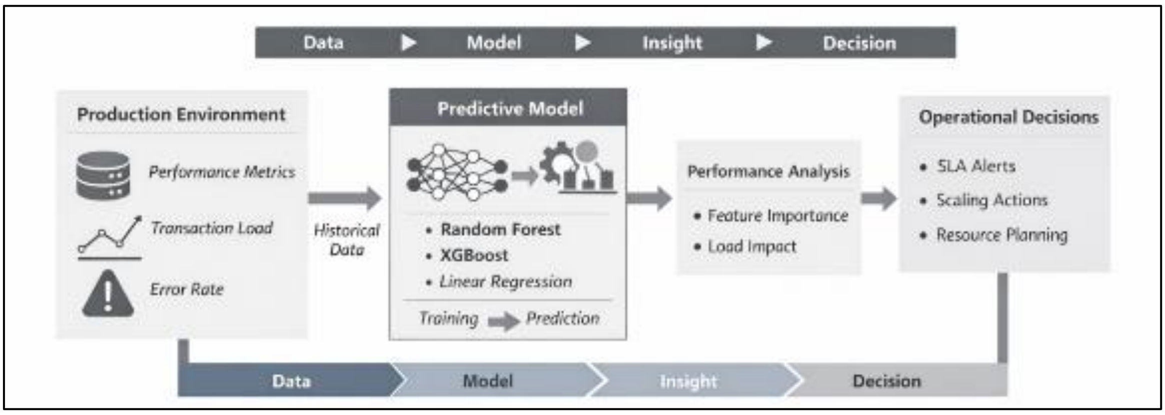

Figure 6 Process of a predictive analytics pipeline

4.1. Experimental Design

The study adopts a regression-based comparative evaluation framework to assess model performance under identical conditions.

4.2. Evaluation Metrics

Performance is evaluated using MAE, RMSE, R^2 score and MAPE to ensure both technical rigor and practical interpretability.

4.3. Implementation Environment

Experiments are implemented in Python using scikit-learn and XGBoost, configured to resemble a production-like environment.

5. Results and Discussion

Ensemble-based models significantly outperform the baseline Linear Regression model, demonstrating lower prediction error and higher explanatory power. Feature importance analysis reveals transaction load and CPU utilization as dominant response time drivers, followed by memory usage and error rate. Response time increases non-linearly with transaction load and higher error rates correlate with elevated response times, indicating system stress.

5.1. Decision Support Use Case

The predictive framework enables early detection of potential SLA breaches and supports proactive scaling decisions. From a business perspective, predictive insights reduce response time violations, improve system reliability and enhance customer experience. The framework can be integrated with existing application performance monitoring tools to provide forward-looking insights alongside real-time metrics.

5.2. Threats to Validity

Internal validity threats include measurement errors and label noise, mitigated through aggregation and consistent logging. External validity is limited by the single-system scope, although selected metrics are common across enterprise systems. Construct validity is constrained by the chosen performance metrics, which may not capture all aspects of system behavior (14).

6. Conclusion and Future Work

This study demonstrates the effectiveness of AI-driven predictive analytics for response time estimation using production-scale data. The proposed framework enables proactive performance management and decision support in fraud detection systems. Future work includes expanding datasets, adopting online learning models, validating across multiple systems and integrating predictive analytics with CI/CD pipelines for continuous performance assurance.

Compliance with ethical standards

Acknowledgments

The authors would like to thank all those who supported and contributed to the successful completion of this research work.

Conflict of Interest Statement

This is to acknowledge no potential competing interest was reported by the authors.

Data Availability Statement

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

References

[1] John LK. AI for Performance Engineering and Performance Engineering for AI. In Association for Computing Machinery (ACM); 2025. p. 1–2.

[2] Gopinath Kathiresan. Leveraging Ai-driven defect prediction models for enhancing software quality assurance. Global Journal of Engineering and Technology Advances. 2023 Jan 30;14(1):136–48.

[3] Saha S. Improving Software Development Using AI Enabled Predictive Analytics. Journal of Artificial Intelligence, Machine Learning and Data Science. 2024 Feb 28;2(1):1050–3.

[4] Zaman Khan S. Automated Test Case Generation and Defect Prediction: Enhancing Software Quality Assurance through AI-Driven Testing Automation. 2023.

[5] Lessmann S, Baesens B, Mues C, Pietsch S. Benchmarking classification models for software defect prediction: A proposed framework and novel findings. In: IEEE Transactions on Software Engineering. 2008. p. 485–96.

[6] Lundberg SM, Allen PG, Lee SI. A Unified Approach to Interpreting Model Predictions [Internet]. Available from: https://github.com/slundberg/shap

[7] Wiese B, Omlin C. Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks.

[8] Thota MK, Shajin FH, Rajesh P. Survey on software defect prediction techniques. International Journal of Applied Science and Engineering. 2020 Jan 1;17(4):331–44.

[9] Almazroi AA, Ayub N. Online Payment Fraud Detection Model Using Machine Learning Techniques. IEEE Access. 2023;11:137188–203.

[10] Jadhao RR, Chitragar P, Kamble D. A chronological review of heat transfer enhancement using inserts in channel flows. Vol. 100, Physica Scripta. Institute of Physics; 2025.

[11] Jadhao RR, Chitragar P, Kamble D. Overview of the future perspective of aerofoil-based passive heat transfer enhancement. Vol. 7, Engineering Research Express. Institute of Physics; 2025.

[12] Garousi V, Jafarov Z, Buğra Keleş A, Değirmenci S, Özdemir Testinium AŞ E, Zarringhalami R. AI-powered software testing tools: A systematic review and empirical assessment of their features and limitations.

[13] Esteves G, Figueiredo E, Veloso A, Viggiato M, Ziviani N. Understanding machine learning software defect predictions. Automated Software Engineering. 2020 Dec 1;27(3–4):369–92.

[14] Jadhao RR, Chitragar P, Kamble D. Comparative Analysis of Aerobeads and Aerofoil Shapes for Heat Transfer Enhancement in Heat Exchanger Systems. In: International Mechanical Engineering Congress and Exposition-India. American Society of Mechanical Engineers; 2025. p. V003T03A031.