ENHANCING FINANCIAL INSIGHT THROUGH MACHINE LEARNING-DRIVEN KPI CORRELATION ANALYSIS

Ilona Veitaitė[1], Audrius Lopata[2], Saulius Gudas[3],
[1] Institute of Social Sciences and Applied Informatics, Vilnius university
[2] Faculty of Informatics, Kaunas University of Technology
[3] Institute of Data Science and Digital Technologies, Vilnius university
ilona.veitaite@knf.vu.lt

The research explores how financial auditing is being reshaped by the rapid expansion of digital data and the integration of artificial intelligence (AI) and machine learning (ML). Traditional methods that rely on manual checks and limited data samples are no longer sufficient to ensure reliability in today's complex financial environment. By leveraging KPI ratio correlation analysis, predictive algorithms, and process mining, auditing can move beyond retrospective reviews and provide continuous, data-driven assurance. This shift, sometimes referred to as "Auditing 2.0," enhances the detection of anomalies and strengthens financial oversight. A central theme is the role of KPI correlation analysis in uncovering subtle links between financial indicators. These relationships are valuable for identifying anomalies, assessing financial health, and predicting potential risks before they escalate.

The study emphasizes that while simple ML models such as decision trees offer transparency, they often fail to capture the complexity of financial data. More advanced techniques, including neural networks and Long Short-Term Memory (LSTM)-based systems, excel in predictive tasks but present challenges in terms of interpretability. Hybrid approaches that combine both simplicity and sophistication appear to be the most effective, offering a balance between accuracy and explainability.

The research also highlights the importance of continuous auditing, where monitoring and testing occur in real-time rather than at periodic intervals. This proactive approach allows errors and fraudulent activities to be detected and addressed immediately, reducing the possibility of them becoming embedded in financial statements. The integration of AI-driven process mining adds further depth, enabling auditors to model and analyze business processes more effectively, compare expected workflows with actual performance, and provide near-instant assurance after critical business events.

Finally, an experimental analysis of KPI datasets revealed that most indicators are not strongly correlated, suggesting that a broad range of them can provide unique insights for ML applications. However, correlation analysis alone is insufficient for identifying the most valuable features; additional statistical and feature-selection methods are needed to refine the inputs and improve model reliability.

The findings point to a future where AI-enhanced auditing combines advanced analytics, transparency, and regulatory alignment to improve trust and effectiveness, while also raising important questions about fairness, security, and ethical use of AI in financial decision-making.

Introduction

Financial auditing: foundation of governance. It ensures transparency and reliability in business operations.

Shift to data-driven auditing: Auditing 2.0. Focused on continuous, real-time oversight enabled by AI and ML.

Challenge: massive financial datasets from ERP, logs, reporting tools. Such data volume overwhelms manual approaches.

Traditional methods insufficient for anomaly detection at scale. They cannot uncover subtle or emerging risks in time.

Auditing 2.0

- **Real-time monitoring vs. periodic reviews.** Continuous oversight reduces delays in fraud detection.
- **KPI correlation**: detect irregularities early. It uncovers unexpected deviations in performance metrics.
- **AI algorithms**: flag anomalies in liquidity, profitability, leverage. Early warning signals guide auditors proactively.
- **Predictive analytics**: forecast risks and inefficiencies. Helps organizations prepare for future challenges before they escalate.

Related Works: Process Mining

- Enables discovery and visualization of financial processes. Auditors gain insights into actual vs. intended workflows.
- Helps detect inconsistencies and ensure compliance. It creates transparency in complex transaction chains.
- Behavioral Change Indicators (BCIs) monitor KPI deviations. BCIs capture subtle shifts in performance that may suggest risk.

Related Works: ML vs AI

- Simple ML (logistic regression, decision trees): interpretability & compliance. Auditors can easily explain model outputs.
- Advanced AI (neural networks, LSTM): higher predictive accuracy. Effective in modeling non-linear financial dynamics.
- Hybrid models: combine interpretability + predictive power. Improve trustworthiness of AI-driven audits.
- XAI (e.g., Anchor XAI, attention-weighted boosting) enhances transparency. These tools bridge the gap between complexity and clarity.

Financial Data Auditing: Continuous Auditing

Continuous auditing modifies nature, timing, extent of audits.

Nature –	automated, ongoing verification. Ensures that processes are always checked.
Timing –	real-time detection. Problems are identified instantly instead of after months.
Extent –	full dataset analysis. No longer limited to small, random samples.

Moves from reactive to proactive auditing. Auditors prevent risks before they cause harm.

Financial Data Auditing: Continuous Assurance

Combines continuous auditing + continuous monitoring. This ensures both oversight and accountability.

Technology-enabled near-instantaneous assurance. Decision-makers get timely, relevant feedback.

Process mining as audit evidence. Data-driven evidence strengthens audit validity.

Strengthens oversight and regulatory compliance. Organizations achieve higher trust and resilience.

Event Logs in Auditing

- **Enable off-line and real-time analysis.** Support both historical reviews and live monitoring.
- **Support process discovery.** Reveals hidden workflows and actual practices.
- **Conformance checking ensures compliance.** Detects deviations from intended process models.
- **Anomaly detection through event data.** Early signals highlight unusual or fraudulent transactions.

Balancing Real-Time Monitoring and Independence

Real-time interventions may blur auditor independence. Intervening too early risks conflicts of interest.

Ensuring objectivity while preventing fraud/errors is challenging. Auditors must maintain oversight roles.

Requires frameworks to separate monitoring from intervention. Protects credibility of audit outcomes.

Machine Learning in Financial Data

Success factors: accuracy, generalization, resistance to overfitting. *These determine practical reliability of models.*

Balance in model complexity is key. *Models must adapt without becoming unstable.*

Risks:
Overfitting – too tailored to training data. *Makes models useless in real-world applications.*
Underfitting – poor predictive performance. *Leads to shallow, inaccurate insights.*

Strategies to Prevent Overfitting

Hold-out & cross-validation	• Allow fair model evaluation on unseen data
Data augmentation	• Expands dataset size artificially when data is scarce
Feature selection	• Reduces unnecessary complexity by focusing on key attributes
Regularization (L1/L2)	• Prevents overemphasis on noise within training data
Early stopping & reduced complexity	• Stops training at optimal point to ensure stability

Strategies to Improve Accuracy

More data samples	• Larger datasets increase statistical confidence
Adding contextual information	• Provides richer insights from each data point
Hyperparameter tuning	• Adjusts model parameters for optimal results
Cross-validation	• Ensures models perform well on multiple subsets of data
Testing different algorithms	• Helps identify the most suitable approach for the problem

KPI Correlation Analysis (Experiments)

- **Scope**: 56 KPIs analyzed across 12 companies. This broad dataset ensures representative financial coverage.
- **Observation**: 49 KPIs not strongly correlated. Indicates that most metrics carry distinct information useful for modeling.
- **Redundancy check**: Some KPIs were correlated by definition (e.g., total equity vs. opening equity). Such links reflect structural dependencies rather than analytical overlap.
- **KPI deltas**: Analysis of changes revealed similar weak correlations. Confirms KPIs remain independent across time-based variations.
- **Visualization**: Correlation heatmaps showed sparse clusters. Most KPIs formed independent patterns, signaling suitability for ML applications.
- **Implication**: A reduced KPI set built on weakly correlated indicators improves ML model stability and avoids redundant features.

Findings from Experiments

Main Result: Correlation analysis alone is insufficient for KPI selection. *Although many KPIs were independent, this method cannot identify the most predictive ones.*

Unique Information: Majority of KPIs retained standalone value. *This strengthens their role in feature sets for ML-driven auditing.*

Cross-company consistency: Results were repeated across multiple firms. *Confirms generalizability and robustness of findings beyond single datasets.*

Impact on ML: Smaller KPI sets improve generalization and reduce overfitting. *Stable KPI inputs also increase predictive accuracy.*

Next Steps: Feature selection and advanced statistical techniques are required to refine KPI subsets further.

Conclusions

- ML & AI transform auditing into continuous, proactive oversight. This shifts audits **from detection to prevention**.
- Hybrid models **balance accuracy**, interpretability, and compliance. They address both technical and regulatory needs.
- KPI analysis shows most indicators provide **unique insights**. It suggests that **redundancy is limited** in financial datasets.
- Continuous auditing improves **fraud detection** and **risk management**. Companies become more resilient to financial shocks.

Conclusions

- **Challenges** remain: explainability, ethical risks, and regulatory alignment. These are barriers to broader adoption.
- **Fairness auditing** and **secure AI deployment** are critical next steps. They help build trust in AI-driven assessments.
- Correlation analysis alone is **insufficient**; **advanced feature selection** is needed. This ensures predictive strength and model stability.
- **Future vision**: integration of process mining, AI, and ML will define auditing. This evolution ensures transparent, reliable, and adaptive financial oversight.

Acknowledgements. *This paper presents part of results of research project "Enterprise Financial Performance Data Analysis Tools Platform (AIFA)". The research project was funded by European Regional Development Fund.*