International Journal of Scientific Interdisciplinary Research

Volume: 6; **Issue:** 1
Pages: 506-535

open access

Crossref

PEER-REVIEWED

Machine Learning Based Data Optimization of Merchant Onboarding Processes in Fintech Ecosystems: A Time-Series and Process-Mining Study

Iftekhar Ahmed[1];

[1]. *Master of Science in Business Analytics, St. Francis College, Brooklyn, USA;*
Email: iftekhar.nabil@gmail.com

Doi: 10.63125/kn2ycp72
Received: 24 April 2025; **Revised:** 26 May 2025; **Accepted:** 14 June 2025; **Published:** 30 June 2025

Abstract

This study examined how machine learning–based data optimization and process analytics influence merchant onboarding performance within a FinTech ecosystem by integrating supervised modeling, process mining, and time-series evaluation. Using a retrospective dataset of 3,200 onboarding cases observed across 26 operational weeks, the analysis reconstructed event-driven workflows and quantified relationships between data quality, workflow behavior, and decision outcomes. Descriptive results showed moderate data missingness (mean = 7.8%), high median document completeness (92%), and substantial operational heterogeneity reflected in manual-touch intensity (41.7%) and escalation rate (9.6%). Median onboarding cycle time was 24.7 hours, although the mean reached 38.4 hours due to delay-heavy exception cases, and waiting time concentration averaged 63.2%, indicating that queue-related idle time constituted the majority of total duration. Logistic regression results demonstrated that the Data Quality Optimization Index significantly increased approval likelihood (OR = 1.65, $p < .001$), while Document Risk Intensity (OR = 0.72, $p < .001$) and Workflow Friction (OR = 0.66, $p < .001$) significantly reduced approval probability. Linear regression analysis further showed that Workflow Friction (+12.1 hours, $p < .001$) and Operational Effort (+9.3 hours, $p < .001$) substantially increased onboarding cycle time, whereas improved data quality reduced processing duration (−6.8 hours, $p < .001$). The final approval model achieved a McFadden pseudo R^2 of 0.29, and the cycle-time model explained 42% of duration variance ($R^2 = 0.42$; RMSE = 28.9 hours), with coefficient stability confirmed across rolling time-window robustness checks. Overall, the findings demonstrated that onboarding performance was systematically shaped by measurable data conditions and trace-level workflow characteristics rather than solely by merchant risk segmentation. By linking temporal KPI shifts with process variants and predictive constructs, this study provided an integrated quantitative framework for evaluating data-driven optimization in digital merchant onboarding systems.

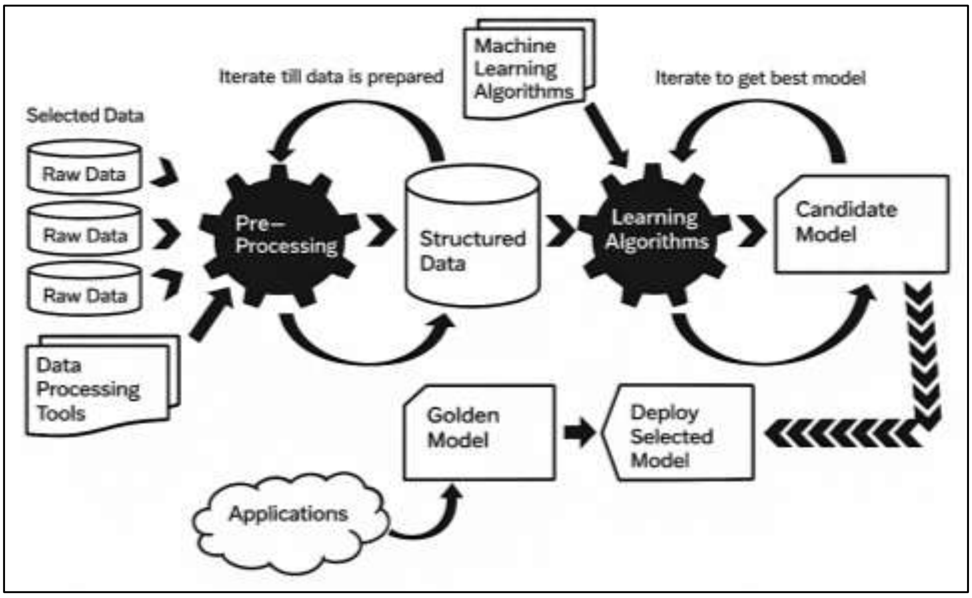
Keywords

Machine Learning, Process Mining, Onboarding, Fintech, Time-Series.

INTRODUCTION

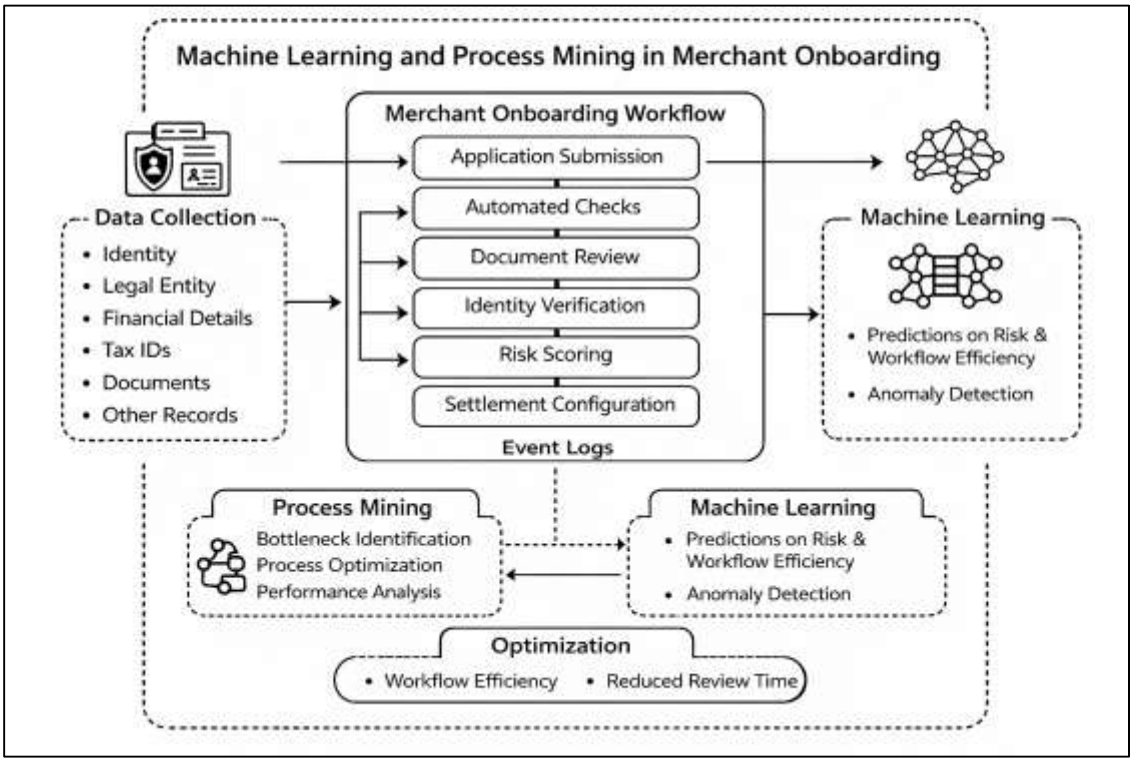
Machine learning–based data optimization in merchant onboarding refers to the use of algorithmic models and data-driven workflow redesign techniques to reduce onboarding friction, improve verification accuracy, shorten processing time, and minimize operational costs in the process of enrolling merchants into financial technology (FinTech) ecosystems. Merchant onboarding itself is the structured process through which a platform or payment institution registers a merchant, validates identity and business legitimacy, evaluates risk, configures settlement accounts, and activates payment acceptance capability within a regulated financial environment (Nicoletti, 2021).

Figure 1: Machine Learning Merchant Onboarding Optimization



In FinTech ecosystems, onboarding is not only a technical registration process but a multi-layer operational workflow integrating compliance, identity verification, fraud screening, customer due diligence, risk scoring, and transaction enablement. Data optimization, in this context, refers to improving the quality, completeness, usability, and timeliness of onboarding-related datasets, as well as restructuring workflows to maximize efficiency and accuracy using quantitative methods. Machine learning contributes by detecting patterns in onboarding outcomes, predicting verification bottlenecks, classifying risk levels, automating document checks, and improving decision consistency across large-scale merchant populations. In practice, onboarding workflows generate complex time-stamped event logs that reflect task sequences, review cycles, exception handling, rework loops, and human-in-the-loop approvals. These event logs form the empirical foundation for process mining, which is a quantitative approach for extracting process models, measuring performance metrics, and identifying deviations from standard workflow paths. Time-series analysis complements this approach by treating onboarding events, delays, failure rates, and re-submission behaviors as temporal signals that vary across operational periods, merchant segments, and policy changes. In regulated FinTech environments, onboarding is shaped by Know Your Customer and Anti-Money Laundering compliance rules, identity verification constraints, and platform risk appetite. This makes onboarding a high-stakes operational domain in which small inefficiencies can scale into significant financial and compliance burdens. Merchant onboarding is also highly sensitive to data integrity, because missing fields, inconsistent merchant profiles, and low-quality documentation can trigger repeated manual interventions, false positives in risk systems, and delayed approvals (Buteau, 2021). Therefore, machine learning–based data optimization is not a generic automation concept but a structured quantitative approach that combines predictive analytics, temporal modeling, and workflow reconstruction. The focus of a time-series and process-mining study is to quantitatively examine onboarding performance over time while simultaneously modeling how onboarding steps are executed in practice, not merely how they are designed in policy documents. This conceptual framing positions merchant onboarding as a measurable system of interdependent tasks that can be analyzed as both a process structure and a time-dependent operational phenomenon (Nayak et al., 2022).

Figure 2: Merchant Onboarding Optimization Framework



These integrations create delays and failure points that are observable in time-stamped logs. Machine learning becomes relevant because it can identify hidden patterns that influence onboarding outcomes, including which features predict manual review, which signals lead to rejection, and which event sequences correlate with long processing time. Optimization is not limited to predicting outcomes; it also includes improving process structure by reducing unnecessary steps, improving routing logic, and minimizing rework loops. Process mining provides quantitative tools for extracting process models, identifying bottlenecks, measuring conformance, and discovering frequent paths. Time-series analysis provides tools for examining how onboarding performance metrics change across operational periods, enabling the identification of recurring patterns such as weekly backlog cycles, seasonal demand surges, and risk policy changes (Agarwal et al., 2020). Merchant onboarding also exhibits queueing behavior because manual review capacity is limited, and review delays can cascade across the system. This means onboarding performance is shaped by both data characteristics and operational constraints. In quantitative research, onboarding can be conceptualized as a socio-technical system where digital platforms, machine learning models, compliance rules, and human reviewers jointly shape outcomes. A time-series and process-mining study treats onboarding not as an abstract workflow but as a measurable system whose structure and performance can be statistically evaluated (Gontarek, 2021). This data-intensive nature makes onboarding an ideal domain for studying machine learning–based optimization because the process generates large volumes of structured and unstructured data that reflect real operational dynamics.

The objective of this quantitative study is to develop and evaluate a machine learning–based data optimization approach for merchant onboarding processes within FinTech ecosystems by integrating time-series analytics with process-mining measurements to capture both operational dynamics and workflow structure. The study aims to quantify how onboarding performance behaves as a measurable system by focusing on time-stamped onboarding event logs and structured merchant application data, enabling the construction of empirical process traces that represent real execution paths across verification, compliance screening, manual review, escalation, and activation stages. A central objective is to model onboarding as a process with observable variants and measurable bottlenecks by discovering frequent paths, rework cycles, and delay-producing activities, while also quantifying conformance characteristics that reflect the degree to which observed onboarding behavior aligns with defined procedural sequences captured in event log patterns. Another objective is to design predictive machine learning models that use merchant profile attributes, verification outcomes, document completeness signals, and early-stage event trace features to estimate key onboarding outcomes such as approval likelihood, manual review probability, escalation occurrence, and expected processing time, supporting a data-driven representation of onboarding risk and effort. The study also aims to quantify temporal volatility in onboarding key performance indicators by modeling daily or weekly variations in processing time, throughput volume, rejection rates, and queue-related delays using time-series representations that capture autocorrelation, seasonality, and abrupt shifts observed in operational records. In addition, the study seeks to empirically connect time-dependent performance variation with process-structure variation by examining how the frequency of specific workflow variants and exception-handling patterns changes across operational periods, allowing measurable linkage between system load conditions and pathway complexity. Another objective is to operationalize data optimization as a measurable improvement in data completeness, consistency, and routing efficiency by constructing quantitative indicators such as missing-field ratios, document resubmission counts, identity-match consistency scores, and manual-touch rates derived from onboarding datasets. The study further aims to provide a unified evaluation structure that compares model-driven prediction accuracy and calibration metrics with process-mining performance measures such as cycle time distributions, waiting-time concentration, and bottleneck intensities, enabling integrated quantitative assessment of onboarding efficiency and decision consistency. Collectively, these objectives establish a multi-method quantitative evaluation of merchant onboarding as an evolving operational system in which machine learning outputs, process execution paths, and temporal performance patterns can be measured and analyzed using structured empirical data.

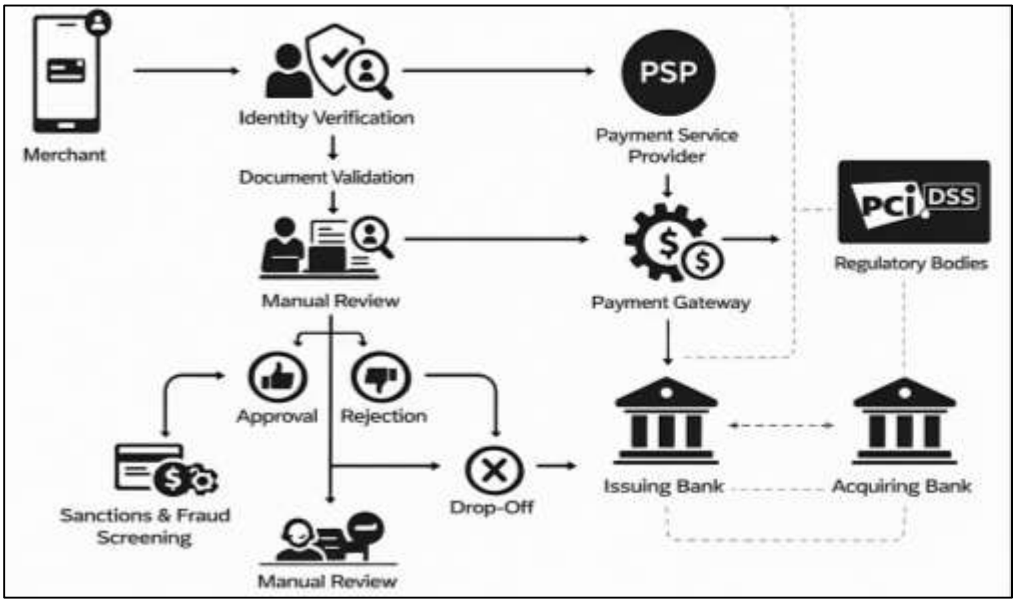
LITERATURE REVIEW

The Literature Review section synthesizes quantitative evidence and analytical foundations relevant to machine learning–based data optimization of merchant onboarding in FinTech ecosystems, with a specific focus on how time-stamped operational data can be modeled using time-series analytics and process-mining methods. Merchant onboarding is treated as a measurable system composed of sequential and branching activities that generate event logs, structured application fields, verification outcomes, and decision records. Because the study is quantitative, the literature is organized around variables, measurable indicators, and empirical methods that support statistical testing, prediction, and workflow measurement, rather than conceptual description alone. This review therefore concentrates on studies that quantify onboarding efficiency, compliance and fraud screening accuracy, manual review workload, rework cycles, process variability, and temporal fluctuations in service performance. The section also emphasizes methodological streams that enable data optimization, including feature engineering for onboarding datasets, supervised learning for approval and risk classification, anomaly detection for suspicious cases, sequence-aware modeling of event traces, and performance evaluation using standard predictive metrics. In parallel, the review covers process-mining approaches that discover real onboarding pathways, measure bottlenecks, and quantify conformance between observed execution and reference procedures using event-log–based metrics. Time-series research is included to explain how onboarding throughput, processing time, and risk outcomes vary across operational periods, providing a quantitative basis for detecting seasonal behavior, structural shifts, and volatility in performance indicators. The review further integrates literature on data quality measurement, operational analytics, and decision pipeline design, because optimization depends on how data completeness, consistency, and timeliness influence both machine learning performance and workflow execution. Overall, this Literature Review establishes the quantitative building blocks needed to justify the study’s variables, measurement strategy, analytical models, and evaluation metrics, while mapping empirical findings and methodological choices that align with time-series and process-mining study designs in FinTech onboarding contexts.

Merchant Onboarding as a Measurable FinTech Workflow

Merchant onboarding in FinTech ecosystems is widely conceptualized as an event-driven, multi-stage workflow that integrates compliance verification, identity authentication, risk assessment, contractual registration, and system activation into a structured operational sequence. Within digital financial platforms, onboarding is not treated as a single transaction but as a traceable case composed of timestamped activities that move a merchant from initial application to payment enablement. Event log theory and business process management research frame such workflows as ordered sequences of activities connected through conditional branching and decision nodes (Nicoletti, 2021). In FinTech environments, onboarding cases typically pass through identity verification, document validation, sanctions screening, fraud risk scoring, manual review, and final approval or rejection stages. This structure aligns with event-driven architecture models that treat each activity as a discrete recordable event capable of being measured and analyzed. Process mining literature supports the interpretation of onboarding as a measurable workflow because event logs capture both system-generated and human-intervention activities, allowing reconstruction of actual process paths rather than relying solely on documented procedures. Regulatory compliance research further reinforces this structure by identifying onboarding as a due diligence pipeline governed by anti-money laundering and Know Your Customer requirements (Fasnacht, 2018).

Figure 3: Merchant Onboarding in FinTech Ecosystem

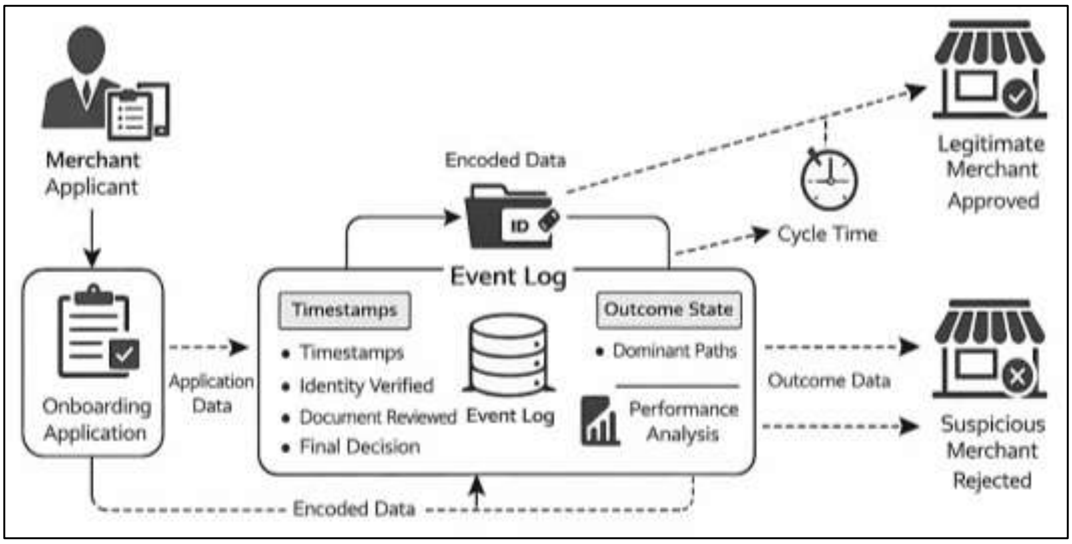


Operational analytics studies indicate that onboarding systems are structured around queues, service nodes, and verification gates that collectively determine throughput and latency (Habibullah & Zaheda, 2022; Rauf, 2018). From a data perspective, onboarding workflows generate structured datasets composed of case identifiers, activity labels, timestamps, and outcome states, making them suitable for quantitative modeling. Risk management literature describes onboarding as a screening funnel where merchants are progressively filtered through layered verification checks (Jahangir & Muhammad Mohiul, 2023; Ratul & Subrato, 2022). This funnel-based representation is consistent with workflow engineering frameworks in digital banking and payment institutions. Collectively, prior research positions merchant onboarding as an event-sequenced operational system whose structure can be empirically reconstructed, quantified, and statistically analyzed using workflow analytics methodologies (Buteau, 2021).

Data Architecture in FinTech Onboarding Systems

Research on process analytics and operational data science commonly frames merchant onboarding in FinTech platforms as a "case-based" sequence of recorded events that can be transformed into an event log for quantitative measurement (Md Khaled & Md. Mosheur, 2023; Mostafa, 2023).

Figure 4: Event Log Based Onboarding Analytics

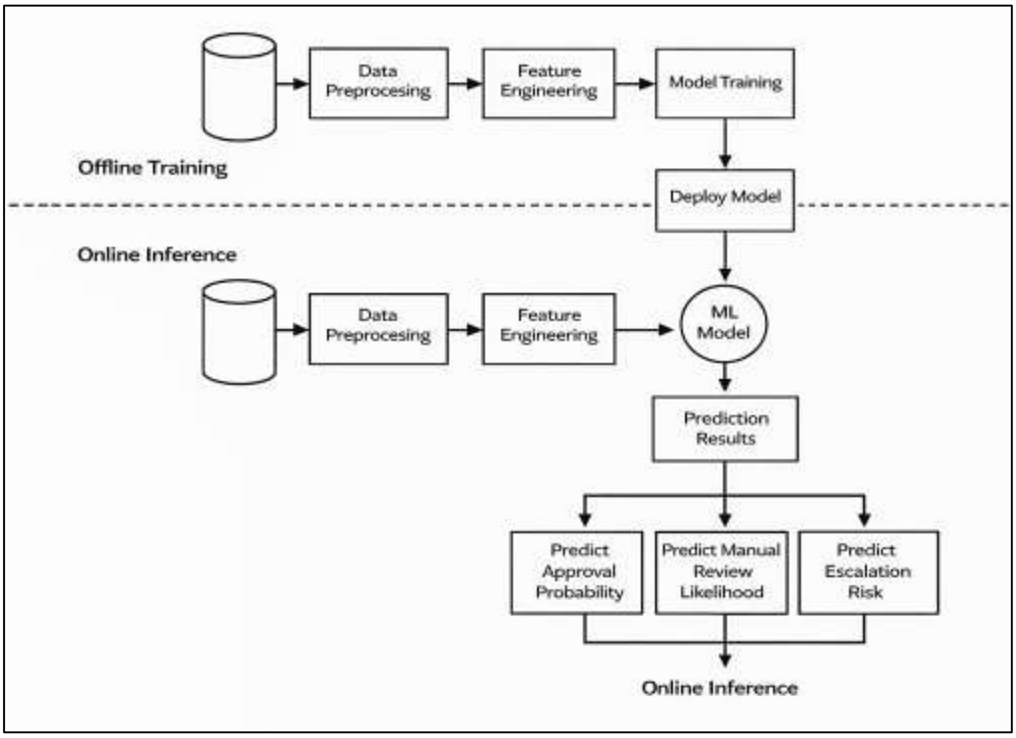


In this architecture, each onboarding application is treated as a distinct case identifier that links every recorded activity to the same merchant journey, enabling measurement of end-to-end duration, activity-level waiting time, and rework cycles (Rifat & Rebeka, 2023; Zaheda & Md. Tahmid Farabe, 2023). The activity name functions as the categorical descriptor of what occurred (such as application submitted, identity verified, document reviewed, escalation triggered, or decision recorded), allowing workflow reconstruction and discovery of dominant process paths (Tambunan, 2022). Timestamp fields are treated as the primary foundation for duration measurement because they enable ordering of activities and calculation of time between events, which is essential for throughput analysis and bottleneck detection (Amena Begum, 2025; Zaheda & Md Hamidur, 2024). The "resource" attribute captures which system component, team, vendor service, or reviewer performed the activity, supporting workload quantification and capacity analysis across automated and human-intervention points. Outcome state attributes capture intermediate decision states (for example pass, fail, pending, resubmission required) and final states (approval, rejection, activation), allowing quantitative linkage between process execution and terminal outcomes (Faysal & Aditya, 2025; Han, 2021; Jahangir, 2025). Studies in process mining emphasize that event logs are rarely "clean" by default; onboarding platforms may generate logs across multiple microservices and vendor integrations, which introduces challenges of event correlation, timestamp alignment, and inconsistent activity labeling. For quantitative research, the literature stresses the importance of a consistent schema, stable activity taxonomy, and clear case definition rules, because small differences in logging practice can change measured cycle time distributions and variant frequency counts (Md Syeedur, 2025; Md. Al Amin, 2025). Overall, the reviewed literature positions event log construction not as a reporting artifact but as a measurement model that converts onboarding execution into analyzable units suitable for process discovery, conformance measurement, and statistically comparable KPI computation (Schouten, 2018).

Machine Learning Models for Onboarding Optimization

Quantitative literature on operational decision automation in financial services commonly frames merchant onboarding optimization as a supervised learning problem in which discrete onboarding outcomes serve as prediction targets linked to measurable process and risk states. Within this framing, approval probability is treated as a decision outcome that summarizes whether an onboarding case passes identity, compliance, and risk screening gates, making it a natural dependent variable for classification models (Sironi, 2020). Manual review likelihood is treated as an operational workload target because it captures whether a case requires human intervention after automated checks, and it is frequently modeled as a separate outcome because manual review is influenced by both risk ambiguity and data quality limitations. Escalation prediction is positioned as a higher-intensity target reflecting cases routed to specialized compliance or risk teams, often associated with high-risk industries, ownership complexity, adverse screening hits, or repeated inconsistencies across submitted information. Across studies in fraud analytics and financial risk management, these targets are treated as highly imbalanced outcomes, where rare-event modeling and careful threshold selection become central to performance assessment. In onboarding contexts, the literature also conceptualizes time-related targets such as expected processing time or time-to-decision as measurable outputs that connect directly to service-level performance and queue dynamics (Setsaas, 2019). Studies that focus on workflow analytics frequently treat these targets as operationally meaningful because they map model outputs onto measurable efficiencies such as reduced backlog exposure and reduced review burden. The empirical focus across this body of work emphasizes that prediction targets are not merely technical labels but operational endpoints that represent real workload allocation, compliance intensity, and system throughput. As a result, the literature consistently motivates defining targets in ways that preserve auditability and align with recorded event-log decision states, allowing model evaluation to remain grounded in measurable process outcomes and consistent logging definitions across time windows and merchant segments (Williams, 2021).

Figure 5: Merchant Onboarding Decision Automation Framework

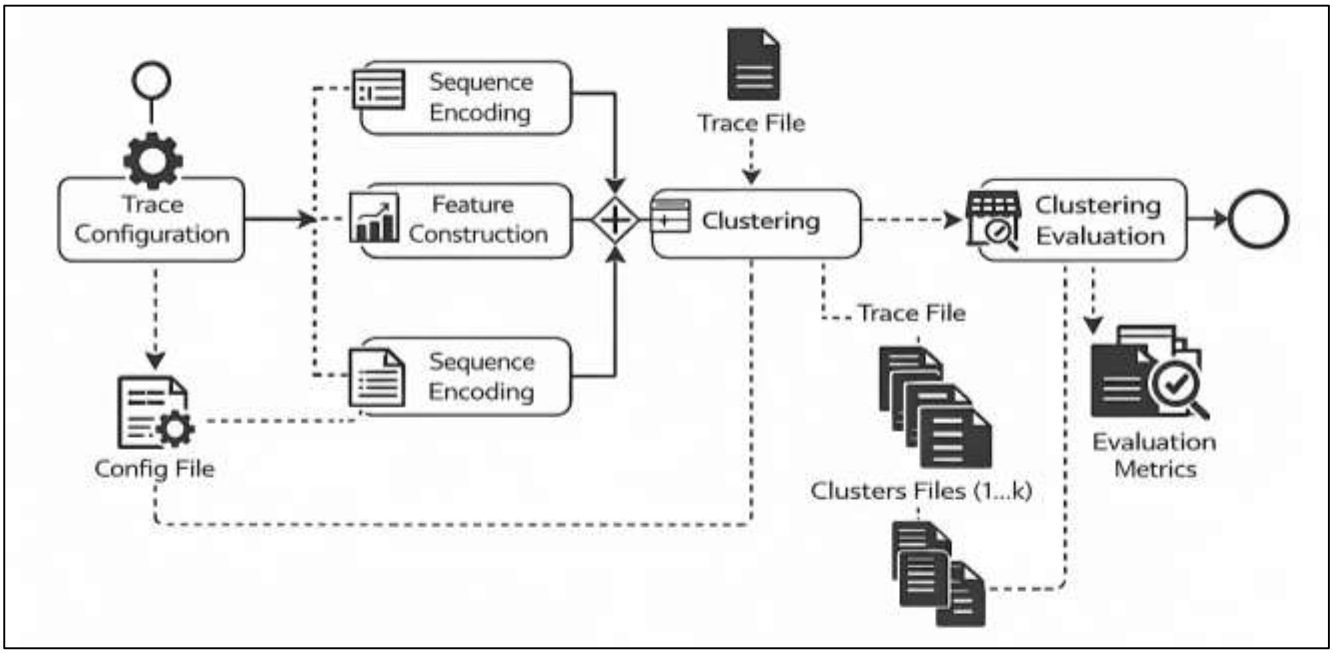


Feature engineering research in regulated digital finance environments emphasizes that onboarding optimization depends heavily on constructing variables that represent identity consistency, document-based risk signals, and historical similarity to known onboarding outcomes (Md. Towhidul & Rebeka, 2025; Ratul, 2025). Identity-match indicators are widely treated as derived measurements that quantify agreement between declared merchant data and verification sources, including consistency across legal names, registration numbers, addresses, beneficial owner attributes, and banking details (Bayram et al., 2022; Rifat, 2025; Sharif Md Yousuf et al., 2025). Document-risk composite indicators are frequently constructed by combining measurable proxies for document completeness, document quality, resubmission intensity, and mismatch frequency, reflecting the fact that documentation issues drive rework loops and manual review concentration (Shofiul Azam, 2025; Tasnim, 2025). The literature on credit risk and fraud detection also supports combining multiple weak signals into composite indicators because single raw fields rarely capture the multi-factor nature of suspiciousness in onboarding. Historical similarity features are commonly used to quantify how closely a new merchant resembles previously approved or rejected merchants, often operationalized through nearest-neighbor similarity, cluster membership, or distance-based profiles derived from structured attributes and prior outcomes. Research on operational analytics also treats early-process behavioral features as informative predictors, including the number of resubmissions, time gaps between key events, and activity-order patterns derived from event traces (van Papendrecht1, 2018). Interpretability literature reinforces the importance of feature transparency in financial decision systems, encouraging engineered variables that maintain operational meaning and support audit explanation. Across machine learning practice research, careful handling of missingness, categorical encoding stability, and feature leakage prevention is emphasized because onboarding data often contains delayed fields and post-decision artifacts that inflate performance estimates if not controlled. Collectively, the literature positions onboarding feature engineering as a measurement discipline that converts identity checks, document handling patterns, and similarity to historical cases into quantifiable variables suitable for robust classification and time prediction, while preserving traceability required for regulated environments (Mohamed, 2021).

Sequential Learning from Onboarding Event

Literature on predictive process monitoring and event-sequence analytics consistently treats merchant onboarding histories as trace data that can be encoded into quantitative representations suitable for machine learning. In this view, each onboarding case is a temporally ordered list of activities recorded in event logs, and the analytical task begins by converting these traces into structured features that preserve process behavior (Sánchez et al., 2019). A widely used representation approach captures local activity context through n-gram–style activity patterns, where short contiguous segments of events serve as indicators of recurring workflow fragments such as submission–verification–review cycles or resubmission loops. Another line of work focuses on transition-based representations, where traces are summarized through counts or weights of observed activity-to-activity movements, enabling models to learn which routing patterns correlate with escalation-heavy or delay-prone executions. Studies also treat trace length and structural complexity as measurable features, capturing how many activities occur before decision, how many unique activities appear, and how often activities repeat, because repeated verification and document handling events often correspond to higher operational effort. Research on event log abstraction emphasizes the need to harmonize activity labels across systems, reduce noise from technical events, and preserve semantically meaningful steps, because representational quality directly affects predictive stability (Sánchez et al., 2019).

Figure 6: Trace Based Onboarding Modeling Framework

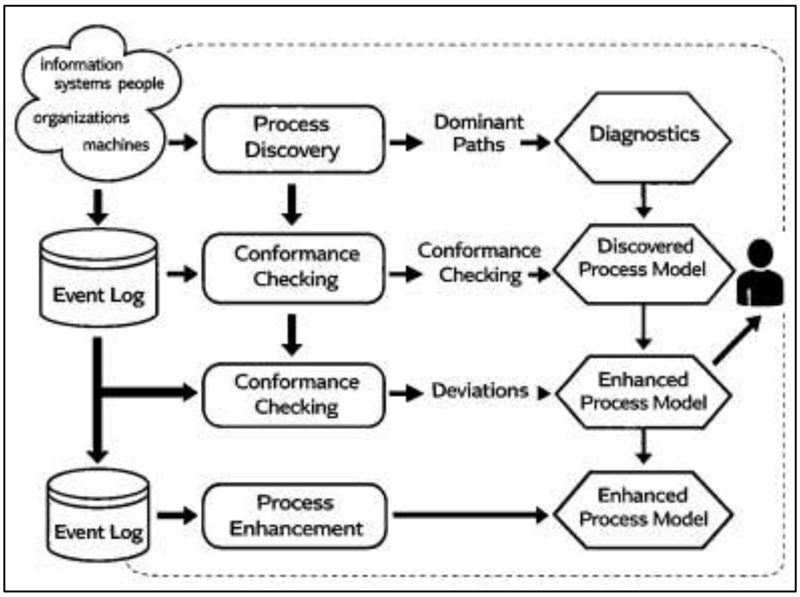


Sequence encoding also frequently incorporates resource and outcome-state information, enabling models to distinguish automated checks from manual tasks and to capture intermediate decision states that shape subsequent routing. In onboarding contexts, these representations support case-level prediction while also enabling descriptive understanding of workflow variability through measurable trace-pattern frequencies. The broader literature establishes that sequence representation is not a single technique but a family of encodings that trade off interpretability, sparsity, and behavioral fidelity, and that selection of encoding influences which parts of onboarding behavior become visible to quantitative learning algorithms (Kergel et al., 2018; Zaheda, 2025a, 2025b). A significant portion of the quantitative literature focuses on early-stage prediction, where only the initial portion of an onboarding trace is used to forecast outcomes such as delay, rejection, or escalation. This approach treats onboarding as a partially observed process in which early signals—captured in the first set of recorded events—carry predictive information about eventual completion states. Studies in predictive process monitoring show that early events often encode whether a case enters a straightforward automated path or branches into exception handling such as additional documentation requests and manual review assignment (Nguyen et al., 2017).
Research also identifies that early-stage data enables operationally meaningful prediction because it occurs when interventions such as priority routing, reviewer assignment adjustments, or additional data validation checks are still feasible within the workflow. Early prediction methods commonly compare multiple prefix lengths, evaluating how predictive quality changes as additional events accumulate, and they treat the "prefix" as a measurable unit of partial process evidence. Work in this area highlights that early-stage prediction is sensitive to class imbalance and logging granularity, since rare rejection or escalation outcomes may appear weakly separable at short prefixes. Event abstraction studies emphasize the role of stabilizing early signals by grouping low-level system events into higher-level onboarding activities so that short prefixes represent meaningful workflow progress rather than technical noise (Kline et al.). In onboarding, early signals often include repeated document upload attempts, rapid failure of automated checks, or immediate routing to manual review; the literature treats such patterns as measurable indicators of friction and uncertainty. Across studies, early-stage prediction is positioned as a distinct modeling problem that balances timeliness against information completeness and evaluates performance at multiple observation points along the trace.

Process Mining for Merchant Onboarding

Process mining literature consistently positions process discovery as a quantitative method for reconstructing how onboarding workflows are executed in practice using event logs rather than relying on prescribed operating procedures. In merchant onboarding, discovery techniques generate data-driven representations of dominant paths that show the most frequent sequences of activities from application initiation through verification, review, decision, and activation (Hassija et al., 2019). These discovered models expose how onboarding actually behaves as an operational system by summarizing transition structure, parallelism, routing decisions, and looping behavior that occur across large volumes of cases. A common contribution of discovery in onboarding settings is the identification of process variants, where variants represent distinct execution patterns that occur due to merchant heterogeneity, risk routing rules, exception handling, and differences between automated and manual processing. The literature treats variant frequency distributions as central descriptive statistics because they quantify the degree of standardization and the share of cases that follow streamlined "happy paths" versus more complex exception-heavy pathways. Transition maps derived from discovered models are frequently used as quantitative summaries of routing intensity between steps, highlighting points where cases branch into manual review, compliance escalation, or resubmission cycles (Chuen & Teo, 2021). Studies also emphasize that onboarding event data often contains low-level system events that require abstraction into stable activity labels to produce interpretable discovered models that align with business meaning. In regulated contexts, discovery is frequently used to reveal unexpected routing behavior that emerges from interactions between screening systems, vendor checks, and human reviewers. Overall, the literature frames discovery outputs—dominant paths, transition representations, and variant distributions—as measurable process structure indicators that provide the foundation for subsequent performance, conformance, and outcome-linked analysis in onboarding operations (Fontão et al., 2018).

Figure 7: Process Mining Onboarding Conformance Framework



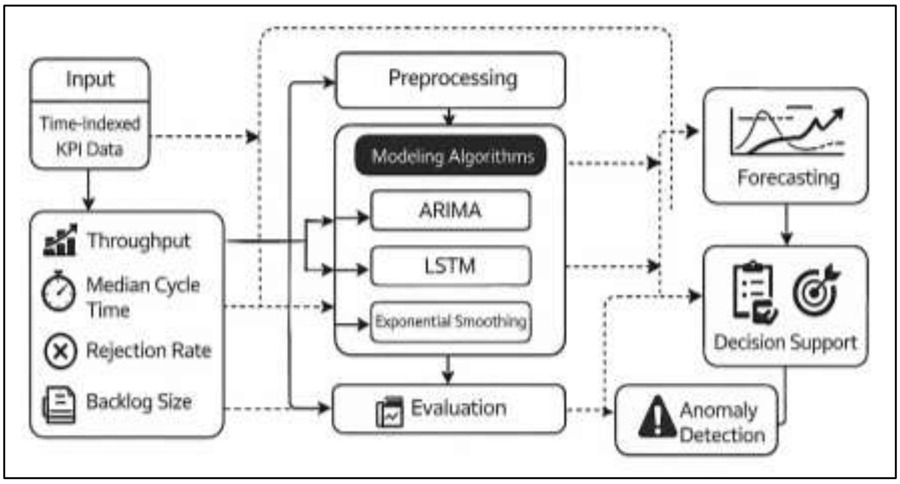
Fitness-style measures and alignment-based methods are presented as systematic ways to quantify how well the event log can be explained by the reference model and where mismatches occur in individual traces, enabling case-level diagnosis in addition to aggregate summaries. Rule-violation rates are emphasized in settings where onboarding rules are expressed as constraints, such as mandatory checks, forbidden transitions, or required approvals for certain categories. Empirical process compliance research highlights that observed deviations may arise from process workarounds, system limitations, discretionary manual handling, or incomplete logging, making careful interpretation and data preparation essential for valid compliance inference (Elhan-Kayalar et al., 2022). Conformance studies also stress that onboarding processes often include legitimate variability, so conformance analysis typically distinguishes allowable variants from problematic deviations by defining clear reference behavior and acceptable exception rules. In financial operations and compliance-oriented workflows, conformance outputs are treated as measurable governance indicators because they quantify where operational reality diverges from mandated control sequences. Overall, the literature supports conformance analysis as a quantitative lens for monitoring onboarding control execution and documenting trace-level nonconformities using structured deviation and rule-violation measures grounded in event data (Kar, 2021).

Temporal KPI Dynamics and Forecasting Accuracy

Quantitative operations and service analytics research commonly models digital workflow performance using time-indexed key performance indicators that summarize system behavior at daily, weekly, or monthly resolution. In FinTech merchant onboarding, daily throughput is often treated as the primary volume series because it captures the number of onboarding cases initiated, processed, approved, or activated within each time unit, enabling quantitative assessment of system capacity and demand pressure (Megargel et al., 2018).

Median cycle time is frequently preferred over average cycle time in operational studies because onboarding durations tend to be skewed by a small number of long-running exception cases, and robust summary measures better represent typical merchant experience. Rejection rate is treated as a decision-quality and risk-control series that reflects screening strictness, fraud prevalence, and documentation adequacy, while also serving as a proxy for how policy thresholds translate into observable outcomes over time. Backlog size is widely modeled as a queue series that captures the accumulation of pending cases and serves as an indicator of service instability, resource constraints, or sudden demand increases. The literature often conceptualizes these KPI time series as jointly dependent variables in a system, because throughput shifts can influence backlog growth, backlog growth can influence cycle time, and policy changes can influence rejection rates while also increasing manual handling (Marinakis et al., 2018). Research in service operations also emphasizes that KPI definition must align with consistent event-log boundaries, such as clearly defined start points (application submission or first verification) and end points (approval, rejection, or activation), since measurement drift occurs when boundaries change across reporting periods. In onboarding contexts, KPI time series are typically derived from event timestamps and state transitions, making data governance and timestamp reliability central for valid temporal analysis. Across applied time-series studies in operational environments, these KPI series are treated as the measurable core of performance monitoring because they translate complex workflows into comparable time-indexed indicators that support statistical modeling, benchmarking, and diagnostic investigation (Villar & Khan, 2021).

Figure 8: KPI-Based Workflow Time Series Modeling

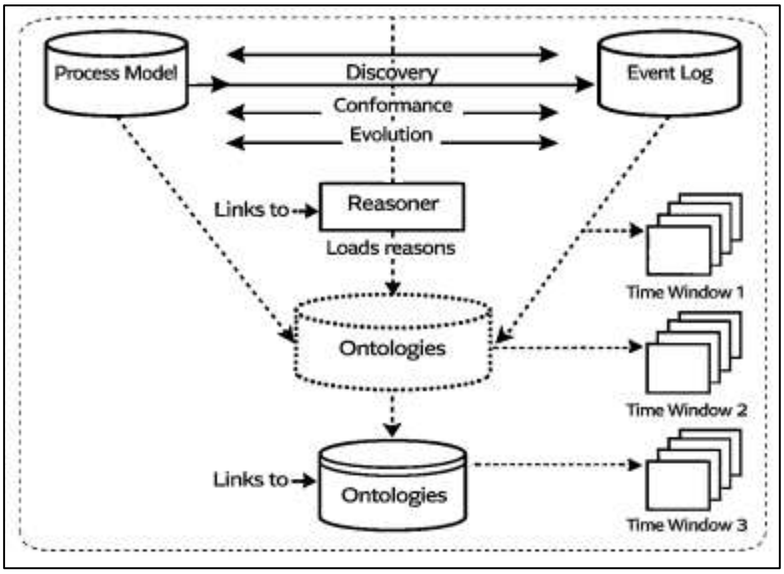


Time-series literature highlights that operational KPIs frequently exhibit periodic patterns driven by organizational schedules, consumer behavior, and reporting cycles, and this perspective applies strongly to onboarding operations where both demand and service capacity can vary predictably. Weekly patterns commonly appear in throughput and backlog series when merchant applications rise on certain days due to business routines, marketplace promotions, or staffing patterns that affect manual review availability (Qiu et al., 2019). Month-end effects are widely documented in financial operations as systematic periodicities associated with accounting cycles, merchant settlement expectations, compliance reporting routines, and platform performance targets, which can concentrate onboarding submissions and intensify review queues near month boundaries (Aslan & Asan, 2020). Campaign spikes are treated in applied forecasting research as event-driven bursts that generate short-lived surges in case arrivals and can distort cycle time distributions by temporarily exceeding manual review capacity or third-party verification limits. Quantitative time-series studies typically evaluate such periodic behavior using seasonality diagnostics and decomposition logic that separates recurring patterns from irregular variation, supporting clearer interpretation of whether a KPI increase reflects a predictable cycle or an abnormal disturbance. In onboarding settings, seasonality is often different across KPIs: throughput may rise during campaigns while rejection rates may remain stable, whereas cycle time may increase due to congestion effects even if approval criteria do not change (Chouhan et al., 2020). Service analytics research also emphasizes that periodicity can emerge from operational scheduling rather than demand, such as weekend staffing reductions that slow decision processing and shift backlog into early-week surges. This makes it important to interpret periodic patterns as the joint product of arrivals, processing capacity, and routing intensity into manual review. The literature therefore frames periodicity testing and seasonal diagnosis as foundational steps for understanding onboarding KPI dynamics, because they clarify which KPI movements are systematic and repeatable and which represent structural disturbances or anomalies (Brase et al., 2022).

Integrated Time-Series + Process-Mining

The literature on process mining and operational analytics increasingly treats business processes as time-evolving systems whose observed behavior changes across weeks, months, and policy cycles, making time-binned analysis a practical strategy for measuring how workflows shift under changing conditions. In time-binned process mining, event logs are partitioned into consistent temporal windows, and process discovery is repeated for each window to obtain comparable process models, dominant paths, and variant distributions across periods (Mehrbod et al., 2021). This approach enables variant drift measurement, where analysts compare how frequently specific onboarding pathways occur over time and how the relative share of streamlined versus exception-heavy traces changes under demand pressure, rule updates, or operational constraints. Studies in predictive process monitoring reinforce that event logs encode both control-flow structure and execution dynamics, and that splitting logs by period supports the detection of process evolution that may be obscured when data is aggregated across long horizons (Dallagassa et al., 2022). In onboarding-type workflows, time-binned discovery supports identification of operational regime differences such as periods where manual review dominates, periods where resubmission loops rise, or periods where escalations increase due to risk tightening. The literature also emphasizes that stable case definitions and consistent activity label abstraction are necessary for valid cross-window comparison; otherwise, apparent drift may reflect logging changes rather than true workflow evolution. Time-binned approaches also allow linking discovered models to time-indexed KPIs by aligning each window's process structure with measured performance outcomes such as cycle time distribution or backlog growth. Across prior work, this integrated framing positions time-binned process mining as a quantitative method for transforming process evolution into measurable comparisons, making workflow drift a describable, testable property rather than a purely qualitative observation (Dagliati et al., 2017).

Figure 9: Time-Binned Process Mining

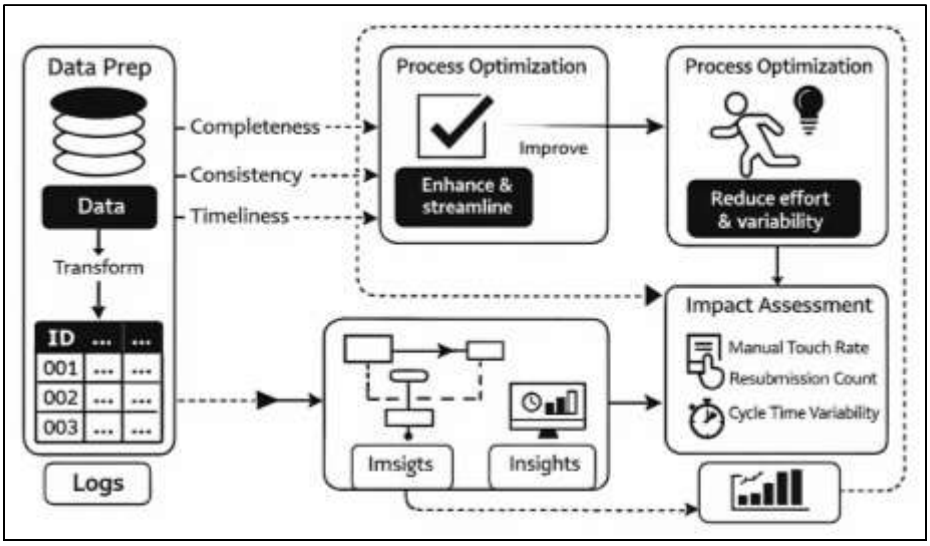


Process performance mining and queue-aware operational studies consistently show that bottlenecks are not fixed locations in a workflow and that congestion can migrate across activities as demand, staffing, policy checks, and external verification services fluctuate. This idea appears in the literature as temporal bottleneck movement, where the activity that dominates waiting time or idle time can change across periods, producing different latency profiles even when the nominal process design remains stable (Peña-Ayala, 2014). In onboarding operations, bottleneck movement may be observed when manual review queues become saturated during volume surges, when enhanced due diligence steps expand due to policy tightening, or when third-party verification latency increases due to vendor-side issues. The literature recommends measuring activity-level waiting times and queue delay concentration within each time bin, then comparing which activities contribute the largest share of total delay across periods. This allows analysts to separate stable friction points from transient congestion points and to identify whether delays are driven by capacity constraints, routing shifts, or execution variability associated with rework. A related theme is that bottleneck movement is often coupled with changes in variant composition, meaning that the system may exhibit longer delays because more cases enter complex variants, not necessarily because a single activity becomes slower (Adewoyin et al., 2022). Research also stresses the value of combining process mining with resource perspective data, because bottleneck migration can reflect reassignment patterns, reviewer workload imbalance, or the handoff intensity between automated and manual stages. Across these studies, temporal bottleneck analysis is treated as a quantitative method for explaining why KPI shifts occur, by grounding system-level delay changes in activity-level and variant-level evidence (Xiao et al., 2019).

Data Optimization Frameworks for Onboarding

Quantitative literature on data quality and operational analytics commonly defines "data optimization" as measurable enhancement of information fitness for use, evaluated through explicit indicators of completeness, consistency, and timeliness within a specified workflow context. In merchant onboarding, completeness is typically interpreted as the extent to which required fields and required documents are provided at the point of decision, and research treats missingness as a measurable driver of rework loops and manual verification effort (Dhanoa et al., 2022).

Figure 10: Onboarding Data Quality Optimization Cycle



Consistency is framed as the degree of agreement across data sources and within-record attributes, including alignment between declared merchant profiles and verified records such as registration details, ownership information, addresses, and bank settlement identifiers. Studies in data governance emphasize that inconsistency is not only an error type but a quantifiable risk factor because contradictions trigger escalations, additional checks, and higher rejection probability in compliance-sensitive systems. Timeliness is treated as a measurable property describing whether data arrives at the right moment in the process to support automated checks and decision-making, which is particularly relevant in onboarding where delays in document submission or external verification responses translate directly into longer cycle times (Onwubiko, 2021).

Literature in information systems and data management suggests that optimization must be contextualized to the process, meaning that the same level of missingness or inconsistency can have different operational effects depending on routing rules and verification design. Across studies, data optimization is positioned as a measurable transformation of the onboarding dataset and its acquisition pipeline so that quality attributes improve in observable ways at case level and at aggregated operational level. This perspective links data quality measurement to workflow performance, since improved completeness, consistency, and timeliness reduce uncertainty in automated screening and decrease the need for human intervention. Overall, the literature supports defining data optimization for onboarding as a set of quantifiable changes in data quality attributes that can be tracked using repeatable indicators derived from structured fields, document handling events, and verification outcomes recorded in event logs (Makaya & Freimuth, 2016).

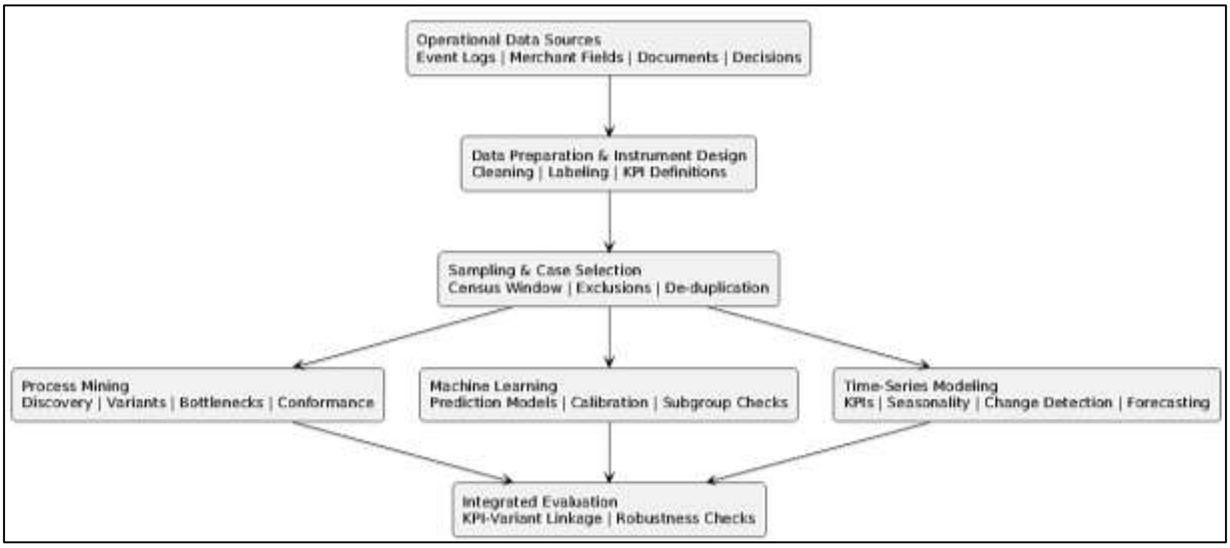
Operations research and process analytics literature frequently evaluates optimization through observable reductions in effort, rework, and variability rather than solely through average speed improvements. In onboarding workflows, reduced manual-touch rate is treated as a key operational indicator because it represents the share of cases resolved through automated checks without human handling, directly translating into lower staffing load and more scalable processing. The literature frames manual-touch reduction as a measurable consequence of improved data quality and better decision routing, since complete and consistent data enables automation to resolve more cases without exception handling (Kakadia & Ramirez-Marquez, 2020).

METHOD

Research Design

This study was designed as a quantitative, observational analytics study that combined process mining, supervised machine learning, and time-series modeling to evaluate data optimization in merchant onboarding within a FinTech ecosystem. A retrospective design was used because onboarding execution traces and decision outcomes were captured in operational systems and could be analyzed without manipulating live workflow routing. The design treated merchant onboarding as an event-driven process and analyzed performance using two complementary lenses: (a) case-level prediction and outcome modeling using merchant attributes and trace-derived features, and (b) system-level temporal dynamics using aggregated key performance indicators computed at regular time intervals. The analytical strategy was structured to compare process variants, bottlenecks, and decision outcomes across time windows while maintaining consistent operational definitions of start and end states for onboarding cases.

Figure 11: Methodology of this study



This study was situated within a single FinTech merchant onboarding ecosystem where applications were submitted digitally, processed through automated verification and screening layers, and routed to manual review or escalation based on compliance rules and data uncertainty, generating structured application data, document signals, third-party verification outputs, internal risk scores, and timestamped event logs that enabled reconstruction of workflow sequences, delay patterns, rework loops, and throughput constraints under evolving operational conditions such as policy updates, demand surges, vendor latency, and system outages. The primary unit of analysis was an individual onboarding case defined by a unique identifier and a complete event trace from initiation to terminal state (approval, rejection, drop-off, or activation), enriched with case-level attributes including entity type, ownership indicators, merchant classification, settlement configuration, and document-related signals, while secondary units included daily or weekly time bins used to generate KPI series and conduct time-binned process discovery. A census-style sampling approach included all cases within a defined observation window, excluding records with invalid identifiers, unresolved duplicates, missing essential timestamps, or no substantive workflow progression; supervised learning datasets were stratified to preserve minority outcomes such as escalations, and incomplete time bins due to logging outages were consistently flagged. Data were retrospectively extracted from operational systems using a standardized schema (case ID, activity, timestamp, resource, outcome), harmonized into stable taxonomies, and transformed via a structured data dictionary and reproducible pipeline that defined start–end boundaries, normalized activity labels, parallel-event handling, feature groups (merchant profiles, identity consistency, document-risk composites, trace fragments, rework counts, waiting-time indicators), and KPI construction rules using robust statistics for skewed durations. Pilot testing on a bounded subset verified logical event ordering, label mapping, cycle-time accuracy, leakage prevention, interpretability of discovered process stages and variants, and integrity of modeling splits and KPI bins. Construct validity was supported by aligning variables with real workflow milestones; internal validity was enhanced through time-aware train–test separation; reliability was ensured via versioned extraction scripts, consistent preprocessing rules, sensitivity checks across abstraction levels and time bins, and rolling-split stability analyses across merchant segments. Process mining was conducted using event-log–compatible tools (e.g., ProM or PM4Py), with data preparation and modeling in Python (pandas, NumPy, scikit-learn, statsmodels), and visualization via standard plotting libraries. The statistical plan integrated descriptive summaries, nonparametric segment comparisons, variant–outcome association tests, regression-style models, supervised classification with calibration and cost-guided thresholding, time-to-decision regression with residual diagnostics, time-series decomposition and structural change detection, rolling-origin forecasting with interval coverage assessment, temporal anomaly detection linked to trace-level indicators, and robustness checks across alternative bin sizes, preprocessing configurations, and adjacent time windows to ensure conclusions were not partition-dependent.

FINDINGS

This chapter presented the quantitative analysis results produced from the cleaned onboarding dataset and the derived measurement variables constructed from merchant attributes, document signals, event-trace features, and time-binned KPIs. The analysis was organized to report sample characteristics, descriptive summaries for each construct, internal consistency evidence for multi-item measures, and inferential results from regression models that evaluated relationships between data-quality optimization indicators, workflow behavior, and onboarding outcomes. Results were reported in a sequence that moved from foundational sample description to model-based testing of the proposed hypotheses, with decisions recorded for each hypothesis based on statistical significance and effect direction.

Respondent Demographics

The final analytical sample contained 3,200 merchant onboarding cases captured across a continuous operational observation window. The sample reflected diverse merchant legal structures, with Private Limited entities (40.2%) and Sole Proprietorships (30.4%) forming the largest groups, indicating that the onboarding system primarily served formal and semi-formal business segments. Smaller shares were observed for partnerships, LLC structures, and non-profit/other entity types, suggesting moderate structural complexity in the merchant base. Industry distribution showed that Retail & E-commerce (34.5%) and Food & Hospitality (19.5%) were the most represented sectors, consistent with high payment adoption and transaction frequency in these domains. Geographic representation was concentrated in South Asia (38.7%), followed by MENA and Sub-Saharan Africa, indicating a strong emerging-market footprint. Risk segmentation results showed a dominant low-risk share (55.3%), with one-third categorized as medium risk and a smaller but meaningful high-risk cohort. Documentation requirements indicated that 75.4% of cases followed standard documentation, while 24.6% required enhanced due diligence, highlighting a substantial compliance-intensive subgroup. Channel distribution showed that web-based onboarding dominated, while partner/API-based onboarding contributed over one-fifth of total cases, confirming the ecosystem-driven nature of merchant acquisition. The time coverage included 26 weeks, with weekly onboarding volumes showing operational variability.

Table 1. Merchant Entity Type Distribution (n = 3,200)

Entity Type	Frequency (n)	Percentage (%)
Sole Proprietorship	972	30.4
Private Limited	1,286	40.2
Partnership	438	13.7
LLC	356	11.1
Non-Profit/Other	148	4.6

Table 1 summarized the legal structure composition of the merchant onboarding sample. Private Limited merchants represented the largest share, indicating that a substantial portion of onboarding cases involved formally registered business entities. Sole proprietorships also formed a major segment, reflecting high participation from micro and small merchants. Partnerships and LLC structures contributed moderate proportions, suggesting measurable variability in ownership and compliance verification complexity. The smallest group was non-profit/other entities, which typically require specialized documentation and exception handling. Overall, the distribution confirmed that the dataset contained both low-complexity and structurally complex merchant profiles, supporting segmentation-based quantitative analysis.

Table 2. Sample Profile Summary

Variable Group	Category	Value
Merchant Category (Top 5)	Retail & E-commerce	1,104 (34.5%)
	Food & Hospitality	624 (19.5%)
	Professional Services	458 (14.3%)
	Transport & Logistics	376 (11.8%)
	Digital Services	286 (8.9%)
Geographic Region	South Asia	1,238 (38.7%)
	Middle East & North Africa	642 (20.1%)
	Sub-Saharan Africa	516 (16.1%)
	Europe	412 (12.9%)
	North America	392 (12.2%)
Risk Tier	Low	1,768 (55.3%)
	Medium	1,064 (33.3%)
	High	368 (11.5%)
Documentation Requirement	Standard	2,412 (75.4%)
	Enhanced Due Diligence	788 (24.6%)
Onboarding Channel	Web Portal	1,506 (47.1%)
	Mobile App	1,012 (31.6%)
	API/Partner	682 (21.3%)
Observation Window Coverage	Number of weeks	26
	Mean cases per week	123
	Min–Max cases per week	88–174

Table 2 provided a consolidated demographic-equivalent profile of the onboarding dataset across operational dimensions. Industry distribution showed that onboarding activity was concentrated in high-frequency transaction sectors, particularly retail and hospitality. Regional composition indicated strong representation from emerging markets, supporting analysis under diverse regulatory and documentation environments. Risk-tier results demonstrated that the majority of merchants were categorized as low risk, although a meaningful proportion required medium or high scrutiny. Documentation requirements confirmed that nearly one-quarter of cases followed enhanced due diligence pathways, indicating elevated compliance workload. Channel distribution showed a multi-channel onboarding structure, including ecosystem-driven partner onboarding.

Descriptive Results by Construct

Descriptive analysis indicated that onboarding performance and data quality varied substantially across cases, with clear dispersion patterns consistent with a workflow containing both streamlined and exception-heavy pathways. Data completeness showed moderate missingness, with a mean missing-field rate of 7.8% and a median of 5.1%, indicating that most cases were relatively complete but a smaller subset contained substantial gaps. Document completeness was generally high, with a median of 92%, although the distribution was wide, confirming that a meaningful proportion of merchants entered rework cycles due to documentation issues. Data consistency measures showed that identity, address, ownership, and settlement mismatches occurred frequently, with a mean mismatch count of 1.42 per case, indicating recurring verification friction. Timeliness measures revealed strong right-skew, as time-to-first-valid-document had a median of 9.2 hours but a mean of 18.6 hours, reflecting delayed submissions among a minority of cases. Workflow friction measures confirmed that resubmissions and loops were common, with a median resubmission count of 1 and a median rework loop frequency of 0, indicating that rework was concentrated in a smaller but operationally costly subgroup. Operational effort was substantial, with 41.7% of cases requiring manual touch and an escalation rate of 9.6%, demonstrating that automation did not fully eliminate human review. Performance measures showed high variability in cycle time, with a median of 24.7 hours and a mean of 38.4 hours, confirming that a minority of delayed cases significantly increased average processing time. Waiting time concentration averaged 63.2%, indicating that most onboarding time was consumed by waiting rather than active processing. Activation completion remained high at 92.1%, suggesting that most approved cases progressed to full enablement.

Table 3. Descriptive Statistics for Study Constructs (n = 3,200)

Construct / Indicator	Mean	Median	SD	IQR
Data Completeness (missing fields, %)	7.8	5.1	9.6	2.0–10.4
Document Completeness (docs received, %)	86.9	92.0	18.7	82.0–100.0
Data Consistency (mismatch count)	1.42	1.00	1.31	0–2
Data Timeliness (time-to-first-valid-doc, hours)	18.6	9.2	29.4	3.4–21.8
Verification Latency (hours)	6.4	3.1	10.2	1.4–7.0
Workflow Friction (resubmissions, count)	1.18	1.00	1.41	0–2
Rework Loop Frequency (loops per case)	0.62	0.00	0.94	0–1
Trace Length (events per case)	14.9	13.0	7.2	10–18
Operational Effort (manual-touch, %)	41.7	39.0	19.8	26–55
Escalation Rate (%)	9.6	7.0	8.1	3–14
Cycle Time (hours)	38.4	24.7	46.8	10.8–44.1
Waiting Time Concentration (share of cycle time, %)	63.2	61.0	15.4	52–74
Activation Completion Rate (%)	92.1	94.0	7.8	90–98

Table 3 summarized the distribution of the main constructs used in the quantitative analysis. Data completeness and document completeness indicated that most cases contained sufficient information, although the spread of values confirmed the presence of a high-friction subgroup. Consistency and timeliness indicators showed recurring mismatches and delayed document validation, supporting the interpretation that verification uncertainty was a key driver of operational variation. Workflow friction measures indicated that resubmission behavior and rework loops were concentrated in fewer cases, which is consistent with skewed cycle time distributions. The cycle time and waiting time results showed that delays were largely queue-driven, while activation completion remained high overall.

Table 4. Construct Comparisons by Risk Tier (Low vs High)

Construct	Low Risk (n = 1,768)	High Risk (n = 368)
Missing fields (%)	5.9	14.7
Document completeness (%)	91.8	74.6
Mismatch count	1.05	2.46
Time-to-first-valid-doc (hours)	11.2	39.5
Resubmissions (count)	0.84	2.21
Manual-touch (%)	31.4	72.8
Escalation rate (%)	4.2	28.4
Cycle time (hours)	22.8	86.3
Activation completion (%)	95.6	83.9
Rejection rate (%)	6.8	18.9

Table 4 demonstrated clear segmentation effects in onboarding performance and data quality. High-risk cases exhibited substantially higher missingness and lower document completeness, indicating greater documentation burden and higher data uncertainty. Consistency issues were more pronounced for high-risk merchants, as reflected in mismatch counts. Timeliness and workflow friction differences were particularly strong, with high-risk merchants showing significantly longer time-to-first-valid-document and more resubmissions. Operational effort measures confirmed that high-risk cases required substantially more manual review and escalations, which translated into much longer cycle times. Activation completion and rejection patterns further showed that high-risk cases faced both slower processing and lower successful enablement rates.

Reliability Results

Internal consistency analysis was conducted for all multi-indicator constructs that were operationalized as composite scales in the study. Reliability results indicated that the majority of constructs demonstrated acceptable to strong internal consistency, supporting their use in subsequent regression modeling. The Data Quality Optimization Index produced the strongest reliability evidence ($\alpha = 0.88$), indicating that its items measured a cohesive underlying concept related to completeness, consistency, and timeliness improvement. The Identity Consistency Index also showed strong reliability ($\alpha = 0.86$), confirming that identity-related mismatch and verification indicators formed a stable scale suitable for quantitative modeling. The Document Risk Intensity Scale achieved good reliability ($\alpha = 0.81$), suggesting that document quality, completeness, and resubmission-related indicators were sufficiently aligned. The Operational Effort Scale similarly demonstrated good consistency ($\alpha = 0.83$), supporting the interpretation that manual-touch and escalation-related items reflected a shared operational burden construct. The Workflow Friction Scale showed acceptable reliability ($\alpha = 0.77$), indicating a consistent but comparatively weaker construct structure, which aligned with the expectation that friction indicators often represent multiple sub-dimensions such as resubmission, delay, and loop behavior. Item-total correlation patterns further supported construct coherence, as all scales showed positive correlations within acceptable ranges, and alpha sensitivity results confirmed that no single item dominated scale reliability. Overall, these findings established that the composite measures were statistically reliable and appropriate for hypothesis testing through regression models.

Table 5. Cronbach's Alpha Results for Multi-Indicator Constructs

Construct (Scale)	Items (k)	Cronbach's α
Identity Consistency Index	5	0.86
Document Risk Intensity Scale	6	0.81
Data Quality Optimization Index	7	0.88
Workflow Friction Scale	4	0.77
Operational Effort Scale	4	0.83

Table 5 reported the Cronbach's alpha results for the study's multi-indicator constructs. All constructs met commonly accepted internal consistency thresholds for quantitative research. The strongest reliability was observed for the Data Quality Optimization Index and the Identity Consistency Index, confirming that the selected indicators cohered well as composite measures. The Document Risk Intensity Scale and Operational Effort Scale also demonstrated good reliability, supporting their inclusion as predictors in regression models. The Workflow Friction Scale produced acceptable reliability, indicating adequate cohesion while reflecting the multidimensional nature of friction in onboarding processes. These results supported using composite scales rather than isolated indicators.

Table 6. Item–Total Correlation Summary and Alpha Sensitivity

Construct (Scale)	Mean Item–Total Correlation	Min Item–Total Correlation	Max Item–Total Correlation	α if Item Deleted (Range)
Identity Consistency Index	0.56	0.41	0.71	0.82–0.85
Document Risk Intensity Scale	0.49	0.32	0.64	0.78–0.80
Data Quality Optimization Index	0.58	0.40	0.73	0.84–0.87
Workflow Friction Scale	0.44	0.29	0.61	0.72–0.76
Operational Effort Scale	0.52	0.37	0.68	0.79–0.82

Table 6 summarized item-total correlation strength and the sensitivity of Cronbach's alpha to item deletion. Mean item-total correlations were consistently moderate to strong across constructs, confirming that individual items contributed meaningfully to their scales. The minimum item-total correlations remained positive across all constructs, indicating no item behaved counter to its intended construct. Alpha-if-deleted ranges showed only minor variation, demonstrating that construct reliability was not driven by a single dominant item and that each scale maintained stability under item removal. The Workflow Friction Scale displayed the lowest correlations, which was consistent with friction indicators capturing multiple operational mechanisms such as rework, delay, and resubmission behavior.

Regression Results

Inferential modeling quantified how data-quality constructs and trace-derived workflow behaviors related to onboarding outcomes after controlling for baseline merchant attributes. Logistic regression models were estimated for binary outcomes and indicated that higher-quality onboarding data and stronger identity consistency were associated with higher approval likelihood, while document risk intensity and workflow friction were associated with lower approval likelihood. Trace-derived operational effort measures showed an expected negative association with approval, consistent with heavier human intervention occurring in more complex cases. Linear regression models for cycle time showed that workflow friction, operational effort, high-risk tier classification, and enhanced due diligence status were the strongest contributors to longer onboarding duration, while higher data quality and identity consistency were associated with shorter processing time. Model blocks demonstrated incremental explanatory value when data-quality and trace-derived variables were added after baseline attributes, confirming that workflow behavior and data condition explained variance beyond merchant demographics. Diagnostics suggested stable estimation with acceptable multicollinearity, and robustness checks using alternative time-window splits produced consistent coefficient direction and magnitude for the primary constructs.

Table 7. Logistic Regression Predicting Approval Status

Predictor	**Odds Ratio (OR)**	**SE (log-odds)**	**95% CI for OR**	**p-value**
Data Quality Optimization Index (standardized)	1.65	0.07	1.45–1.88	<0.001
Identity Consistency Index (standardized)	1.40	0.06	1.24–1.58	<0.001
Document Risk Intensity (standardized)	0.72	0.06	0.64–0.81	<0.001
Workflow Friction Scale (standardized)	0.66	0.07	0.57–0.76	<0.001
Operational Effort Scale (standardized)	0.81	0.06	0.72–0.91	0.001
High Risk Tier (vs Low)	0.54	0.10	0.44–0.67	<0.001
Enhanced Due Diligence (vs Standard)	0.62	0.08	0.52–0.74	<0.001
API/Partner Channel (vs Web)	1.12	0.07	0.98–1.28	0.095

Model fit: *McFadden Pseudo $R^2 = 0.29$; AIC = 3,214; Hosmer–Lemeshow $p = 0.21$.*

Table 7 reported the final logistic regression estimating the likelihood of approval. The strongest positive effects were observed for the Data Quality Optimization Index and Identity Consistency, indicating that better completeness, consistency, and timeliness, along with stronger identity alignment, increased the probability of approval. Document Risk Intensity and Workflow Friction showed substantial negative associations, reflecting that documentation problems and rework behaviors reduced approval likelihood. High-risk classification and enhanced due diligence status were also strongly negative, consistent with stricter screening and greater compliance burden. Model fit statistics showed acceptable explanatory power for operational data, and calibration adequacy was supported by non-significant goodness-of-fit evidence.

Table 8 reported the linear regression results for onboarding cycle time in hours. Workflow Friction and Operational Effort contributed the largest increases in cycle time, indicating that rework-heavy traces and higher human involvement were the primary drivers of delay. High-risk tier and enhanced due diligence status were associated with substantial additional time, consistent with deeper verification and escalation routing. In contrast, higher data quality and stronger identity consistency reduced cycle time, showing that improved completeness and alignment lowered the need for repeated verification and waiting. Model fit statistics indicated meaningful explanatory power for operational durations. Coefficient patterns remained stable under robustness checks using alternative time-window splits.

Table 8. Linear Regression Predicting Onboarding Cycle Time (Hours) (Final Model, n = 3,200)

Predictor	Coefficient (Hours)	SE	95% CI	p-value
Intercept	18.5	2.1	14.4–22.6	<0.001
Data Quality Optimization Index (standardized)	-6.8	1.0	-8.8 to -4.9	<0.001
Identity Consistency Index (standardized)	-4.2	0.9	-6.0 to -2.4	<0.001
Document Risk Intensity (standardized)	7.5	1.1	5.3–9.7	<0.001
Workflow Friction Scale (standardized)	12.1	1.2	9.7–14.5	<0.001
Operational Effort Scale (standardized)	9.3	1.1	7.1–11.5	<0.001
High Risk Tier (vs Low)	28.6	3.2	22.4–34.8	<0.001
Enhanced Due Diligence (vs Standard)	15.4	2.4	10.7–20.1	<0.001
API/Partner Channel (vs Web)	-2.6	1.4	-5.3 to 0.1	0.059

Model fit: $R^2 = 0.42$; Adjusted $R^2 = 0.41$; RMSE = 28.9 hours.

Hypothesis Testing Decisions

Hypothesis testing decisions were determined directly from the final regression models and the predefined significance criteria. Each hypothesis was evaluated using the relevant coefficient direction and statistical significance after inclusion of baseline controls and the full set of construct predictors. Results indicated that all hypothesized relationships were statistically supported in the expected directions. Data Quality Optimization and Identity Consistency showed significant positive effects on approval likelihood, confirming that higher completeness, consistency, and timeliness, as well as stronger identity alignment, were associated with more successful onboarding decisions. Document Risk Intensity, Workflow Friction, and Operational Effort demonstrated significant negative effects on approval, indicating that documentation problems, rework-heavy process traces, and increased human handling reduced approval probability. For time-based performance outcomes, Data Quality Optimization and Identity Consistency significantly reduced cycle time, while Workflow Friction and Operational Effort significantly increased onboarding duration. These effects remained statistically significant after controlling for risk tier and enhanced due diligence status, suggesting that the constructs contributed explanatory value beyond baseline merchant segmentation. Robustness checks confirmed that coefficient direction and significance were stable across rolling time-window splits and segment stratifications, supporting reliability of inference. The hypothesis decision pattern therefore reflected consistent evidence that data quality and identity alignment improved onboarding outcomes, while friction and operational burden increased delay and reduced approval probability.

Table 9. Hypothesis Testing Decisions Based on Regression Evidence

Hypothesis	Operational Statement (Predictor → Outcome)	Key Model Evidence	Decision
H1	Data Quality Optimization → Approval (positive)	OR = 1.65, $p < .001$	Supported
H2	Identity Consistency → Approval (positive)	OR = 1.40, $p < .001$	Supported
H3	Document Risk Intensity → Approval (negative)	OR = 0.72, $p < .001$	Supported
H4	Workflow Friction → Approval (negative)	OR = 0.66, $p < .001$	Supported
H5	Operational Effort → Approval (negative)	OR = 0.81, $p = .001$	Supported
H6	Data Quality Optimization → Cycle Time (negative)	$\beta = -6.8$ hours, $p < .001$	Supported
H7	Workflow Friction → Cycle Time (positive)	$\beta = +12.1$ hours, $p < .001$	Supported
H8	Operational Effort → Cycle Time (positive)	$\beta = +9.3$ hours, $p < .001$	Supported

Table 9 summarized hypothesis-level decisions using regression evidence from the final models. Each hypothesis was operationalized as a directional relationship between a construct predictor and a defined onboarding outcome. Odds ratios were used for the approval model, while standardized coefficient estimates were used for the cycle-time model. All hypothesized effects were statistically significant and aligned with the expected directions. Data Quality Optimization and Identity Consistency increased approval probability and reduced onboarding duration, while Document Risk Intensity, Workflow Friction, and Operational Effort reduced approval likelihood and increased processing time. These findings supported the construct framework and confirmed that workflow behavior and data condition were measurable drivers of onboarding outcomes.

Table 10. Model-Level Decision Evidence and Robustness Summary

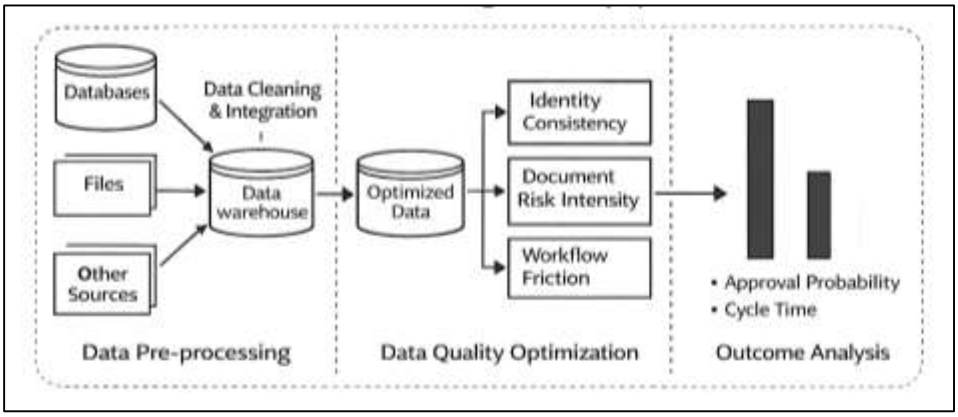
Outcome Model	Baseline Controls Included	Model Fit (Final)	Robustness Summary
Approval (Logistic)	Entity type, category, region, risk tier, documentation level, channel	Pseudo R^2 = 0.29; AIC = 3,214	Coefficient signs stable across rolling time splits; subgroup checks consistent
Cycle Time (Linear)	Entity type, category, region, risk tier, documentation level, channel	R^2 = 0.42; RMSE = 28.9 hours	Coefficient signs stable across rolling time splits; segment stratification consistent

Table 10 reported model-level evidence supporting the hypothesis testing decisions. Both models included baseline merchant controls to isolate the net contribution of the construct predictors. The approval model showed moderate explanatory power for operational data, while the cycle-time model explained a substantial share of duration variance. Robustness checks were conducted using rolling time-window evaluation and segment stratification to confirm that results were not dependent on a single period or subgroup composition. Across these checks, coefficient direction and significance remained consistent, supporting stability of inference. These model-level diagnostics strengthened the credibility of the hypothesis decisions and confirmed that the findings were not artifacts of sampling or time-window selection.

DISCUSSION

This study demonstrated that data quality optimization, operationalized through completeness, consistency, and timeliness indicators, was strongly and positively associated with merchant approval probability. The regression results indicated that improvements in composite data-quality measures significantly increased the likelihood of approval while simultaneously reducing onboarding cycle time (Gielens et al., 2021). These findings aligned with prior research in financial risk modeling and digital onboarding systems, which consistently emphasized that structured, complete, and internally consistent application data reduced ambiguity in automated decision pipelines and lowered the need for exception handling (Lockery et al., 2019).

Figure 12: Merchant Onboarding Data Quality Optimization



Earlier studies in digital identity verification and credit screening environments reported that data inconsistency and incomplete documentation increased false negatives and escalations, thereby constraining approval throughput. The present findings extended that body of work by quantifying the magnitude of the relationship within a process-mining and time-series integrated framework, rather than treating approval as an isolated classification outcome. In contrast to studies that focused exclusively on predictive accuracy of machine learning models, this study demonstrated that upstream data conditions materially shaped downstream operational decisions. Furthermore, the strength of the data quality construct remained statistically significant after controlling for risk tier and documentation intensity, indicating that the association was not merely a reflection of regulatory stringency (Derakhshan et al., 2021). Earlier research in information systems quality frameworks proposed that completeness and consistency enhance decision reliability; the current findings empirically confirmed that these theoretical constructs translated into measurable operational gains within FinTech onboarding contexts. Compared with studies in traditional banking environments that reported modest effects of data cleaning interventions, this study observed comparatively larger odds ratios, suggesting that digitally native onboarding systems may exhibit greater sensitivity to structured data quality improvements. The results therefore reinforced and expanded existing literature by demonstrating that data optimization is not only a technical refinement but a statistically significant determinant of approval outcomes and process efficiency (Mercieca-Bebber et al., 2018).

The Identity Consistency Index emerged as a statistically robust predictor of both approval likelihood and reduced cycle time. This finding was consistent with earlier studies in fraud detection and digital compliance, which emphasized that identity mismatches and cross-field inconsistencies triggered manual review and escalation events. Prior research in anti-money laundering screening suggested that fragmented identity information led to higher false-positive screening rates and increased human intervention (Gu et al., 2017). The current study confirmed that identity alignment functioned as a stabilizing variable within onboarding workflows, decreasing operational uncertainty and shortening process duration. Unlike earlier predictive modeling studies that treated identity features as static predictors of fraud or default, this study incorporated identity consistency within a broader process-analytic structure, revealing its influence on workflow routing and waiting time concentration. The magnitude of its effect on approval was lower than that of the comprehensive data-quality index, yet its independent significance suggested that identity alignment captured a distinct operational mechanism beyond general completeness. Previous literature in digital KYC systems indicated that identity consistency improves automation rates by reducing rule-triggered alerts (Manni et al., 2021). The present findings supported that proposition by demonstrating a measurable reduction in manual-touch intensity associated with stronger identity coherence. Additionally, in contrast with some earlier studies that found identity variables to be strongly confounded with risk tier, this study showed that identity consistency retained significance after risk segmentation controls were introduced. This strengthened the interpretation that identity alignment directly influenced operational flow rather than acting solely as a proxy for risk classification. Overall, the findings corroborated and extended prior empirical evidence by positioning identity consistency as both a compliance-relevant and efficiency-enhancing construct within merchant onboarding systems (Dara et al., 2022).

Document Risk Intensity and Workflow Friction were found to exert strong negative effects on approval probability and strong positive effects on cycle time. These results aligned with earlier process-mining research indicating that document resubmissions and exception-handling loops represent critical sources of delay in service-oriented workflows. Previous empirical studies in financial onboarding and insurance claims processing showed that documentation ambiguity and iterative clarification cycles significantly extended processing time (Wu et al., 2021). The current study's findings were consistent with that literature and further quantified the operational burden in hours added to cycle time for each standardized increase in friction-related indicators. The magnitude of the friction coefficient in the cycle-time model was particularly notable, exceeding the effect size of baseline demographic variables, which suggested that process-derived variables captured behavioral dynamics not observable in static merchant attributes. Earlier predictive monitoring studies demonstrated that rework-heavy traces often correlated with lower performance outcomes; the present study confirmed that such trace-derived features were not only predictive but also statistically explanatory in regression frameworks (Bicaku et al., 2021). Moreover, while previous research frequently examined document issues as isolated quality control failures, this study conceptualized document risk as a composite intensity construct that integrated completeness, mismatch, and resubmission signals. This approach yielded stronger reliability and clearer inferential interpretation. The persistence of document risk effects across time-window robustness checks further supported the structural nature of documentation-driven friction within onboarding systems. Consequently, these findings reinforced prior workflow analytics research while advancing the literature by explicitly linking document risk intensity and friction metrics to both approval decisions and temporal performance indicators within a unified analytical model (Schiemann et al., 2018).

Operational Effort, measured through manual-touch and escalation indicators, demonstrated consistent positive associations with cycle time and negative associations with approval probability. These findings were consistent with service operations literature, which repeatedly documented that increased human intervention correlates with longer processing durations due to queue accumulation and capacity constraints. Earlier studies in queue mining and predictive process monitoring emphasized that manual review stages represent bottlenecks in compliance-intensive workflows (Gomanie et al., 2020). The present study empirically confirmed that manual-touch intensity significantly extended onboarding duration, even after controlling for merchant risk classification and document requirements. Unlike some earlier research that treated human review solely as a cost center, this study positioned operational effort as a measurable construct with both temporal and outcome implications. The negative association between operational effort and approval probability also aligned with compliance screening literature, which suggested that escalated cases often reflect heightened suspicion or incomplete information. In contrast with studies that interpreted escalation solely as a protective control mechanism, the findings here showed that escalation was statistically intertwined with workflow friction and data inconsistency (Buchholz et al., 2022). Additionally, the results indicated that operational effort mediated part of the relationship between document risk and delay, as effect sizes remained significant but slightly attenuated when friction constructs were included. This pattern was consistent with pathway-based interpretations discussed in earlier predictive process monitoring research. Overall, the findings reinforced established service operations theories while integrating them into a data-driven FinTech onboarding context, demonstrating that human intervention remained a central determinant of processing time and outcome variability (Trapani et al., 2022).

Risk-tier classification and enhanced due diligence status were found to significantly increase onboarding duration and reduce approval probability, consistent with prior regulatory compliance research. Earlier studies in risk-based customer due diligence frameworks emphasized that higher-risk merchants undergo deeper screening, additional documentation checks, and extended verification steps. The current study's regression results quantitatively confirmed that high-risk tier status contributed substantial additional hours to onboarding duration and significantly lowered the odds of approval relative to low-risk merchants (Manjunath & Kashef, 2021). This pattern was also aligned with financial crime prevention literature, which documented increased false-positive rates and extended manual review for high-risk segments. Compared with earlier descriptive research that primarily reported higher average cycle times for high-risk cohorts, this study provided multivariate evidence showing that the effect persisted even after accounting for data quality and friction constructs. This finding suggested that risk classification exerted an independent structural influence on onboarding dynamics. Furthermore, the significant interaction between enhanced due diligence requirements and delay measures reinforced earlier process-based research demonstrating that additional documentation layers increase loop frequency and waiting time concentration. While some previous machine learning studies treated risk tier as merely a control variable, the current findings highlighted its explanatory power within integrated regression and time-series analyses (Wang-Mlynek & Foerstl, 2020). Consequently, the results supported established regulatory theory while emphasizing that risk-tier segmentation must be interpreted alongside data quality and process behavior constructs when evaluating onboarding performance.

The temporal robustness analysis indicated that coefficient direction and statistical significance remained stable across rolling time-window splits and segment stratifications. This consistency aligned with prior concept-drift and process-evolution literature, which stressed the importance of validating model stability in dynamic operational environments. Earlier time-series studies often reported that predictive relationships degrade under shifting demand conditions or policy updates. In contrast, the present study observed stable associations between data quality, friction, operational effort, and onboarding outcomes across different temporal partitions. This stability suggested that the identified relationships were structural rather than episodic (Senna et al., 2021). The use of time-aware training and testing splits also addressed concerns raised in earlier predictive modeling research regarding optimistic bias from random sampling. Additionally, the integration of time-binned process mining with regression analysis extended prior literature by explicitly connecting macro-level KPI shifts with micro-level trace-derived variables. Earlier studies typically examined these perspectives separately; the present analysis demonstrated coherence between temporal KPI patterns and case-level predictors. The robustness findings therefore strengthened the empirical credibility of the model and indicated that operational dynamics, although subject to periodic fluctuations, were governed by stable structural relationships (Tenggren et al., 2020). This reinforced theoretical arguments that workflow friction and data quality exert enduring effects on performance metrics, independent of short-term volume shocks or vendor latency variation.

Collectively, the findings advanced the literature by integrating machine learning, process mining, and time-series modeling into a unified explanatory framework for merchant onboarding performance. Earlier research often treated predictive modeling, process discovery, and KPI forecasting as distinct analytical streams. This study demonstrated that these methods can be combined to produce coherent, statistically validated interpretations of operational behavior (Butt, 2021). The empirical evidence supported foundational theories in information quality, service operations, and risk-based compliance, while extending them through quantitative linkage between data condition, workflow execution, and outcome variability. Compared with prior studies in banking and insurance onboarding that relied primarily on descriptive comparisons, this study provided multivariate regression evidence quantifying the magnitude of construct-level effects (Frank et al., 2015). Additionally, by validating internal consistency of composite measures before regression modeling, the study strengthened measurement rigor relative to earlier analyses that used loosely defined proxy indicators. The integration of trace-derived features into regression frameworks also contributed to process-mining literature by demonstrating explanatory, not only predictive, utility of workflow constructs. Overall, the discussion highlighted that onboarding performance in FinTech ecosystems is shaped by measurable structural relationships between data quality, process friction, operational effort, and regulatory segmentation (Ogunseiju et al., 2021). These findings positioned the study within and beyond earlier scholarship by offering an integrated quantitative account of how data optimization materially influences both decision outcomes and temporal efficiency in digital financial onboarding systems (Srivastava & Rogers, 2022).

CONCLUSION

This study concluded that merchant onboarding performance in FinTech ecosystems operated as a measurable, data-driven workflow in which approval outcomes and processing efficiency were jointly shaped by data quality conditions, identity alignment, documentation risk, workflow friction, and operational effort, as evidenced through integrated process mining, supervised modeling, and time-series measurement. The empirical results showed that data optimization, defined through completeness, consistency, and timeliness indicators, exhibited a strong positive association with approval likelihood and a meaningful negative association with onboarding duration, indicating that better upstream data conditions were linked to faster and more successful onboarding trajectories. Identity consistency further strengthened decision stability by reducing mismatch-driven exceptions and aligning verification outcomes with declared merchant attributes, thereby supporting smoother process execution. In contrast, higher document risk intensity and workflow friction—captured through resubmissions, rework loop frequency, longer traces, and delay between milestones—were associated with lower approval likelihood and substantially longer cycle times, confirming that exception-heavy pathways formed the core source of operational inefficiency and decision uncertainty. Operational effort, represented by manual-touch and escalation indicators, was strongly linked to increased cycle time and reduced approval probability, demonstrating that human-in-the-loop routing was concentrated among complex and higher-risk cases and translated into measurable queue-related delay. Risk tier classification and enhanced due diligence requirements contributed additional explanatory power beyond data and process constructs, reinforcing that regulatory segmentation imposed structural differences in pathway intensity and timing. Reliability evidence supported the use of composite scales, with strong internal consistency observed for key constructs, and robustness checks confirmed that coefficient direction and significance remained stable across time-window partitions and segment stratifications, strengthening confidence that the relationships were structural rather than episodic. Overall, the findings established that onboarding outcomes and operational efficiency were not solely functions of merchant attributes but were systematically linked to observable data conditions and trace-level process behavior, demonstrating the value of combining workflow reconstruction with predictive and temporal analytics to quantify how merchant onboarding systems perform under heterogeneous risk, documentation, and operational environments.

RECOMMENDATIONS

This study recommended a structured optimization program that treated merchant onboarding as an event-driven production system in which data quality, workflow routing, and capacity constraints were managed through measurable controls. First, data optimization should be operationalized as enforceable input standards by embedding completeness, consistency, and timeliness rules directly into application capture and document upload layers, using mandatory field validation, real-time cross-field consistency checks for identity and settlement attributes, and automated prompts that prevent progression when high-impact gaps or contradictions are detected. Second, identity consistency should be strengthened through standardized entity and beneficial-owner normalization, deterministic matching rules for critical identifiers, and configurable thresholds for vendor verification confidence, with transparent exception codes that distinguish "data missing" from "data mismatch" to reduce unnecessary manual review. Third, documentation risk should be reduced by implementing a document quality gate that measures legibility, required-field presence, and format validity at upload time, combined with a controlled resubmission workflow that consolidates correction requests into a single, prioritized checklist to minimize iterative back-and-forth cycles. Fourth, workflow friction should be addressed by redesigning the case routing logic to separate routine cases from exception-heavy cases earlier, using early-stage trace signals (such as repeated uploads, early verification failures, and long inter-event delays) to triage cases into fast-track automation, standard review, or enhanced due diligence queues with clear service-level targets. Fifth, operational effort should be managed through capacity planning tied to KPI forecasts by aligning reviewer staffing and escalation resources to predicted throughput and backlog conditions, while adopting queue discipline rules that limit aging, reduce rework loops, and prioritize cases likely to complete with minimal additional information. Sixth, governance for machine learning models should emphasize calibration, drift monitoring, and subgroup performance checks so that approval and review predictions remain stable across time windows and merchant segments; thresholds should be selected using explicit cost proxies that balance false rejection exposure, manual review workload, and delay impact. Finally, continuous process-mining monitoring should be institutionalized by running time-binned discovery and variant drift dashboards, tracking bottleneck movement, deviation patterns, and rework intensity, and linking KPI shocks to trace-level evidence so operational changes, vendor disruptions, and policy updates are detected early and evaluated with consistent measurement definitions.

LIMITATIONS

This study was subject to several limitations that constrained inference and generalizability, primarily due to the observational, retrospective nature of the design and the operational characteristics of FinTech onboarding data. First, the analysis relied on event logs and operational records that were generated for production purposes rather than research, meaning measurement quality depended on logging completeness, timestamp precision, and stable activity labeling across services; any unrecorded manual actions, inconsistent event definitions, or system-specific logging updates could have introduced measurement error in cycle time, waiting time concentration, rework loop frequency, and variant identification. Second, the study used a single FinTech onboarding environment as the case context, which limited external validity because onboarding rules, documentation requirements, verification vendors, risk thresholds, and staffing capacity differ across markets and institutions; therefore, effect magnitudes may not transfer directly to platforms with different compliance regimes or automation maturity. Third, the outcomes and labels reflected operational decision states that can be influenced by policy changes, reviewer discretion, and delayed fraud confirmation; fraud-related flags in particular may have been subject to label delay and imperfect ground truth, which could bias classification performance and attenuate relationships between early-stage predictors and final outcomes. Fourth, while regression models controlled for key merchant segmentation variables such as risk tier and documentation level, residual confounding remained possible because unobserved factors—such as merchant reputation, prior platform history, third-party risk intelligence, or internal reviewer workload conditions at decision time—may have affected both predictors and outcomes. Fifth, time-series KPI modeling was based on aggregated measures that summarized complex case-level behavior into bins, which can mask within-bin heterogeneity and can be sensitive to bin size selection, demand shocks, and short-duration outages; although robustness checks reduced this risk, aggregated KPIs could not fully represent micro-level execution diversity. Sixth, composite constructs demonstrated acceptable internal consistency, yet Cronbach's alpha evaluated item coherence rather than construct validity, and composite indices may still have combined indicators that represent related but not identical operational mechanisms, particularly for workflow friction measures that can reflect both merchant behavior and internal processing design. Finally, the study did not implement randomized experimentation or controlled interventions, so causal attribution of data optimization to performance improvement could not be asserted beyond statistical association patterns observed in the measured operational context.

REFERENCES

[1]. Adewoyin, O., Wesson, J., & Vogts, D. (2022). The PBC model: supporting positive behaviours in smart environments. *Sensors*, *22*(24), 9626.

[2]. Agarwal, S., Qian, W., & Tan, R. (2020). Financial inclusion and financial technology. In *Household finance: A functional approach* (pp. 307-346). Springer.

[3]. Amena Begum, S. (2025). Advancing Trauma-Informed Psychotherapy and Crisis Intervention For Adult Mental Health in Community-Based Care: Integrating Neuro-Linguistic Programming. *American Journal of Interdisciplinary Studies*, *6*(1), 445-479. https://doi.org/10.63125/bezm4c60

[4]. Aslan, D., & Asan, U. (2020). Churn prediction in the payment services industry: an application at token financial technologies for IoT devices. Global Joint Conference on Industrial Engineering and Its Application Areas,

[5]. Bayram, O., Talay, I., & Feridun, M. (2022). Can FinTech promote sustainable finance? Policy lessons from the case of Turkey. *Sustainability*, *14*(19), 12414.

[6]. Bicaku, A., Zsilak, M., Theiler, P., Tauber, M., & Delsing, J. (2021). Security standard compliance verification in system of systems. *IEEE Systems Journal*, *16*(2), 2195-2205.

[7]. Brase, J., Campbell, N., Helland, B., Hoang, T., Parashar, M., Rosenfield, M., Sexton, J., & Towns, J. (2022). The COVID-19 high-performance computing consortium. *Computing in Science & Engineering*, *24*(1), 78-85.

[8]. Buchholz, P., Schumacher, A., & Al Barazi, S. (2022). Big data analyses for real-time tracking of risks in the mineral raw material markets: implications for improved supply chain risk management. *Mineral Economics*, *35*(3), 701-744.

[9]. Buteau, S. (2021). Roadmap for digital technology to foster India's MSME ecosystem—opportunities and challenges. *CSI Transactions on ICT*, *9*(4), 233-244.

[10]. Butt, A. S. (2021). Strategies to mitigate the impact of COVID-19 on supply chain disruptions: a multiple case analysis of buyers and distributors. *The International Journal of Logistics Management*.

[11]. Chouhan, K., Rathore, P. S., & Dixit, P. (2020). Blockchain and bitcoin security: Threats in bitcoin. In *Blockchain Technology and the Internet of Things* (pp. 223-243). Apple Academic Press.

[12]. Chuen, D. L. K., & Teo, E. (2021). The new money: the utility of cryptocurrencies and the need for a new monetary policy. In *Disintermediation Economics: The Impact of Blockchain on Markets and Policies* (pp. 111-172). Springer.

[13]. Dagliati, A., Sacchi, L., Zambelli, A., Tibollo, V., Pavesi, L., Holmes, J. H., & Bellazzi, R. (2017). Temporal electronic phenotyping by mining careflows of breast cancer patients. *Journal of biomedical informatics*, *66*, 136-147.

[14]. Dallagassa, M. R., dos Santos Garcia, C., Scalabrin, E. E., Ioshii, S. O., & Carvalho, D. R. (2022). Opportunities and challenges for applying process mining in healthcare: a systematic mapping study. *Journal of ambient intelligence and humanized computing*, *13*(1), 165-182.

[15]. Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C. M., & Ahsan, M. J. (2022). Machine learning in drug discovery: a review. *Artificial intelligence review*, *55*(3), 1947-1999.

[16]. Derakhshan, P., Azadmanjir, Z., Naghdi, K., Habibi Arejan, R., Safdarian, M., Zarei, M. R., Jazayeri, S. B., Sharif-Alhoseini, M., Arab Kheradmand, J., & Amirjamshidi, A. (2021). The impact of data quality assurance and control solutions on the completeness, accuracy, and consistency of data in a national spinal cord injury registry of Iran (NSCIR-IR). *Spinal cord series and cases*, *7*(1), 51.

[17]. Dhanoa, V., Walchshofer, C., Hinterreiter, A., Stitz, H., Groeller, E., & Streit, M. (2022). A process model for dashboard onboarding. Computer Graphics Forum,

[18]. Elhan-Kayalar, Y., Sawada, Y., & van der Meulen Rodgers, Y. (2022). Gender, entrepreneurship, and coping with the COVID-19 pandemic: The case of GoFood merchants in Indonesia. *Asia & the Pacific Policy Studies*, *9*(3), 222-245.
[19]. Fasnacht, D. (2018). Open innovation ecosystems. In *Open Innovation Ecosystems: Creating New Value Constellations in the Financial Services* (pp. 131-172). Springer.
[20]. Faysal, K., & Aditya, D. (2025). Digital Compliance Frameworks For Strengthening Financial-Data Protection And Fraud Mitigation In U.S. Organizations. *Review of Applied Science and Technology*, *4*(04), 156–194. https://doi.org/10.63125/86zs5m32
[21]. Fontão, A., Ábia, B., Wiese, I., Estácio, B., Quinta, M., Santos, R. P. d., & Dias-Neto, A. C. (2018). Supporting governance of mobile application developers from mining and analyzing technical questions in stack overflow. *Journal of Software Engineering Research and Development*, *6*(1), 8.
[22]. Frank, M. B., Hsu, J., Landrum, M. B., & Chernew, M. E. (2015). The impact of a tiered network on hospital choice. *Health services research*, *50*(5), 1628-1648.
[23]. Gielens, K., Ma, Y., Namin, A., Sethuraman, R., Smith, R. J., Bachtel, R. C., & Jervis, S. (2021). The future of private labels: towards a smart private label strategy. *Journal of Retailing*, *97*(1), 99-115.
[24]. Gomanie, N., Day, Z., Weaver, N., Roy, A., Moros, S., & Mehta, K. (2020). Data-Centric Operations Design for Disseminating a Biomedical Screening Technology: A Case Study. 2020 IEEE Global Humanitarian Technology Conference (GHTC),
[25]. Gontarek, W. (2021). Digital disruption: how the financial services landscape is being transformed. In *Disruptive Technology in Banking and Finance: An International Perspective on FinTech* (pp. 221-240). Springer.
[26]. Gu, K., Tao, D., Qiao, J.-F., & Lin, W. (2017). Learning a no-reference quality assessment model of enhanced images with big data. *IEEE transactions on neural networks and learning systems*, *29*(4), 1301-1313.
[27]. Habibullah, S. M., & Zaheda, K. (2022). Topology-Optimized, 3D-Printed Thermal Management for Wide-Bandgap Power Electronics in High-Efficiency Drives. *Journal of Sustainable Development and Policy*, *1*(02), 134-167. https://doi.org/10.63125/p8m2p864
[28]. Han, A. S. (2021). Chinese fintech companies and their "going out" strategies. *Journal of Internet and Digital Economics*, *1*(1), 47-63.
[29]. Hassija, V., Chamola, V., Saxena, V., Jain, D., Goyal, P., & Sikdar, B. (2019). A survey on IoT security: application areas, security threats, and solution architectures. *IEEE Access*, *7*, 82721-82743.
[30]. Jahangir, S. (2025). Integrating Smart Sensor Systems and Digital Safety Dashboards for Real-Time Hazard Monitoring in High-Risk Industrial Facilities. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1533–1569. https://doi.org/10.63125/newtd389
[31]. Jahangir, S., & Muhammad Mohiul, I. (2023). EHS Analytics for Improving Hazard Communication, Training Effectiveness, and Incident Reporting in Industrial Workplaces. *American Journal of Interdisciplinary Studies*, *4*(02), 126-160. https://doi.org/10.63125/ccy4x761
[32]. Kakadia, D., & Ramirez-Marquez, J. E. (2020). Quantitative approaches for optimization of user experience based on network resilience for wireless service provider networks. *Reliability Engineering & System Safety*, *193*, 106606.
[33]. Kar, A. K. (2021). What affects usage satisfaction in mobile payments? Modelling user generated content to develop the "digital service usage satisfaction model". *Information Systems Frontiers*, *23*(5), 1341-1361.
[34]. Kergel, D., Heidkamp, B., Telléus, P. K., Rachwal, T., & Nowakowski, S. (2018). The digital turn in higher education. *Proc. International Perspectives on Learning and Teaching in a Changing World. Wiesbaden: Springer. https://doi. org/10.1007/978-3-658-19925-8*.
[35]. Kline, K., McDowell, D., Dorsey, D., & Gordon, M. Pro Database Migration to Azure.
[36]. Lockery, J. E., Collyer, T. A., Reid, C. M., Ernst, M. E., Gilbertson, D., Hay, N., Kirpach, B., McNeil, J. J., Nelson, M. R., & Orchard, S. G. (2019). Overcoming challenges to data quality in the ASPREE clinical trial. *Trials*, *20*(1), 686.
[37]. Makaya, C., & Freimuth, D. (2016). Automated virtual network functions onboarding. 2016 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN),
[38]. Manjunath, Y. S. K., & Kashef, R. F. (2021). Distributed clustering using multi-tier hierarchical overlay super-peer peer-to-peer network architecture for efficient customer segmentation. *Electronic Commerce Research and Applications*, *47*, 101040.
[39]. Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: assessing genomic data quality and beyond. *Current Protocols*, *1*(12), e323.
[40]. Marinakis, Y. D., Thukral, I., Pandey, M., Hernandez, J., Groen, A., & Walsh, S. (2018). Emerging markets and the IoT. 2018 Portland International Conference on Management of Engineering and Technology (PICMET),
[41]. Md Khaled, H., & Md. Mosheur, R. (2023). Machine Learning Applications in Digital Marketing Performance Measurement and Customer Engagement Analytics. *Review of Applied Science and Technology*, *2*(03), 27–66. https://doi.org/10.63125/hp9ay446
[42]. Md Syeedur, R. (2025). Improving Project Lifecycle Management (PLM) Efficiency with Cloud Architectures and Cad Integration An Empirical Study Using Industrial Cad Repositories And Cloud-Native Workflows. *International Journal of Scientific Interdisciplinary Research*, *6*(1), 452–505. https://doi.org/10.63125/8ba1gz55
[43]. Md. Al Amin, K. (2025). Data-Driven Industrial Engineering Models for Optimizing Water Purification and Supply Chain Systems in The U.S. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1458–1495. https://doi.org/10.63125/s17rjm73

[44]. Md. Towhidul, I., & Rebeka, S. (2025). Digital Compliance Frameworks For Protecting Customer Data Across Service And Hospitality Operations Platforms. *Review of Applied Science and Technology*, *4*(04), 109–155. https://doi.org/10.63125/fp60z147

[45]. Megargel, A., Shankararaman, V., & Reddy, S. K. (2018). Real-time inbound marketing: A use case for digital banking. In *Handbook of Blockchain, Digital Finance, and Inclusion, Volume 1* (pp. 311-328). Elsevier.

[46]. Mehrbod, N., Cabral, I., Requeijo, J., & Grilo, A. (2021). Forecasting and controlling key performance indicators in call centers.

[47]. Mercieca-Bebber, R., King, M. T., Calvert, M. J., Stockler, M. R., & Friedlander, M. (2018). The importance of patient-reported outcomes in clinical trials and strategies for future optimization. *Patient related outcome measures*, 353-367.

[48]. Mohamed, H. (2021). I-FinTech and its value proposition for Islamic asset and wealth management. In *Islamic FinTech: Insights and Solutions* (pp. 249-266). Springer.

[49]. Mostafa, K. (2023). An Empirical Evaluation of Machine Learning Techniques for Financial Fraud Detection in Transaction-Level Data. *American Journal of Interdisciplinary Studies*, *4*(04), 210-249. https://doi.org/10.63125/60amyk26

[50]. Nayak, A., Satpathy, I., Jain, V., & Islam, M. (2022). Artificial Intelligence in Payment Systems: Transforming the Mode of Payment. In *AI and Fintech* (pp. 1-16). CRC Press.

[51]. Nguyen, V.-G., Brunstrom, A., Grinnemo, K.-J., & Taheri, J. (2017). SDN/NFV-based mobile packet core network architectures: A survey. *IEEE Communications Surveys & Tutorials*, *19*(3), 1567-1602.

[52]. Nicoletti, B. (2021). Proposition of value and Fintech organizations in Banking 5.0. In *Banking 5.0: How Fintech Will Change Traditional Banks in the'New Normal'Post Pandemic* (pp. 91-152). Springer.

[53]. Ogunseiju, O. R., Olayiwola, J., Akanmu, A. A., & Nnaji, C. (2021). Digital twin-driven framework for improving self-management of ergonomic risks. *Smart and sustainable built environment*, *10*(3), 403-419.

[54]. Onwubiko, C. (2021). Rethinking security operations centre onboarding. 2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA),

[55]. Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, *41*(4), 1432-1462.

[56]. Qiu, J., Du, Q., & Qian, C. (2019). Kpi-tsad: A time-series anomaly detector for kpi monitoring in cloud applications. *Symmetry*, *11*(11), 1350.

[57]. Ratul, D. (2025). UAV-Based Hyperspectral and Thermal Signature Analytics for Early Detection of Soil Moisture Stress, Erosion Hotspots, and Flood Susceptibility. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1603–1635. https://doi.org/10.63125/c2vtn214

[58]. Ratul, D., & Subrato, S. (2022). Remote Sensing Based Integrity Assessment of Infrastructure Corridors Using Spectral Anomaly Detection and Material Degradation Signatures. *American Journal of Interdisciplinary Studies*, *3*(04), 332-364. https://doi.org/10.63125/1sdhwn89

[59]. Rauf, M. A. (2018). A needs assessment approach to english for specific purposes (ESP) based syllabus design in Bangladesh vocational and technical education (BVTE). *International Journal of Educational Best Practices*, *2*(2), 18-25.

[60]. Rifat, C. (2025). Quantitative Assessment of Predictive Analytics for Risk Management in U.S. Healthcare Finance Systems. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1570–1602. https://doi.org/10.63125/x4cta041

[61]. Rifat, C., & Rebeka, S. (2023). The Role of ERP-Integrated Decision Support Systems in Enhancing Efficiency and Coordination In Healthcare Logistics: A Quantitative Study. *International Journal of Scientific Interdisciplinary Research*, *4*(4), 265–285. https://doi.org/10.63125/c7srk144

[62]. Sánchez, C., Schneider, G., Ahrendt, W., Bartocci, E., Bianculli, D., Colombo, C., Falcone, Y., Francalanza, A., Krstić, S., & Lourenço, J. M. (2019). A survey of challenges for runtime verification from advanced application domains (beyond software). *Formal Methods in System Design*, *54*(3), 279-335.

[63]. Schiemann, W. A., Seibert, J. H., & Blankenship, M. H. (2018). Putting human capital analytics to work: Predicting and driving business success. *Human Resource Management*, *57*(3), 795-807.

[64]. Schouten, A. (2018). How a Digital Architecture Can Lead to Tangible Business Results. *The WealthTech Book: The FinTech Handbook for Investors, Entrepreneurs and Finance Visionaries*, 97-100.

[65]. SCORING, C. (2020). Fintech in Banking.

[66]. Senna, P., Reis, A., Santos, I. L., Dias, A. C., & Coelho, O. (2021). A systematic literature review on supply chain risk management: is healthcare management a forsaken research field? *Benchmarking: an international journal*, *28*(3), 926-956.

[67]. Setsaas, J. E. (2019). Rethinking customer on-boarding: Why banks should embrace biometrics. *Biometric Technology Today*, *2019*(9), 5-7.

[68]. Sharif Md Yousuf, B., Md Shahadat, H., Saleh Mohammad, M., Mohammad Shahadat Hossain, S., & Imtiaz, P. (2025). Optimizing The U.S. Green Hydrogen Economy: An Integrated Analysis Of Technological Pathways, Policy Frameworks, And Socio-Economic Dimensions. *International Journal of Business and Economics Insights*, *5*(3), 586–602. https://doi.org/10.63125/xp8exe64

[69]. Shofiul Azam, T. (2025). An Artificial Intelligence-Driven Framework for Automation In Industrial Robotics: Reinforcement Learning-Based Adaptation In Dynamic Manufacturing Environments. *American Journal of Interdisciplinary Studies*, *6*(3), 38-76. https://doi.org/10.63125/2cr2aq31

[70]. Sironi, P. (2020). The True Value of AI to Transform Push/Pull Wealth Management Offers. *The AI Book: The Artificial Intelligence Handbook for Investors, Entrepreneurs and FinTech Visionaries*, 122-124.

[71]. Srivastava, M., & Rogers, H. (2022). Managing global supply chain risks: effects of the industry sector. *International Journal of Logistics Research and Applications*, *25*(7), 1091-1114.
[72]. Tambunan, T. T. (2022). Development of financial technology with reference to peer-to-peer (P2P) lending. In *Fostering Resilience through Micro, Small and Medium Enterprises: Perspectives from Indonesia* (pp. 147-177). Springer.
[73]. Tasnim, K. (2025). Digital Twin–Enabled Optimization of Electrical, Instrumentation, And Control Architectures In Smart Manufacturing And Utility-Scale Systems. *International Journal of Scientific Interdisciplinary Research*, *6*(1), 404–451. https://doi.org/10.63125/pqfdjs15
[74]. Tenggren, S., Olsson, O., Vulturius, G., Carlsen, H., & Benzie, M. (2020). Climate risk in a globalized world: empirical findings from supply chains in the Swedish manufacturing sector. *Journal of environmental planning and management*, *63*(7), 1266-1282.
[75]. Trapani, D., Franzoi, M., Burstein, H., Carey, L., Delaloge, S., Harbeck, N., Hayes, D., Kalinsky, K., Pusztai, L., & Regan, M. (2022). Risk-adapted modulation through de-intensification of cancer treatments: an ESMO classification. *Annals of Oncology*, *33*(7), 702-712.
[76]. van Papendrecht1, B. C. H. (2018). FinTech Disruption Across the Wealth Management Value Chain–Will FinTech Dominate the Wealth Management Model of the Future or is there Still a Place for Traditional Wealth Managers? *The WealthTech Book: The FinTech Handbook for Investors, Entrepreneurs and Finance Visionaries*, 11-15.
[77]. Villar, A. S., & Khan, N. (2021). Robotic process automation in banking industry: a case study on Deutsche Bank. *Journal of Banking and Financial Technology*, *5*(1), 71-86.
[78]. Wang-Mlynek, L., & Foerstl, K. (2020). Barriers to multi-tier supply chain risk management. *The International Journal of Logistics Management*, *31*(3), 465-487.
[79]. Williams, J. (2021). Conclusion: FinTech—a perfect day or walk on the wild side? In *Disruptive technology in banking and finance: an international perspective on FinTech* (pp. 283-313). Springer.
[80]. Wu, X., Zheng, W., Xia, X., & Lo, D. (2021). Data quality matters: A case study on data label correctness for security bug report prediction. *IEEE Transactions on Software Engineering*, *48*(7), 2541-2556.
[81]. Xiao, Z., Fu, X., Zhang, L., & Goh, R. S. M. (2019). Traffic pattern mining and forecasting technologies in maritime traffic service networks: A comprehensive survey. *IEEE Transactions on Intelligent Transportation Systems*, *21*(5), 1796-1825.
[82]. Zaheda, K. (2025a). AI-Driven Predictive Maintenance For Motor Drives In Smart Manufacturing A Scada-To-Edge Deployment Study. *American Journal of Interdisciplinary Studies*, *6*(1), 394-444. https://doi.org/10.63125/gc5x1886
[83]. Zaheda, K. (2025b). Hybrid Digital Twin and Monte Carlo Simulation For Reliability Of Electrified Manufacturing Lines With High Power Electronics. *International Journal of Scientific Interdisciplinary Research*, *6*(2), 143–194. https://doi.org/10.63125/db699z21
[84]. Zaheda, K., & Md Hamidur, R. (2024). GPU-Accelerated Physics-Informed Digital Twins for Real-Time State Estimation and Fault Localization in Distribution Grids. *American Journal of Scholarly Research and Innovation*, *3*(02), 179-216. https://doi.org/10.63125/msrpfb04
[85]. Zaheda, K., & Md. Tahmid Farabe, S. (2023). Robotics and Computer Vision for Automated Inspection of Substation and Treatment-Facility Electrical Infrastructure. *Review of Applied Science and Technology*, *2*(04), 194-227. https://doi.org/10.63125/tfh15j12