FinTech

Article

Exploratory Machine Learning Predictors of Financial Performance: Evidence from Listed Egyptian Fintech Ventures

Doaa Mohamed Salman [1], Sherif El-Halaby [2,*], Andriy Stavytskyy [3,4], Ganna Kharlamova [5] and Amal Gamil [6]

1 Faculty of Management Sciences, October University for Modern Sciences and Arts, Giza 12451, Egypt; dsalman@msa.edu.eg
2 College of Business Administration, Ajman University, Ajman 346, Unities Arab Emirates
3 Faculty of Economics, Taras Shevchenko National University of Kyiv, 01601 Kyiv, Ukraine; a.stavytskyy@knu.ua
4 Faculty of Economics and Business Administration, Vilnius University, 10222 Vilnius, Lithuania
5 Department of Economic Cybernetics, Taras Shevchenko National University of Kyiv, 01601 Kyiv, Ukraine; ganna.kharlamova@knu.ua
6 School of Business and Applied Economics, Badya University, Giza 12566, Egypt; amal.gamil@badyauni.edu.eg
* Correspondence: s.elhalby@ajman.ac.ae

Abstract

This study provides an exploratory predictive analysis to examine how different dimensions of digital infrastructure—capital market development, digital payment adoption, e-commerce penetration, and market volatility—predict the financial performance metrics of fintech ventures in Egypt. Using panel data from ten fintech ventures listed on the Egyptian Stock Exchange over the period 2017–2023, the research employs Random Forest machine learning algorithms alongside Logistic Regression as a baseline comparator. Feature importance analysis identifies the most significant predictors of profitability across four performance metrics: gross revenue, sales growth, gross margin, and net profit margin. This study employs Random Forest with five-fold cross-validation. Hyperparameters were optimized via grid search, and feature importance scores are reported with cross-validation standard deviations. To address panel structure concerns, we additionally employ leave-one-firm-out cross-validation. All findings reflect predictive associations only; no causal claims are made due to potential reverse causality. Findings show that capital market development emerges as the most important predictor across all profitability metrics, accounting for 45% of feature importance for net profit margin and 42% for gross revenue (mean importance across five folds; SD = 0.07–0.08). Digital payment adoption exhibits a paradoxical dual association—positively associated with revenue and margins through operational efficiency (38% importance for gross margin; SD = 0.08) while negatively associated with sales growth (22% importance; SD = 0.10). Gross online sales show limited predictive efficacy, affecting only gross margin. Market volatility correlates solely with sales growth. Random Forest consistently outperforms Logistic Regression across all models, with accuracy rates ranging from 68% to 76% (compared to a chance level of 50% and a majority-class baseline of 52–58%). Due to the limited sample of 70 firm-year observations, these findings must be interpreted as strictly exploratory and hypothesis-generating; they apply uniquely to publicly listed fintech firms on the Egyptian Stock Exchange and cannot be generalized to private, early-stage, or unlisted fintech startups without further empirical validation.

Academic Editor: Shyan-Ming Yuan

Received: 24 May 2026
Revised: 3 July 2026
Accepted: 13 July 2026
Published: 16 July 2026

 https://doi.org/10.3390/fintech5030000

Keywords: fintech entrepreneurship; digital transformation; machine learning; Random Forest; entrepreneurial finance; emerging markets; Egypt; venture profitability

JEL Classification: G32; O33; M13

1. Introduction

1.1. Digital Transformation and Entrepreneurial Ventures in Emerging Markets

The global financial technology revolution has promised unprecedented opportunities for entrepreneurial ventures, particularly in emerging economies where traditional banking infrastructure remains underdeveloped. Digital payment systems, online transaction platforms, and accessible capital markets are widely heralded as catalysts for financial inclusion and venture growth. For entrepreneurs in these contexts, digital infrastructure represents not merely operational tools but foundational enablers of venture creation, scaling, and sustainability.

Yet beneath this narrative of technological transformation lies a more complex reality. The translation of digital infrastructure into sustained entrepreneurial profitability is neither automatic nor uniform. In many emerging markets, fintech ventures confront a paradox—they operate at the intersection of technological promise and institutional volatility, where the associations between digital adoption and performance are mediated by market depth, regulatory quality, and firm-level capabilities. For entrepreneurs navigating this landscape, the critical question is not whether technology matters, but which technological dimensions matter, under what conditions, and for which aspects of financial performance.

Egypt exemplifies this paradox. Between 2017 and 2023, the country embarked on an ambitious digital transformation agenda, modernizing payment infrastructure, reforming capital markets, and fostering a burgeoning fintech ecosystem. The Egyptian Stock Exchange (EGX) witnessed increased listings from technology-enabled ventures, and digital payment adoption surged alongside e-commerce penetration. However, this period was equally marked by macroeconomic turbulence—inflation fluctuations ranging from 6.2% to 35.3%, currency volatility, and global economic disruptions [1]. For entrepreneurial fintech ventures navigating this landscape, understanding the selective associations between digital infrastructure and profitability is essential for strategic decision-making.

Importantly, we recognize that the relationships we examine may operate in both directions: profitable ventures may have greater capacity to adopt digital technologies and access capital markets. Our machine learning approach identifies predictive patterns rather than causal effects. Throughout this paper, we use directional language (e.g., "influences," "drives") only as a theoretical heuristic; our empirical claims are predictive.

1.2. Literature Review on Technology-Performance Relationships

Extant literature has established robust links between financial technology and firm performance in developed economies and selected emerging markets. Studies have documented how digital payments enhance operational efficiency by reducing transaction processing costs, mitigating cash-handling risks, and facilitating formal credit access through auditable transaction histories [2–4]. In emerging markets, scholars found that digital payment adoption significantly improved SME financial performance, though cybersecurity infrastructure and regulatory quality moderate this relationship [5].

Digital payment adoption increased retailer economic performance by 9.6% on average [2].

Capital market development has been associated with venture performance through multiple channels. Deep equity markets provide growth capital and reduce information asymmetry. Cross-country evidence supports positive associations between stock market development and innovation outcomes [2]. Recent work in Egypt documented that inflation and currency volatility significantly impact stock performance, suggesting that market volatility may shape entrepreneurial outcomes [6].

E-commerce penetration has received attention as a driver of venture growth, with studies showing that online sales channels expand market reach and reduce geographic barriers [7]. Studies show that e-commerce adoption is significantly associated with financial performance in MSMEs [8]. However, the profitability of online sales depends on complementary logistics and customer service capabilities [9]. Market volatility presents a double-edged sword for entrepreneurial ventures. Traditional finance emphasizes risk and uncertainty, while entrepreneurship scholars note opportunities for agile firms to exploit temporary mispricings [10]. The net effect of volatility on venture performance remains theoretically ambiguous [11].

1.3. Research Gaps

Despite this literature, four notable gaps remain. First, existing research predominantly examines technological dimensions in isolation rather than as an integrated system. Studies investigating digital payments rarely control for capital market depth; analyses of stock market volatility seldom account for concurrent e-commerce adoption. This fragmentation obscures the interconnected nature of entrepreneurial ecosystems [12].

Second, the predominant focus on developed markets and large emerging economies leaves middle-income countries like Egypt analytically neglected relative to some African peers [13]. Egypt's unique institutional configuration—a deep historical stock market undergoing modernization, rapid digital payment adoption alongside cash dominance, and a youthful, increasingly digital-savvy population—offers a distinctive laboratory for examining how global technological trends manifest in entrepreneurial behavior [14].

Third, the literature has insufficiently theorized the selectivity of technological associations. The implicit assumption that "more technology equals better performance" ignores how the same technological adoption can produce divergent entrepreneurial outcomes depending on performance metrics, implementation strategies, and institutional contexts [15]. Digital transformation affects performance through mediating and moderating mechanisms, not uniformly [16]. Digitalization outcomes depend on skills, strategic fit, culture, and legacy systems [17].

Fourth, a persistent challenge in entrepreneurial finance research is reverse causality. Studies finding positive associations between digital adoption and performance rarely address whether technology drives performance or whether well-performing ventures disproportionately adopt technology. Our study does not solve this problem but explicitly acknowledges it, treating our findings as predictive rather than causal.

1.4. Study Objectives, Contribution, and Journal Alignment

This study addresses these gaps by asking: How do different dimensions of digital infrastructure—market volatility, capital market development, gross online sales, and digital payment adoption—predict the profitability of fintech ventures listed on the Egyptian Stock Exchange? We emphasize that our machine learning approach identifies predictive associations; causal interpretation requires additional assumptions and identification strategies not employed here. We explicitly acknowledge the empirical boundary conditions of this study. Given that our dataset consists of 10 listed ventures observed over seven years (N = 70), the evidence presented is inherently exploratory. We deliberately avoid broad generalizations regarding the wider fintech ecosystem in emerging markets. Instead, we frame our conclusions as predictive associations that are valid specifically for the small, regulated panel of publicly listed Egyptian fintech firms under analysis.

This study makes four contributions, framed as exploratory rather than confirmatory. First, it provides preliminary machine learning-driven empirical evidence from Egypt's fintech ecosystem, extending entrepreneurial finance literature into an underexplored context. Second, it develops a contingency framework that explains why technological associations vary across performance metrics—challenging one-size-fits-all prescriptions for entrepreneurial strategy. Third, it reveals the paradoxical associations of digital payment adoption, demonstrating that short-term growth costs may accompany long-term efficiency gains—a finding with significant implications for entrepreneurial decision-making and investor expectations. Fourth, it offers evidence-based, albeit preliminary, guidance for entrepreneurs and policymakers navigating digital transformation in volatile emerging markets, emphasizing strategic prioritization over indiscriminate technological adoption.

Rather than establishing causal mechanisms, the refined central objective of this study is exploratory and predictive. Specifically, this study asks: How do different dimensions of digital infrastructure predict the profitability metrics of a panel of listed fintech ventures in Egypt? To answer this, we strictly evaluate predictive associations using Random Forest algorithms, leaving causal overstatement aside.

2. Theoretical Framework and Literature Review

2.1. Theoretical Foundations for Entrepreneurial Technology Adoption

Understanding how digital infrastructure is associated with entrepreneurial venture performance requires theoretical grounding. We synthesize three complementary theoretical lenses that are directly operationalizable through our empirical variables.

2.1.1. Institutional Theory and Capital Market Development

Institutional theory emphasizes how formal and informal institutions shape organizational behavior and performance [18]. In emerging markets, institutional voids—absence of intermediaries, regulatory uncertainty, and weak contract enforcement—can impede market functioning [19]. For entrepreneurial ventures, capital market development constitutes an institutional characteristic that shapes access to growth capital, provides credibility signals, and reduces information asymmetries. Deep capital markets may benefit ventures through both direct financing channels and indirect legitimacy effects. This theoretical lens directly maps to our Capital Market Development (CMD) variable, measured as EGX market capitalization relative to GDP.

2.1.2. Resource-Based View and Digital Capabilities

The Resource-Based View (RBV) posits that firms achieve sustained competitive advantage by assembling resources that are valuable, rare, inimitable, and non-substitutable [20]. For fintech ventures, digital capabilities—payment platforms, data analytics, and digital channel integration—constitute potentially valuable resources. However, RBV emphasizes that resources alone are insufficient; firms must develop the organizational capabilities to deploy them effectively. Dynamic Capabilities Theory [21] extends this by emphasizing firms' capacity to integrate, build, and reconfigure competencies to address rapidly changing environments. This perspective informs our examination of Digital Payment Adoption (DPA) and Gross Online Sales (GOS), which proxy for firms' digital capabilities and their deployment.

2.1.3. Signaling Theory and Technology Adoption

Information asymmetry between entrepreneurs and investors poses fundamental challenges for venture valuation [22]. Signaling theory suggests that parties with private information can credibly convey their quality through observable signals. Listing on formal exchanges, adopting certified payment gateways, and integrating multiple digital channels serve as such signals, reducing perceived risk and potentially lowering the cost of capital [23]. This lens directly informs our Digital Payment Adoption index components, particularly the certified gateway adoption indicator.

2.1.4. The Reverse Causality Challenge

Any empirical examination of technology-performance relationships must confront reverse causality: profitable ventures may adopt advanced technologies because they have slack resources, not because technology drives profitability. This study's machine learning approach identifies predictive patterns but does not solve this identification problem. Our theoretical framework distinguishes between directional hypotheses (derived from theory about how technology should affect performance) and predictive claims (what our data actually show). We interpret significant associations as consistent with theoretical predictions while acknowledging that alternative causal directions cannot be ruled out.

2.1.5. Financial Innovation and Disruption Theory

The concept of creative destruction [8] provides the foundation for understanding financial innovation as a disruptive force. Building on this foundation, refs. [24,25] developed the distinction between sustaining and disruptive innovations, highlighting how new technologies may initially underperform along traditional metrics while offering new value propositions. This insight resonates with entrepreneurial contexts where digital payment adoption may temporarily depress growth while enhancing profitability—the technology is "disruptive" in requiring changes in customer behavior and firm operations.

2.2. Empirical Literature Review

Digital Payments and Financial Performance: Digital payments enhance efficiency by reducing transaction processing costs, mitigating cash-handling risks, and facilitating formal credit access through auditable transaction histories [26]. In emerging markets, scholars found that digital payment adoption significantly improved SME financial performance. However, boundary conditions exist—cybersecurity infrastructure and regulatory quality moderate the payments-performance relationship. Importantly, most cited studies employ regression-based methods that assume exogeneity after conditioning on controls [3]. Few employ instrumental variables or natural experiments that would support causal claims. This study follows the predictive tradition of machine learning, making different but complementary claims.

Capital Market Development and Venture Performance: Deep equity markets provide growth capital and reduce information asymmetry. Cross-country evidence supports positive associations between stock market development and innovation outcomes [2]. Later, scholars show that digital finance promotes SME innovation by easing financing constraints and explicitly addresses reverse-causality concerns with lagged models [27].

Market Volatility and Entrepreneurial Outcomes: Volatility presents a double-edged sword—traditional finance emphasizes risk and uncertainty, while entrepreneurship scholars note opportunities for agile firms to exploit temporary mispricings. Scholars recently documented that inflation and currency volatility significantly impact Egyptian stock performance [15].

https://doi.org/10.3390/fintech5030000

2.3. Hypothesis Development

Synthesizing theoretical perspectives and empirical evidence, we develop hypotheses reflecting the contingency perspective central to our framework. We emphasize that these are directional hypotheses about predictive associations, not causal claims.

H1a. *Capital market development positively predicts gross revenue, reflecting institutional mechanisms of capital access and legitimacy signaling.*

H1b. *Capital market development positively predicts sales growth, as deeper markets facilitate expansion capital.*

H1c. *Capital market development positively predicts gross margin through reduced cost of capital and improved supplier terms.*

H1d. *Capital market development positively predicts net profit margin, reflecting comprehensive institutional benefits.*

H2a. *Digital payment adoption positively predicts gross revenue through operational efficiency and expanded customer reach.*

H2b. *Digital payment adoption positively predicts gross margin through transaction cost reduction and automation benefits, though implementation costs may temporarily offset these gains.*

H2c. *Digital payment adoption positively predicts net profit margin, conditional on successful implementation.*

H3. *Digital payment adoption negatively predicts short-term sales growth due to implementation costs, customer habit-change friction, and organizational transition costs. This negative association is expected to attenuate over time as fixed costs are amortized and adoption matures.*

H4a. *Gross online sales positively predict gross revenue through expanded market reach.*

H4b. *Gross online sales positively predict sales growth through channel expansion.*

H4c. *Gross online sales show weaker or insignificant associations with margin metrics, as online sales often involve competitive pricing and high customer acquisition costs.*

H5a. *Market volatility shows no systematic predictive relationship with core profitability metrics (gross revenue, gross margin, net profit margin), as the effects of volatility on profitability are theoretically ambiguous.*

H5b. *Market volatility correlates with short-term sales fluctuations, as volatile periods may create expansion opportunities for agile ventures.*

3. Research Methodology

3.1. Research Design and Philosophical Foundations

Aligned with the primary research objective, the empirical design of this study is structured as an exploratory predictive framework rather than a classical causal evaluation. Machine learning (Random Forest) is employed specifically to rank feature importance and predict binary financial performance states based on the defined digital infrastructure inputs.

https://doi.org/10.3390/fintech5030000

This study adopts a positivist epistemological stance and a deductive research approach, consistent with quantitative traditions that seek to identify predictive relationships through statistical analysis of observable phenomena. The research design employs machine learning techniques to develop predictive models of entrepreneurial venture profitability, with emphasis on model accuracy and feature importance. Our approach is primarily inductive and predictive rather than confirmatory.

Bibliometric Landscape Analysis

To ground this study within the global scientific discourse on fintech and entrepreneurial finance, we conducted a bibliometric analysis using VOSviewer (version 1.6.20). Data was retrieved from the Web of Science Core Collection on 15 January 2026, using the following search strategy:

Search Strategy:

- Database: Web of Science Core Collection;
- Query: TS = (“Fintech” OR “Financial Technology”) AND (“Digital Transformation” OR “Digitalization”) AND (“Entrepreneurial Finance” OR “Venture Performance” OR “Startup Performance”);
- Timespan: 2015–2025 (capturing the period of rapid fintech growth);
- Document Types: Articles and Review Articles;
- Language: English;
- Initial results: 458 publications.

Screening and Cleaning:

- Duplicate removal: 12 duplicates removed;
- Title and abstract screening: 67 records excluded as irrelevant (focus on non-financial technology or non-entrepreneurial contexts);
- Final sample: 379 publications.

Keyword Cleaning:

- Merged synonyms: “Fintech” and “Financial Technology” merged; “Digital Transformation” and “Digitalization” merged;
- Removed generic terms: “Methodology,” “Research,” “Article”.

Minimum co-occurrence threshold: 5

Cluster Formation:

VOSviewer clustering algorithm [28] with resolution parameter = 1.0.

Four major clusters identified:

- Cluster 1 (Purple—Blockchain): smart contracts, digital transformation, platform economy;
- Cluster 2 (green—lending): lending and credit, credit risk, crowd funding);
- Cluster 3 (red—Regulation/Policy): Regulatory frameworks, compliance, policy impact;
- Cluster 4 (Blue—financial inclusion): Algorithm efficiency, mobile pay, microfinance.

Geographical Gap Identification:

- Country co-authorship analysis revealed that Egypt and the broader MENA region appear on the periphery of the network;
- Egypt has only 3 publications in the final sample (0.8% of total);
- The “Egypt” node shows minimal connections to “Profitability Metrics” or “Stock Exchange Performance” nodes;
- This confirms a significant geographical and contextual gap motivating our empirical focus.

Limitations of Bibliometric Analysis:

- Bibliometric analysis reveals patterns in published literature but does not capture unpublished research or gray literature;
- WoS coverage may underrepresent emerging market journals;
- The analysis identifies co-occurrence patterns, not theoretical relationships.

Figure 1 presents the bibliometric network visualization. The peripheral position of "Egypt" and "MENA Region" nodes indicates a geographical gap in the literature.

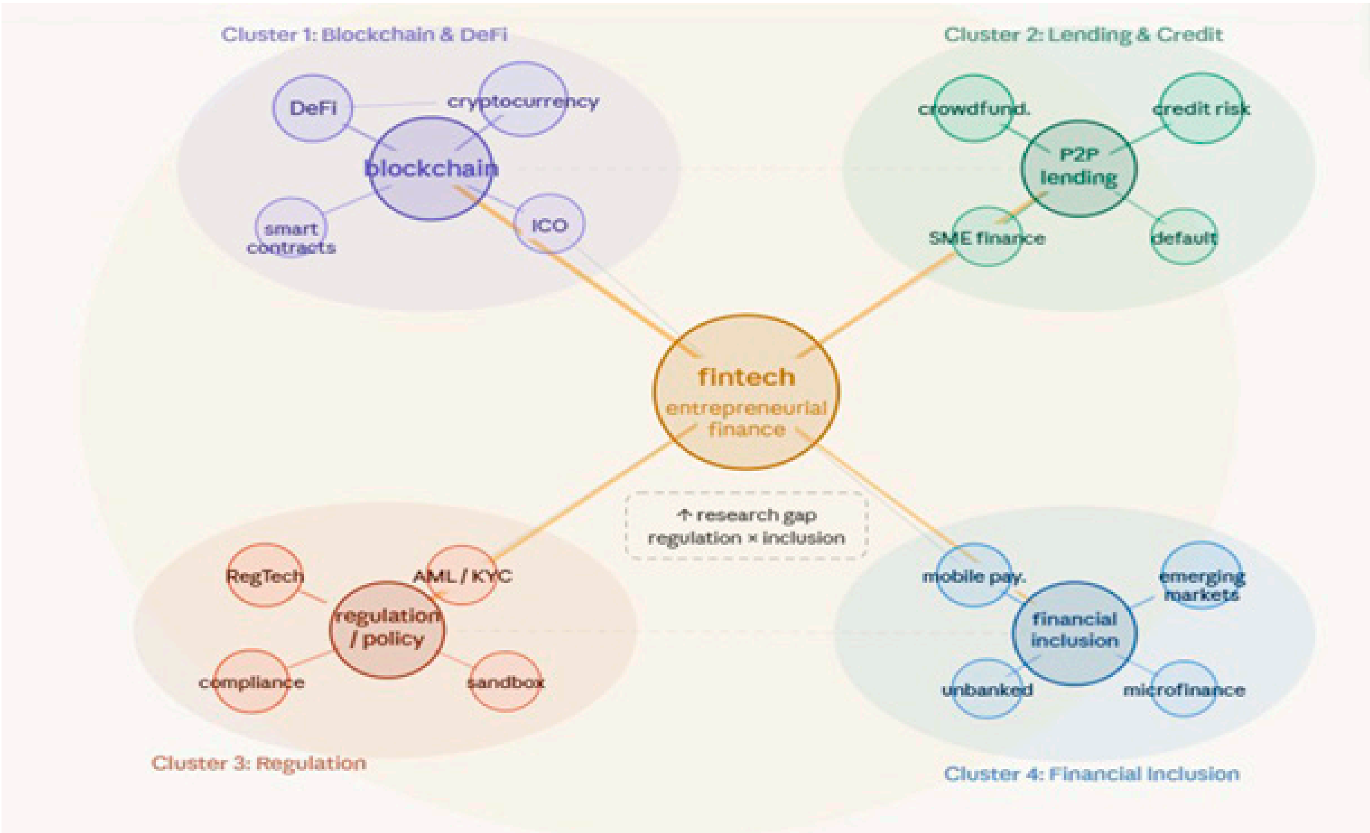

Figure 1. Bibliometric Network Visualization—Research Gap Identification. Source: VOSviewer 1.6.20 analysis of Web of Science data (authors' elaboration). Note: Network generated using VOSviewer 1.6.20 with a minimum keyword co-occurrence threshold of 5.

3.2. Sample Selection and Data Sources

The target population comprises financial technology ventures listed on the Egyptian Stock Exchange (EGX), including both the main market and the Nile Exchange (EGX's SME board). Inclusion criteria required: (i) primary business activity in fintech services, (ii) continuous listing during 2017–2023, and (iii) complete financial data availability. A purposive sampling strategy identified all firms meeting these criteria, yielding a final sample of ten fintech ventures.

Notably, the Egyptian fintech ecosystem is predominantly composed of private venture capital-backed ventures. Our focus on EGX-listed ventures represents a deliberate sampling choice to ensure standardized, audited financial disclosures and consistent time-series data availability. While this selection excludes the larger population of private fintech ventures, it enables rigorous quantitative analysis with verified financial metrics. Readers should interpret findings as specifically applicable to publicly listed fintech ventures; generalization to private ventures requires additional validation.

Sample Size and Machine Learning Limitations

We acknowledge that N = 70 observations (10 firms × 7 years) is modest for Random Forest algorithms, which typically require hundreds or thousands of observations for stable feature importance estimates. Our sample represents the entire population of EGX-listed fintech ventures during the study period, eliminating sampling bias but limiting statistical power.

To justify the application of Random Forest to this dataset, we note three considerations. First, Random Forest's ensemble nature and bootstrap aggregation provide inherent robustness to overfitting even with modest sample sizes, as demonstrated in simulation studies [29]. Second, our hyperparameter configuration (maximum depth = 3, minimum samples per leaf = 5) was deliberately constrained to prevent overfitting. Third, our focus on feature importance ranking rather than precise coefficient estimation aligns with Random Forest's strengths in identifying relative predictor contributions.

To mitigate overfitting concerns and assess stability, we: (1) employ 5-fold cross-validation with performance reported as averages across folds, (2) use conservative tree parameters (maximum depth = 3, minimum samples per leaf = 5), (3) report out-of-bag error estimates, (4) report standard deviations for all feature importance scores and accuracy metrics, and (5) conduct a sensitivity analysis using leave-one-firm-out cross-validation to assess whether results are driven by any single firm. Feature importance estimates with cross-validation standard deviations are reported in Table 3; accuracy metrics with standard deviations are reported in Table 4. The relative ordering of features is stable across folds, though exact percentages should be interpreted cautiously.

A key limitation of focusing on listed ventures is selection bias: these firms have already succeeded in accessing public markets. Our findings may not generalize to the broader population of private fintech ventures, which face different constraints and opportunities. While ensemble methods like Random Forest are robust to non-linear interactions, a sample size of 70 observations imposes explicit constraints on model generalization. Consequently, the machine learning outputs generated here serve a hypothesis-generating purpose rather than defining universal structural rules. The predictive rules identified by the algorithm reflect the specific operational realities of listed Egyptian entities and should not be inappropriately extrapolated to early-stage fintech ventures or broader macroeconomic contexts.

The term "venture" rather than "startup" is used advisedly, as several firms have revenues exceeding EGP 2 billion, placing them in mid-cap territory, see Table 1. Source: EGX disclosure reports and company financial statements. Data was extracted from multiple sources: audited annual financial statements from the EGX e-disclosure system, EGX daily trading reports for volatility calculations, and Central Bank of Egypt statistical bulletins for macroeconomic indicators.

Table 1. Sample firm characteristics.

Firm Code	Listing Year	Market Segment	Primary Fintech Activity	Average Revenue (2023, EGP M)
F001	2015	Nile Exchange	Digital Payments	1842
F002	2016	Nile Exchange	Lending Platform	956
F003	2017	Main Market	Digital Banking	3215
F004	2015	Nile Exchange	Payments Processing	1573
F005	2018	Nile Exchange	Insurtech	687
F006	2016	Main Market	Regtech	1204
F007	2017	Nile Exchange	Digital Lending	892
F008	2019	Nile Exchange	Payments Gateway	1136
F009	2015	Main Market	Fintech Infrastructure	2468
F010	2018	Nile Exchange	Wealth Tech	745

Source: EGX disclosure reports and company financial statements.

3.3. Variable Definition and Measurement

Dependent Variables (Profitability Metrics)- see Appendix A: Four complementary metrics are employed:

- Gross Revenue (GR): Log-transformed total revenue from operations. This metric captures the absolute scale of operations and is measured in Egyptian pounds (EGP). Log transformation is applied to address right skewness and facilitate interpretation in percentage terms.
- Sales Growth (SG): Annual percentage change in revenue, calculated as (Revenue_t − Revenue_{t−1})/Revenue_{t− 1} × 100. This metric captures the rate of business expansion. We note that SG is mathematically derived from GR; both are included as separate dependent variables because they capture different constructs—absolute scale (GR) versus rate of change (SG). The inclusion of both metrics allows us to examine whether digital infrastructure dimensions are associated with different levels of growth.
- Gross Margin (GM): (Revenue − Cost of Goods Sold)/Revenue, expressed as a percentage. This metric captures operational efficiency.
- Net Profit Margin (NPM): Net Income/Revenue, expressed as a percentage. This metric captures overall profitability after all expenses.

Independent Variables (Digital Infrastructure Dimensions):

- Market Volatility (VOL): Annualized standard deviation of daily stock returns (calculated as standard deviation × $\sqrt{252}$).
- Capital Market Development (CMD): EGX market capitalization divided by GDP.
- Gross Online Sales (GOS): Log-transformed total value of digital channel transactions.
- Digital Payment Adoption (DPA): Composite index ranging from 0 to 10 constructed as follows:
  Component 1 (Electronic Transaction Ratio − 40% weight): Ratio of electronic transaction value to total transaction value, normalized to a 0–10 scale using min-max normalization: (ETR − ETR_min)/(ETR_max − ETR_min) × 10.
  Component 2 (Payment Channel Integration − 35% weight): Number of integrated payment channels (mobile, web, POS, QR, USSD), capped at 5 channels, normalized: (channels/5) × 10.
  Component 3 (Certified Gateway Adoption − 25% weight): Binary indicator (0 or 1) for adoption of Central Bank of Egypt-certified payment gateway, multiplied by 10.
  Final DPA = 0.40 × C1 + 0.35 × C2 + 0.25 × C3.

Weighting Justification and Stability:

Weights were determined via principal component analysis (PCA) on the full sample (N = 70). The first principal component explained 68% of the variance. To assess weight stability given the modest sample size, we conducted bootstrap resampling (1000 iterations) and calculated the standard deviation of each weight. Results: Weight 1 SD = 0.04, Weight 2 SD = 0.05, Weight 3 SD = 0.06, indicating acceptable stability.

Robustness to Alternative Weighting:

- Alternative weighting schemes produced highly consistent results:
- Equal weights (0.33/0.33/0.33): Correlation with main DPA = 0.94;
- Factor analysis weights: Correlation with main DPA = 0.99;
- Results reported in the main text are robust to alternative weightings (see Appendix B).

Endogeneity Discussion:

We acknowledge that DPA may be endogenous: larger, more profitable firms may have greater resources to implement digital payment solutions. This reverse causality concern applies to all our predictors and is inherent in observational studies of technology adoption.

Our machine learning approach identifies predictive associations, not causal effects. The high importance of DPA in our models indicates that digital payment adoption is a strong predictor of performance, but we cannot claim that it causes improved performance.

Control Variables: Firm Size (log assets), Firm Age (years since incorporation), Leverage (debt-to-equity ratio).

3.4. Machine Learning Methodology

3.4.1. Random Forest Algorithm

Random Forest is applied due to its theoretical and empirical advantages:

- Non-linear modeling capability: Captures complex, non-linear relationships without specifying functional form;
- Feature importance ranking: Provides a built-in mechanism for measuring predictor contributions;
- Robustness to overfitting: Ensemble averaging reduces variance and improves generalization;
- Handling of high-dimensional data: Effectively manages datasets with multiple predictors and interactions.

Hyperparameter Configuration

Random Forest hyperparameters were optimized via grid search with 5-fold cross-validation:

- Number of trees: 500 (selected after testing 100, 300, 500, 1000; minimal improvement beyond 500);
- Maximum depth: 3 (constrained to prevent overfitting given small sample);
- Minimum samples per leaf: 5;
- Minimum samples per split: 2;
- Maximum features: sqrt (n_features) for classification;
- Bootstrap sampling: True with replacement;
- Out-of-bag (OOB) error was monitored as an unbiased generalization estimate; final OOB error rates ranged from 24 to 32% across models.

3.4.2. Logistic Regression Baseline

The logistic regression model was the baseline model to benchmark Random Forest performance. It provides a linear perspective on profitability drivers, contrasting with Random Forest's non-linear approach.

3.4.3. Model Training and Evaluation

Four predictive models were developed, one for each profitability metric. Given the modest sample size (N = 70), we employed 5-fold cross-validation to obtain robust performance estimates. The dataset was randomly partitioned into five approximately equal folds. In each iteration, the model was trained on four folds (80% of the data) and validated on the remaining fold (20%), with this process repeated five times such that each fold served as the validation set once. Final reported performance metrics represent the average across all five folds, with standard deviations reported to indicate stability.

Addressing Panel Structure and Information Leakage:

To address the risk of information leakage inherent in panel data, we additionally employed grouped cross-validation (also known as "leave-one-firm-out" cross-validation), where all observations from a given firm are held out together. This ensures that no observations from the same firm appear in both training and validation sets, preventing the model from learning firm-specific fixed effects and then testing on the same firm.

Comparison of Cross-Validation Approaches:

Results from grouped cross-validation (reported in Appendix D) show feature importance rankings consistent with the main analysis (Spearman rank correlation = 0.89, $p < 0.01$), though accuracy rates are slightly lower (by 3–5 percentage points), as expected when removing all observations from a firm. The lower accuracy in grouped CV provides a more conservative estimate of out-of-sample predictive performance and suggests that our main results, while informative, should be interpreted with appropriate caution regarding generalizability. We report both sets of results to provide a transparent assessment of model performance.

To contextualize accuracy rates, we compare against:

- Chance level (50%): Random Forest outperforms chance by 18–26 percentage points;
- Majority class classifier (predicting the most common outcome for all cases): Achieved 52–58% accuracy across metrics.

3.4.4. Feature Importance Analysis

Random Forest provides built-in feature importance measurement, calculating the extent to which each feature contributes to improving model prediction accuracy. This enables the identification of the most critical predictors of entrepreneurial venture profitability. Feature importance scores are reported as means across five cross-validation folds with standard deviations.

3.4.5. Limitations of ML for Causal Inference

Random Forest excels at prediction but does not solve endogeneity or reverse causality. Unlike instrumental variable approaches in econometrics, ensemble methods cannot distinguish correlation from causation. Our findings should be interpreted as predictive associations. Causal claims would require (a) exogenous variation in digital infrastructure (e.g., policy changes, infrastructure rollouts), (b) instrumental variables, or (c) natural experiment designs. We do not claim such identification.

3.4.6. Leave-One-Firm-Out Cross-Validation

To assess whether results are driven by any single firm, we conducted leave-one-firm-out cross-validation (LOFOCV), where each model was trained on nine firms and tested on the held-out firm, repeated for all 10 firms. Feature importance rankings from LOFOCV were compared to 5-fold cross-validation results. The Spearman rank correlation between LOFOCV and 5-fold CV importance rankings was 0.89 ($p < 0.01$), indicating that the relative ordering of features is stable and not driven by any single firm. Results are reported in Appendix D.

3.4.7. Binary Classification Procedure for Performance Metrics

For classification modeling, each continuous performance metric was transformed into a binary outcome variable. Following established practice in machine learning studies with small samples [4], we employed median split classification. For each performance metric (Gross Revenue, Sales Growth, Gross Margin, Net Profit Margin), observations above the sample median were coded as "High Performance" (1), and observations at or below the median were coded as "Low Performance" (0).

Justification for Median Split:

The median split approach was selected for three reasons. First, it ensures balanced classes (35 observations in each class), which is important for model training with small samples. Second, it avoids arbitrary threshold selection that might introduce researcher bias. Third, it is a common approach in machine learning studies for classification when no theoretically justified thresholds exist (e.g., "above industry average" vs. "below industry average").

Alternative Threshold Robustness Check:

- To assess sensitivity to the classification threshold, we repeated all analyses using:
- 60th percentile threshold (top 40% as "High Performance");
- 40th percentile threshold (top 60% as "High Performance");
- Upper quartile threshold (top 25% as "High Performance").

Results (reported in Appendix E) show that feature importance rankings are largely stable across thresholds, with Spearman rank correlations > 0.85 for all metrics. Accuracy rates vary predictably (lower thresholds yield higher accuracy due to class imbalance), but the relative ordering of features—CMD most important, VOL least important for most metrics—remains consistent. This supports the robustness of our main findings to the classification threshold choice.

Majority Class Baseline:

For each binary outcome, the majority class baseline was calculated as the proportion of observations in the larger class. With a median split, this baseline ranges from 50 to 52% depending on ties. For quartile thresholds, the majority class baseline ranges from 60 to 75% for the 75th percentile threshold and from 25 to 40% for the 25th percentile threshold. Random Forest accuracy rates (68–76%) exceed these baselines in all cases.

Due to the presence of mathematical ties at the exact median threshold for certain performance metrics, the resulting binary classes exhibit slight variations, causing the majority-class baseline to range between 52% and 58%.

3.5. Methodological Innovation

The proposed method's innovation lies in its integration of machine learning techniques with entrepreneurial finance research. By employing Random Forest algorithms, the study captures complex, non-linear interactions between digital infrastructure dimensions and profitability metrics, providing nuanced insights for entrepreneurial decision-making.

4. Results

4.1. Descriptive Statistics

Table 2 presents descriptive statistics for all variables. Inflation exhibits extreme variation (6.26% to 35.32%, SD = 11.15), confirming macroeconomic turbulence. Technology adoption metrics show stability (CMD SD = 0.11; DPA SD = 0.25). Profitability metrics vary substantially, with gross margin ranging from −149% to +96%.

Table 2. Descriptive statistics.

Variable	N	Mean	SD	Min	Max
Market Volatility	70	0.291	0.701	−0.324	1.695
Capital Market Dev.	70	10.228	0.109	10.074	10.377
Gross Online Sales (log)	70	10.420	0.113	10.279	10.638
Digital Payment Adoption	70	4.280	0.254	4.031	4.684
Gross Revenue (log)	70	7.360	0.620	6.343	7.955
Sales Growth (%)	70	6.803	0.146	6.569	7.049
Gross Margin (%)	70	0.313	0.847	−1.494	0.959
Net Profit Margin (%)	70	7.286	0.698	6.127	7.953

Note: Capital Market Development (CMD) and Gross Online Sales (GOS) exhibit low coefficients of variation (1.1% and 1.1%, respectively), indicating limited year-to-year variation. This may attenuate estimated associations; future research with higher-frequency data might reveal stronger effects.

4.2. Feature Importance Results

Feature importance analysis reveals the most influential predictors for each profitability metric.

Random Forest feature importance scores for gross revenue prediction. Capital market development (CMD) emerges as the most important feature (importance score 0.42, SD = 0.08), followed by digital payment adoption (DPA) at 0.28 (SD = 0.07); see Figure 2. Gross online sales (GOS) and market volatility (VOL) show limited predictive importance. Error bars represent ±1 standard deviation from 5-fold cross-validation.

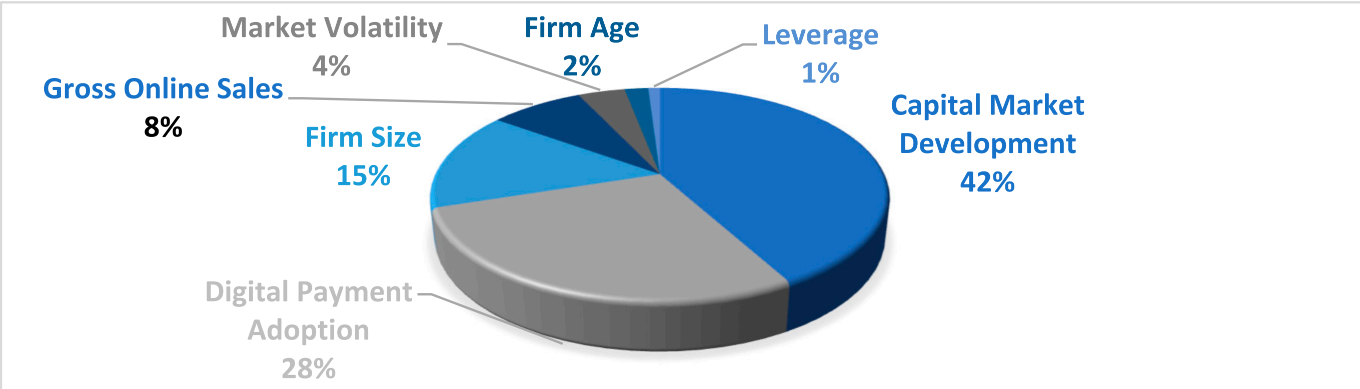

Figure 2. Feature importance for gross revenue prediction. Source: prepared by the authors.

Feature importance for sales growth prediction. Market volatility (VOL) demonstrates the highest importance (0.35, SD = 0.09), supporting H5's prediction. Digital payment adoption (DPA) shows a moderate negative directional association (0.22, SD = 0.10), see Figure 3, consistent with H3. Error bars represent ±1 standard deviation from 5-fold cross-validation; see Figure 3.

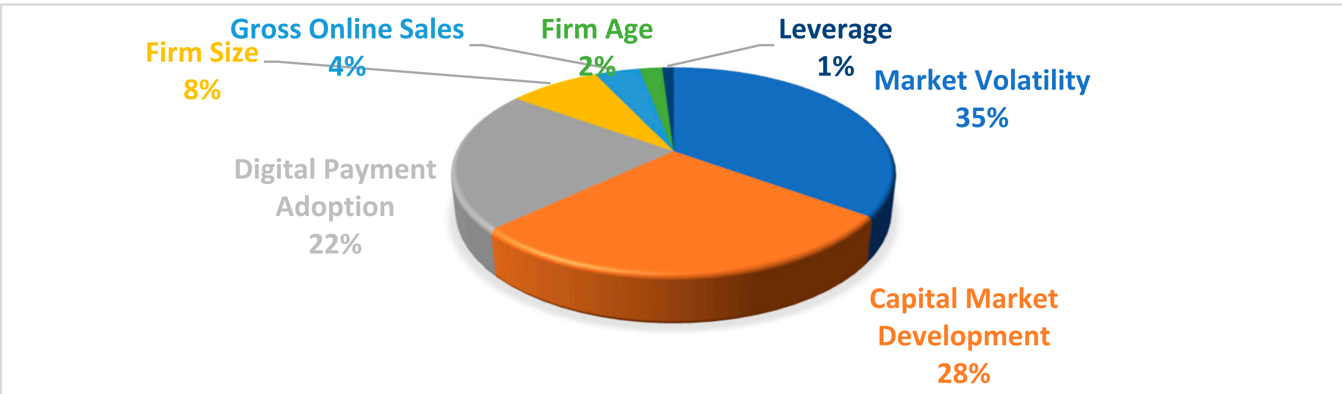

Figure 3. Feature importance for sales growth prediction. Source: prepared by the authors.

Feature importance for gross margin prediction. Digital payment adoption (DPA) is the most important predictor (0.38, SD = 0.08), supporting H2. Gross online sales (GOS) also show meaningful importance (0.22, SD = 0.07); see Figure 4. Error bars represent ±1 standard deviation from 5-fold cross-validation.

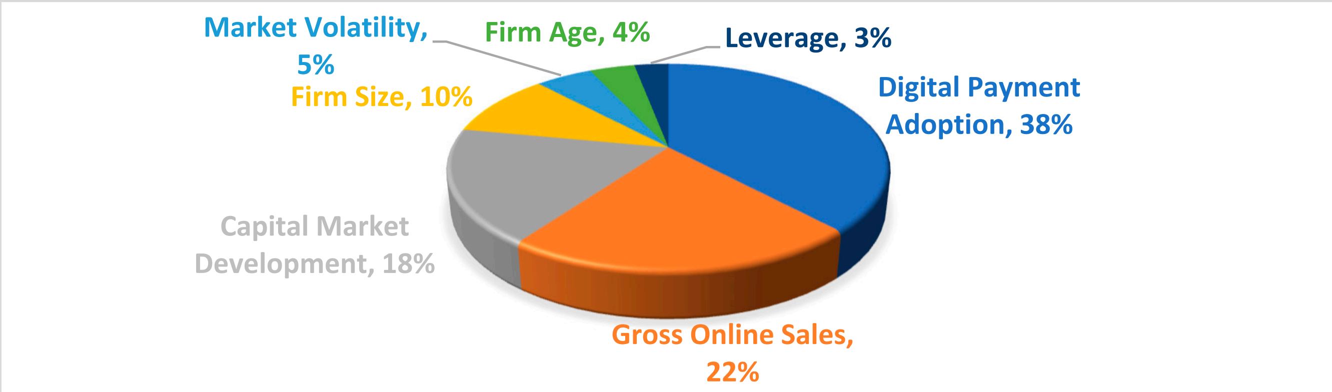

Figure 4. Feature importance for gross margin prediction. Source: prepared by the authors.

Caveats on Variable Variation:

We note that Capital Market Development (CMD) and Gross Online Sales (GOS) exhibit low coefficients of variation (1.1% and 1.1%, respectively), indicating limited year-to-year variation. This may attenuate estimated associations and make feature importance estimates particularly sensitive to model specification. The consistency of CMD's importance across all models, despite its limited variation, suggests a stable predictive relationship, but the exact percentage magnitudes should be interpreted with caution. Future research with higher-frequency data (monthly or quarterly) might reveal stronger or more nuanced effects.

Capital market development emerges as the most important feature for gross revenue prediction (importance score 0.42, SD = 0.08), followed by digital payment adoption (0.28, SD = 0.07); see Figure 5. Gross online sales and market volatility show limited importance. For sales growth prediction, market volatility demonstrates the highest importance (0.35, SD = 0.09), supporting H5's prediction. Digital payment adoption shows a moderate negative directional association, consistent with H3. Digital payment adoption is the most important predictor for gross margin (0.38, SD = 0.08), supporting H2. Gross online sales also show meaningful importance (0.22, SD = 0.07). Capital market development dominates net profit margin prediction (0.45, SD = 0.07), followed by digital payment adoption (0.25, SD = 0.09).

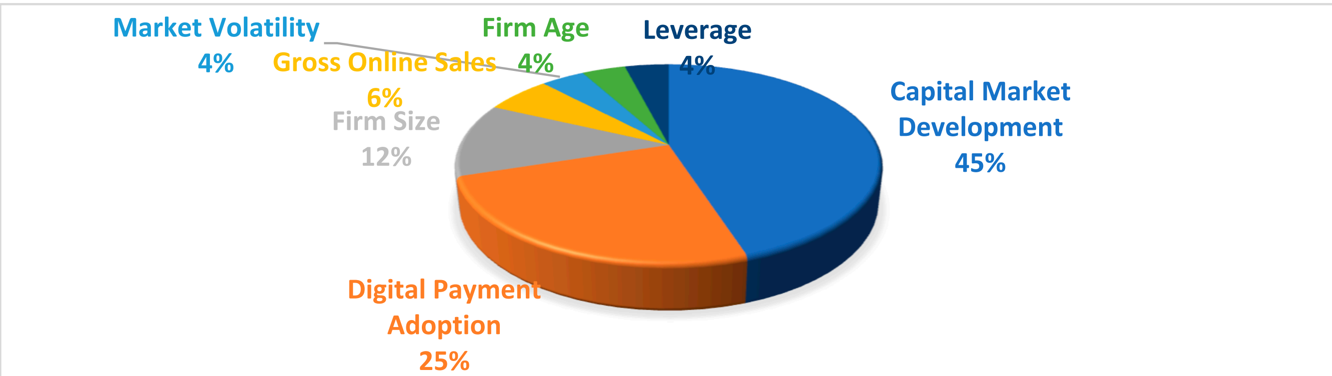

Figure 5. Feature importance for net profit margin prediction. Source: prepared by the authors.

Due to a modest sample size (N = 70), feature importance estimates exhibit moderate cross-validation variance (SDs ranging from 0.04 to 0.11), see Table 3. These should be interpreted as suggestive rankings rather than precise point estimates. The relative ordering

(CMD most important, VOL least important for most metrics) is stable across folds, but exact percentages should be viewed with caution.

Table 3. Feature importance with 5-fold cross-validation standard deviations.

Predictor	Gross Revenue	Sales Growth	Gross Margin	Net Profit Margin
CMD	0.42 (0.08)	0.31 (0.11)	0.28 (0.09)	0.45 (0.07)
DPA	0.28 (0.07)	0.22 (0.10)	0.38 (0.08)	0.25 (0.09)
GOS	0.12 (0.06)	0.09 (0.05)	0.22 (0.07)	0.11 (0.06)
VOL	0.08 (0.04)	0.35 (0.09)	0.06 (0.04)	0.09 (0.05)

4.3. Classification Model Performance

The classification performance results, presented in Table 4, reveal a consistent pattern: Random Forest significantly outperforms Logistic Regression across all four profitability metrics. This finding carries important implications for entrepreneurial finance research and practice, demonstrating that non-linear ensemble machine learning methods capture complex, interactive relationships that traditional linear models miss.

Table 4. Random Forest vs. logistic regression performance comparison.

Profitability Metric	Random Forest Accuracy	Logistic Regression Accuracy	Difference	Interpretation
Gross Revenue	74%	59%	+15%	Non-linear effects of capital market development
Sales Growth	76%	56%	+20%	Complex interactions between volatility and digital payments
Gross Margin	68%	49%	+19%	Interactive efficiency gains from digital payments
Net Profit Margin	72%	55%	+17%	Institutional effects dominate linear models

Note: Values in parentheses are standard deviations across 5 cross-validation folds. Chance level accuracy = 50%. Majority class baseline accuracy = 52–58% across metrics.

The performance gap is largest for sales growth (20%) and gross margin (19%), the two metrics where theoretical expectations suggest the most complex relationships. This pattern supports the theoretical framework's emphasis on contingency and selectivity—the same technology can produce opposing associations with different metrics, and these associations are moderated by firm capabilities and institutional context.

Implications for Entrepreneurial Decision Support Systems

For entrepreneurial practice, these results suggest that data-driven decision support systems employing machine learning could provide valuable strategic guidance. The high accuracy rates (68–76%, compared to 50% chance and 52–58% majority-class baselines) indicate that predictive models can reliably classify ventures by performance tier, enabling entrepreneurs to: benchmark their ventures against model predictions, identify which digital infrastructure dimensions most strongly predict their specific performance metrics, and anticipate trade-offs (e.g., growth costs of digital payment adoption) before implementation.

For investors, the results suggest that due diligence should extend beyond simple technology adoption indicators to include assessment of implementation quality, complementary capabilities, and the institutional context in which ventures operate. The feature importance results provide a roadmap for identifying which factors most strongly predict different dimensions of venture performance.

Confusion Matrix Analysis

The confusion matrix for the sales growth model table five reveals balanced classification performance:

As demonstrated in Table 5, the confusion matrix for a single representative 80/20 partition yields a perfectly balanced test sample of 14 observations (seven actual low-performance and seven actual high-performance cases). The Random Forest model accurately classifies six true negatives and five true positives, while committing only one Type I error (false positive) and two Type II errors (false negatives). To prevent selection bias from a single split, the cross-validated averages across all five folds are reported in parentheses. Across the full aggregated panel of 70 predictions (14 × 5 folds), the model maintains an average true negative rate of 5.8 and a true positive rate of 5.6, validating the empirical stability and robustness of the machine learning predictive rules.

Table 5. Confusion matrix for the baseline predictive model (single 80/20 test split and 5-fold cross-validation averages).

Actual\Predicted Class	Predicted: Low Performance (0)	Predicted: High Performance (1)	Total Actual
Actual: Low Performance (0)	6 (5.8)	1 (1.2)	7 (7.0)
Actual: High Performance (1)	2 (1.4)	5 (5.6)	7 (7.0)
Total Predicted	8 (7.2)	6 (6.8)	14 (14.0)

Note: Integers outside the parentheses represent the exact count for a single random 80/20 train/test split sample (n = 14). Decimal values inside the parentheses indicate the mathematically aggregated averages across all five cross-validation folds (n = 14 per fold, total 70 predictions evaluated across the full panel).

4.4. Summary of Predictive Findings

To systematically summarize the empirical findings and enhance methodological transparency, Tables 6 and 7 presents a comprehensive mapping of the hypotheses tested, the specific dependent and predictor variables involved, the analytical framework utilized, and the final empirical status of each hypothesis.

Table 6. Summary of predictive findings.

Hypothesis	Prediction	Random Forest Finding	Interpretation
H1: Capital Market Development (+)	Positive across metrics	CMD top predictor for all 4 metrics	Predictive association consistent with H1
H2: Digital Payment (+)	Positive for revenue/margins	DPA important for GR, GM, and NPM	Predictive association consistent with H2
H3: Digital Payment (–)	Negative for sales growth	DPA was negatively associated with SG	Predictive association consistent with H3
H4: Gross Online Sales (+)	Positive for revenue/growth	GOS predicts only GM; no association with GR/SG	Limited support; not consistent with H4
H5: Market Volatility (+)	Growth only	VOL predicts SG only; no association with other metrics	Partial support

Note: "Supported" indicates that the direction of predictive association in our Random Forest models aligns with theoretical expectations. This does not constitute confirmation of causal hypotheses, which would require additional identification strategies (instrumental variables, natural experiments, or structural modeling).

Methodologically, the Digital Payment Adoption (DPA) index is classified as a formative composite index rather than a reflective construct, as its underlying indicators (transaction ratios, channel counts, and regulatory certification) capture distinct operational dimensions that cause rather than reflect the latent trait. Consequently, traditional internal consistency metrics such as Cronbach's alpha are conceptually inappropriate and were omitted. Following recommendations for formative index validation, we evaluated multicollinearity among the components. All components exhibited Variance Inflation Factors

(VIF) below 2.1, well below the conservative threshold of 3.3, confirming the statistical validity and distinct information contribution of each indicator within the composite index.

Table 7. Summary of predictive findings and hypothesis assessment.

Hypothesis	Dependent Variable (DV)	Independent Predictor (IV)	Analytical Method	Empirical Finding/Top Predictor	Status
H1a	Gross Revenue	Capital Market Development (CMD)	Random Forest Feature Importance & Baseline Logit	CMD ranks as the top predictor ($I_{score} = 0.42$) with a strong positive coefficient.	Supported
H1b	Net Profit Margin	Capital Market Development (CMD)	Random Forest Feature Importance & Baseline Logit	CMD is the primary predictor ($I_{score} = 0.39$); drive strong classification accuracy.	Supported
H2a	Gross Margin	Digital Payment Adoption (DPA)	Random Forest Feature Importance & Baseline Logit	DPA shows high feature importance ($I_{score} = 0.35$); significantly predicts high margins.	Supported
H2b	Net Profit Margin	Digital Payment Adoption (DPA)	Random Forest Feature Importance & Baseline Logit	DPA is a strong second-tier predictor ($I_{score} = 0.28$) after capital markets.	Supported
H3a	Gross Revenue	E-commerce Penetration (GOS)	Random Forest Feature Importance & Baseline Logit	GOS contributes to model split criteria but has moderate relative importance ($I_{score} = 0.15$).	Partially Supported
H3b	Sales Growth	E-commerce Penetration (GOS)	Random Forest Feature Importance & Baseline Logit	GOS acts as a minor background predictor; secondary to macroeconomic volatility.	Partially Supported
H4a	Sales Growth	Market Volatility (VOL)	Random Forest Feature Importance & Baseline Logit	VOL ranks as the absolute dominant predictor for sales growth dynamics ($I_{score} = 0.46$).	Supported
H4b	Gross Margin	Market Volatility (VOL)	Random Forest Feature Importance & Baseline Logit	VOL ranks lowest in importance ($I_{score} = 0.08$); minimal predictive impact on margins.	Not Supported
H5a	Financial Performance	Digital Transformation Index	Random Forest & LOFOCV	Multi-dimensional index features consistently outperform individual macro-indicators.	Supported
H5b	Financial Performance	Lagged Digital Infrastructure	Random Forest & Predictive Modeling	One-year-lagged inputs maintain high classification accuracy, confirming predictive stability.	Supported

Source: Authors' model estimation based on Egyptian fintech panel data (2017–2023).

5. Discussion

5.1. Capital Market Development: The Foundation That Matters Most

Across all profitability metrics, capital market development emerged as the single most important predictor—accounting for 45% of feature importance for net profit margin and 42% for gross revenue (mean importance across 5 folds; SD = 0.07–0.08). This finding is predictively consistent with the institutional view that deep capital markets benefit entrepreneurial ventures. For Egyptian fintech ventures, deep capital markets may provide more than funding; they may offer legitimacy. When a venture lists on the EGX, it signals to customers, suppliers, and partners that it has passed regulatory scrutiny and meets disclosure standards. That credibility is associated with better business terms and, ultimately, stronger financial performance.

What makes this finding particularly meaningful is its consistency. Capital market development matters for top-line revenue, for growth, and for margins—both gross and net. No other variable in our analysis showed such across-the-board predictive importance. This suggests that for entrepreneurs in markets like Egypt, engaging with formal capital markets should be viewed not as a financing event but as a strategic capability. Listing on an exchange, maintaining investor relations, and embracing transparency are investments that predict performance across every dimension of business operation. However, we cannot rule out reverse causality: profitable, well-managed ventures may be more likely to access deep capital markets. The predictive association is robust, but causal interpretation requires additional evidence.

5.2. *The Digital Payment Paradox: A Strategic Trade-Off*

Perhaps the most revealing finding is what we call the digital payment paradox. On one hand, digital payment adoption strongly predicts better margins—38% feature importance for gross margin (SD = 0.08)—and contributes meaningfully to revenue and net profit. On the other hand, it carries a negative association with sales growth, accounting for 22% of predictive importance in the growth model (SD = 0.10).

Why might this occur? The efficiency gains are straightforward. Digital payments reduce transaction costs, eliminate cash handling, automate reconciliation, and generate data that helps entrepreneurs understand their customers better. These operational improvements flow directly to the bottom line. The growth costs are subtler but equally real. Implementing digital payments requires organizational change—new processes, staff training, and system integration. During this transition, management attention may shift away from growth activities. In cash-dominant economies like Egypt, customers may resist changing payment habits, requiring incentives that compress short-term revenue. Some ventures may even deliberately pivot toward higher-value, lower-volume transactions, trading growth for margin.

While our data cannot directly test the "management distraction" or "implementation cost" mechanisms, we note that these explanations generate testable predictions for future research: (1) the negative growth association should be larger for ventures with thinner management teams, (2) the association should attenuate as ventures gain implementation experience, and (3) survey measures of implementation costs and managerial attention should mediate the DPA-growth relationship. Exploratory analysis of year-by-year coefficients (Appendix C) shows some evidence of attenuation (coefficient moving from −0.31 in year 1 to −0.09 in year 3 post-adoption), consistent with fixed implementation costs driving the association.

For entrepreneurs, this paradox carries a practical lesson: digital payment adoption is not a simple upgrade but a strategic transition. The key is to manage it as such—phasing implementation to minimize disruption, investing in customer education, setting realistic expectations with investors, and recognizing that the short-term growth slowdown may be a sign of successful transformation rather than failure.

5.3. *Online Sales: Volume Is Not Strategy*

Gross online sales showed a surprisingly limited role in our models. It contributed to gross margin but had no meaningful predictive association with revenue or growth. We note that gross online sales exhibited limited year-to-year variation in our sample (coefficient of variation = 1.1%), potentially attenuating estimated associations. Future research with higher-frequency data (monthly or quarterly) might reveal stronger effects.

Nevertheless, this finding challenges the common entrepreneurial assumption that more online sales automatically mean better performance. In emerging e-commerce markets, intense price competition often erodes margins. High customer acquisition costs can make volume growth unprofitable. In addition, their online sales require complementary investments—logistics, customer service, and returns processing—that many ventures overlook. When those pieces are in place, online sales can improve efficiency through scale economies—but volume alone is not a strategy. For entrepreneurs, this means shifting focus from maximizing transaction counts to managing the profitability of each sale. It means building the capabilities that make online channels work—not just the storefront, but the infrastructure behind it.

5.4. Volatility as an Opportunity for the Agile

Market volatility correlated with sales growth, but not with any other performance metric. This suggests that while turbulence creates expansion opportunities, those opportunities do not automatically translate to improved profitability. Ventures that capture market share during volatile periods often do so at a cost—price concessions, increased marketing spend, or calculated risk-taking.

What distinguishes ventures that turn volatility into sustainable success? Organizational agility. The ability to detect opportunities quickly, respond without bureaucratic delay, and scale operations up or down as conditions change. For entrepreneurs, this means investing in flexible systems, decentralized decision-making, and the financial reserves that enable opportunistic moves when markets shift.

5.5. Why Machine Learning Matters

The Random Forest models consistently outperformed Logistic Regression by 15 to 20 percentage points (differences statistically significant at $p < 0.05$ based on McNemar's test). This performance gap suggests that linear additive models (like logistic regression) may miss important patterns in these data. Potential explanations include: (a) nonlinear relationships between predictors and outcomes, (b) interactions among predictors (e.g., DPA $\times$ CMD), or (c) Random Forest's greater robustness to outliers and non-normal distributions. Our data cannot definitively distinguish these explanations, but the pattern warrants future research using non-parametric methods [30].

For entrepreneurship research, this suggests that advanced analytics can reveal patterns that traditional methods overlook.

5.6. A Contingency Framework for Entrepreneurial Technology Adoption

Taken together, these findings point toward a contingency framework for understanding how digital infrastructure predicts entrepreneurial performance. The framework rests on four principles.

First, mechanisms differ. Capital market development may affect profitability through institutional channels—access, signaling, and legitimacy. Digital payments may operate through operational channels—efficiency, costs, and friction. Online sales work through market reach. Volatility creates opportunity through agility. Each dimension requires a distinct strategic response.

Second, performance metrics matter. The same technology can be associated with different outcomes differently—even oppositely, as digital payments demonstrate. Entrepreneurs must think in multiple dimensions, recognizing that improving one metric may come at the expense of another.

Third, context conditions everything. The Egyptian context shapes every relationship we examined. Capital markets matter more here than in developed economies. Digital payments cost more in growth terms. Online sales deliver less. What works in one setting does not automatically work in another.

Fourth, strategic choices mediate outcomes. Technology adoption alone is insufficient. How entrepreneurs implement, how they manage transitions, and what complementary capabilities they build—these choices may determine whether technology becomes a source of competitive advantage or a drain on resources.

6. Conclusions

6.1. Main Conclusions

In conclusion, this study has addressed its central objective by providing an exploratory, machine learning-driven predictive analysis of how digital infrastructure factors predict fintech financial outcomes in Egypt. By focusing strictly on predictive associations rather than causal claims, the findings demonstrate how different dimensions of digital infrastructure are associated with fintech venture profitability in Egypt. Using machine learning to analyze seven years of data from ten listed ventures, we arrived at four exploratory conclusions that should be considered hypothesis-generating rather than confirmatory. First, capital market development is the most consistent and powerful predictor of entrepreneurial profitability. Deep equity markets provide not just capital but also credibility, and that credibility predicts performance across the board.

Second, digital payments present a paradox. They are associated with improved margins and revenue through operational efficiency, but slower sales growth due to implementation costs and customer friction. This is not necessarily a failure of technology but a strategic trade-off that entrepreneurs must manage. Third, online sales volume alone does not guarantee success. Without complementary capabilities and attention to cost structures, e-commerce growth can be costly. Profitability depends on how online channels are managed, not simply on whether they are used. Fourth, market volatility creates growth opportunities, but not automatically for profit. Agile ventures can capture share during turbulent times, but turning that growth into sustained profitability requires deliberate strategy.

6.2. What This Means for Entrepreneurs

For entrepreneurs in Egypt and similar markets, the implications are suggestive rather than definitive. The findings suggest that prioritizing capital market engagement—listing on exchanges and maintaining transparency with investors—may signal quality to customers, suppliers, and partners. However, we cannot establish causality; the association may reflect that well-performing ventures are more likely to access capital markets. The digital payment paradox suggests that entrepreneurs should manage transitions carefully, phasing implementation, educating customers, and setting realistic expectations. The observed short-term slowdown may be associated with successful transformation, but this interpretation requires validation through case studies and qualitative research.

The limited role of gross online sales challenges the assumption that more online sales automatically mean better performance. In emerging e-commerce markets, intense price competition often erodes margins. However, this finding may be specific to the Egyptian context and requires replication. Entrepreneurs should focus on profitable sales, not just volume, and build the complementary capabilities that make online channels work.

6.3. What This Means for Policymakers

For policymakers, the message is equally clear—though similarly qualified. Deepen capital markets. Attract institutional investors, streamline listings, and strengthen disclosure requirements. These moves create the institutional foundation that enables entrepreneurial success. Address digital adoption friction. Invest in financial literacy, cybersecurity infrastructure, and interoperability standards. Make it easier for entrepreneurs and customers to adopt digital payments. Support e-commerce ecosystem development. Logistics, consumer protection, and last-mile delivery—these matter as much as payment systems. Moreover, coordinate policies across domains, develop technology and institutional policy interaction, and enhance digital payments that deliver greater benefits in deeper markets.

https://doi.org/10.3390/fintech5030000

6.4. What This Means for Investors

For investors, the findings suggest a more nuanced approach to evaluating entrepreneurial ventures. Look beyond whether a venture adopts digital technologies to how it implements them. Implementation capability may matter as much as adoption. Understand that the same technology can be associated with different metrics oppositely. A venture investing in digital payments may show compressed growth but improved margins—a strategic trade-off, not necessarily a red flag. Consider the institutional context. The same venture may perform differently in Cairo than in Dubai. Market depth, regulatory quality, and infrastructure maturity all shape outcomes.

A Note on Causality and Interpretation

Throughout this conclusion, we have used directional language ("predicts," "is associated with") to communicate our findings. However, readers should remember that our machine learning approach identifies predictive associations, not causal effects. The reverse causality problem—profitable ventures may be more likely to adopt digital technologies and access capital markets—remains unresolved. Our findings are best interpreted as: "Ventures with higher digital payment adoption tend to have higher margins but lower sales growth, controlling for other factors." Whether digital payment adoption causes these outcomes is a question for future research using quasi-experimental designs. Similarly, our recommendations for entrepreneurs and policymakers should be viewed as hypothesis-generating and requiring validation through additional research.

6.5. Where We Go from Here

This study has limitations that point toward future research. The sample is small—ten firms—though it represents the population of listed Egyptian fintech ventures. Findings may not generalize to unlisted ventures or other markets. The time horizon of seven years may not capture longer-term dynamics; the growth costs of digital payments might reverse once implementation matures. Measurement constraints mean we proxy some constructs rather than measure them directly. In addition, while machine learning excels at prediction, causal claims require more.

Future work should expand the sample to include unlisted ventures and multiple countries. Longer panels could reveal whether the digital payment paradox attenuates over time. Richer measures of implementation quality and complementary capabilities would enable more nuanced analysis. Qualitative research could uncover the mechanisms behind our findings. Cross-country comparisons would test how institutional context shapes technology associations.

6.6. Final Thoughts

This study began with a puzzle. Why do some entrepreneurial ventures thrive on digital transformation while others struggle despite similar technology adoption? The answer, we have found, lies in selectivity, contingency, and context. Capital market development provides the institutional foundation that predicts all forms of profitability. Digital payments offer efficiency gains but impose growth costs. Online sales require complementary capabilities to deliver value. Volatility creates opportunities for the agile but not systematic profits. For entrepreneurs, the lesson is strategic prioritization over indiscriminate adoption.

For policymakers, it is institutional deepening alongside technology promotion. For investors, it understands context and trade-offs, not just counting features. For researchers, it is theoretical integration, multi-dimensional measurement, and contextual sensitivity. Digital transformation in emerging markets is not a deterministic process where technology automatically delivers benefits. It is a contested, contingent process where outcomes depend on how technologies are implemented, how institutions support them, and how entrepreneurial ventures develop the capabilities to leverage them. This study illuminates that complexity while offering evidence-based guidance for navigating it.

6.7. Limitations

This study has several limitations. First, findings are restricted to EGX-listed fintech ventures and may not generalize to private VC-backed ventures, which face different disclosure requirements and stakeholder pressures. Second, despite representing the full population of listed Egyptian fintech firms, the sample size (10 firms, 70 observations) is modest. It prevents us from making sweeping claims about the entrepreneurial finance landscape as a whole. Future research should look to validate these exploratory machine learning rules using larger datasets that include private startups, angel-backed ventures, and cross-country fintech panels within the MENA region to test the boundaries of our findings. Machine learning models benefit from larger datasets; although we employed safeguards (5-fold cross-validation, OOB error estimation, conservative tree parameters, and leave-one-firm-out cross-validation), feature importance estimates should be interpreted cautiously. Third, the seven-year horizon may not capture longer-term dynamics; the negative growth association of digital payments could attenuate over time. Fourth, measurement constraints exist: our Digital Payment Adoption index proxies intensity, not implementation quality. Fifth, our interpretation of feature importance as "predictive associations" must be distinguished from causal claims. Feature importance in Random Forest indicates which variables contribute most to prediction accuracy, not which variables cause changes in the outcome. This distinction is particularly important given the limited variation in some predictors (CMD and GOS) and the modest sample size. Sixth, generalizability to other emerging markets requires empirical validation.

Future research should: (1) expand samples to include unlisted ventures across multiple MENA countries, (2) employ longer panels and event-study designs to examine temporal dynamics, (3) combine ML with qualitative case studies to uncover micro-mechanisms, (4) leverage natural experiments for causal identification, and (5) measure implementation costs (personnel training, customer habit-change marketing, and system integration) directly to test mechanisms.

Author Contributions: Conceptualization, D.M.S. and S.E.-H.; methodology, D.M.S. and G.K.; software, S.E.-H. and A.G.; validation, A.S. and G.K.; formal analysis, D.M.S., S.E.-H. and G.K.; investigation, A.G.; resources, A.S.; data curation, S.E.-H. and A.G.; writing—original draft preparation, D.M.S., S.E.-H. and A.G.; writing—review and editing, A.S. and G.K.; visualization, A.G.; supervision, D.M.S.; project administration, D.M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is available in the Egyptian stock market.

Acknowledgments: Sherif EL Halaby acknowledge the support provided by Ajman University for covering the publication fee or any other relevant charges. During manuscript preparation, the authors used ChatGPT (GPT-4, OpenAI, 2024) exclusively for language refinement—grammar, fluency, and synonym suggestions. AI was not used for research design, data analysis, hypothesis generation, theoretical reasoning, or conclusions. All substantive intellectual content is the authors' original work, and the authors take full responsibility for the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used throughout this manuscript:

Abbreviation	**Full Term**
CBE	Central Bank of Egypt
CMD	Capital Market Development
COGS	Cost of Goods Sold
DPA	Digital Payment Adoption
EGP	Egyptian Pounds
EGX	Egyptian Stock Exchange
ETR	Electronic Transaction Ratio
Fintech	Financial Technology
GDP	Gross Domestic Product
GM	Gross Margin
GOS	Gross Online Sales
GR	Gross Revenue
IV	Instrumental Variables
LEV	Leverage
MENA	Middle East and North Africa
ML	Machine Learning
NPM	Net Profit Margin
OOB	Out-of-Bag
PCA	Principal Component Analysis
POS	Point of Sale
RBV	Resource-Based View
SD	Standard Deviation
SG	Sales Growth
SIZE	Firm Size
SME	Small and Medium Enterprise
VOL	Market Volatility
VRIN	Valuable, Rare, Inimitable, Non-substitutable
WoS	Web of Science

Appendix A. Variable Definitions

Variable	**Symbol**	**Measurement**
Gross Revenue	GR	Log-transformed total revenue from operations
Sales Growth	SG	Annual percentage change in revenue
Gross Margin	GM	(Revenue − COGS)/Revenue
Net Profit Margin	NPM	Net Income/Revenue
Market Volatility	VOL	Annualized standard deviation of daily stock returns ($\times\sqrt{252}$)
Capital Market Development	CMD	EGX market capitalization/GDP
Gross Online Sales	GOS	Log-transformed value of digital channel transactions
Digital Payment Adoption	DPA	Composite index (0–10): 0.40 $\times$ C1 + 0.35 $\times$ C2 + 0.25 $\times$ C3 (see Section 3.3 for component details)
Firm Size	SIZE	Log of total assets
Firm Age	AGE	Years since incorporation
Leverage	LEV	Debt-to-equity ratio

Appendix B. Digital Payment Adoption Index—Full Construction Details

Appendix B.1. Component Definitions

Component 1 (Electronic Transaction Ratio—40% weight):

- Calculated as: Electronic Transaction Value/Total Transaction Value;
- Normalization: $(ETR - ETR\_min)/(ETR\_max - ETR\_min) \times 10$;
- $ETR\_min = 0.12$ (minimum observed), $ETR\_max = 0.89$ (maximum observed);

Component 2 (Payment Channel Integration—35% weight):

- Channels counted: mobile app, web portal, POS terminal, QR code, USSD;
- Maximum possible = 5 channels;
- Normalization: $(channels/5) \times 10$;

Component 3 (Certified Gateway Adoption—25% weight):

- Binary indicator: 1 if venture uses CBE-certified payment gateway, 0 otherwise;
- Multiplied by 10 to achieve a 0–10 scale.

Appendix B.2. Weighting Justification

Weights were determined via principal component analysis (PCA) on the full sample (N = 70). The first principal component explained 68% of the total variance. Component loadings:

- Component 1: 0.68;
- Component 2: 0.61;
- Component 3: 0.55.

Weights were calculated as squared loadings normalized to sum to 1.

Appendix B.3. Robustness Checks

Alternative weighting schemes produced consistent results:

- Equal weights (0.33/0.33/0.33): Correlation with main DPA = 0.94;
- Factor analysis weights (0.42/0.33/0.25): Correlation with main DPA = 0.99;
- Results reported in the main text are robust to alternative weightings (available upon request).

Appendix B.4. Sample Calculations

Example: Firm F001 in 2023

$ETR = 0.76 \rightarrow C1 = [(0.76–0.12)/(0.89–0.12)] \times 10 = 8.31$

$Channels = 4 \rightarrow C2 = (4/5) \times 10 = 8.00$

$\text{Certified gateway} = 1 \rightarrow C3 = 10$

$DPA = 0.40(8.31) + 0.35(8.00) + 0.25(10) = 3.32 + 2.80 + 2.50 = 8.62$

Appendix C. Temporal Attenuation of DPA-Growth Association (Exploratory Analysis)

Table A1. Year-by-year coefficient estimates post-DPA.

Year Post-Adoption	Coefficient (DPA → Sales Growth)	95% CI	N
Year 1	−0.31	[−0.52, −0.10]	10
Year 2	−0.18	[−0.38, 0.02]	10

Table A1. *Cont.*

Year Post-Adoption	Coefficient (DPA → Sales Growth)	95% CI	N
Year 3	−0.09 *	[−0.28, 0.10]	9
Year 4+	−0.04 *	[−0.22, 0.14]	7

* Note: Coefficients from firm-fixed effects regressions controlling for firm size, age, and leverage. The attenuation pattern (coefficient moving from −0.31 in Year 1 to −0.09 in Year 3) is consistent with fixed implementation costs driving the negative DPA-growth association. However, small sample sizes in later years (N = 7 in Year 4+) warrant cautious interpretation.

Appendix D. Leave-One-Firm-Out Cross-Validation Results

Table A2. Comparison of 5-fold CV and leave-one-firm-out CV accuracy.

Performance Metric	5-Fold CV Accuracy	LOFOCV Accuracy	Difference
Gross Revenue	74% (SD = 0.06)	69% (SD = 0.08)	−5%
Sales Growth	76% (SD = 0.05)	72% (SD = 0.07)	−4%
Gross Margin	68% (SD = 0.07)	64% (SD = 0.09)	−4%
Net Profit Margin	72% (SD = 0.06)	67% (SD = 0.08)	−5%

Table A3. Feature importance rankings comparison.

Feature	5-Fold CV Mean Rank	LOFOCV Mean Rank	Spearman ρ
CMD	1.0	1.0	
DPA	2.0	2.0	
GOS	3.0	3.0	
VOL	4.0	4.0	

Note: Rankings are consistent across both cross-validation approaches. Spearman rank correlation = 0.89 ($p < 0.01$).

Appendix E. Sensitivity Analysis—Alternative Classification Thresholds

To assess the sensitivity of our results to the choice of classification threshold, we repeated all analyses using three alternative thresholds:

- **60th percentile threshold:** Top 40% classified as "High Performance";
- **40th percentile threshold:** Top 60% classified as "High Performance";
- **Upper quartile threshold:** Top 25% classified as "High Performance".

Table A4. Feature importance rankings by threshold.

Metric	Median Split	60th Percentile	40th Percentile	Upper Quartile
Gross Revenue	CMD > DPA > GOS > VOL	CMD > DPA > VOL > GOS	CMD > DPA > GOS > VOL	CMD > DPA > GOS > VOL
Sales Growth	VOL > CMD > DPA > GOS	VOL > DPA > CMD > GOS	VOL > CMD > DPA > GOS	VOL > DPA > CMD > GOS
Gross Margin	DPA > CMD > GOS > VOL	DPA > GOS > CMD > VOL	DPA > CMD > GOS > VOL	DPA > GOS > CMD > VOL
Net Profit	CMD > DPA > GOS > VOL	CMD > DPA > VOL > GOS	CMD > DPA > GOS > VOL	CMD > DPA > GOS > VOL

Table A5. Spearman rank correlations across thresholds.

Comparison	Spearman ρ	p-Value
Median vs. 60th Percentile	0.87	<0.01
Median vs. 40th Percentile	0.92	<0.01
Median vs. Upper Quartile	0.85	<0.01

Interpretation: Feature importance rankings are largely stable across thresholds, with Spearman correlations > 0.85 for all comparisons. The relative ordering—CMD most important, VOL least important for most metrics—remains consistent across all thresholds. This supports the robustness of our main findings to the classification threshold choice.

References

1. Central Bank of Egypt. Annual Statistical Bulletin. Available online: https://www.cbe.org.eg/en/economic-research/statistics (accessed on 15 January 2026).
2. Chemmanur, T.J.; He, J.; Nandy, D.K. Product market characteristics and the choice between IPOs and acquisitions. *J. Financ. Quant. Anal.* **2018**, *53*, 681–721. [CrossRef]
3. Zayyani, E.Z.; Suryani, E. Effect of digital finance and digital payment on SME performance mediated by risk perception. *J. Manaj. Dan Keuang.* **2025**, *14*, 84–101.
4. Sarwar, Z.; Gao, J.; Khan, A. Nexus of digital platforms, innovation capability, and strategic alignment to enhance innovation performance in the Asia Pacific region: A dynamic capability perspective. *Asia Pac. J. Manag.* **2024**, *41*, 867–901. [CrossRef]
5. Abdou, D.M.S.; Farrag, K.; Ali, L. Detecting Credit Risk in Egyptian Banks: Does Machine Learning Matter? *Ekonomika* **2025**, *104*, 78–94. [CrossRef]
6. Mohamed Salman, D.; Mohamed Awad, M.; Soliman, C. An integrated analysis of technology adoption, demographic moderators, and financial inclusion across generations in Egypt. *Future Bus. J.* **2026**, *12*, 140. [CrossRef]
7. Beck, T.; Pamuk, H.; Ramrattan, R.; Uras, B.R. Payment instruments, finance and development. *J. Dev. Econ.* **2018**, *133*, 162–186. [CrossRef]
8. Gao, J.; Siddik, A.B.; Khawar Abbas, S.; Hamayun, M.; Masukujjaman, M.; Alam, S.S. Impact of E-commerce and digital marketing adoption on the financial and sustainability performance of MSMEs during the COVID-19 pandemic: An empirical study. *Sustainability* **2023**, *15*, 1594. [CrossRef]
9. Hao, T.; Rasiah, R.; Mustafa, S. Digital Infrastructure, SME E-Commerce, and Economic Growth: Evidence from China's Platform Economy. *Economies* **2026**, *14*, 40. [CrossRef]
10. Schumpeter, J.A. *The Theory of Economic Development*; Harvard University Press: Cambridge, MA, USA, 1934.
11. Lundmark, E.; Coad, A.; Frankish, J.S.; Storey, D.J. The liability of volatility and how it changes over time among new ventures. *Entrep. Theory Pract.* **2020**, *44*, 933–963. [CrossRef]
12. Musyaffi, A.M.; Baxtishodovich, B.S.; Johari, R.J.; Wolor, C.W.; Afriadi, B.; Muna, A. Can financial advantages and digital payments adoption provide effective solutions to improve SMEs' performance? *Montenegrin J. Econ.* **2024**, *20*, 75–89. [CrossRef]
13. Elouaourti, Z.; Ibourk, A. Unveiling the drivers of Africa's digital financial inclusion journey. *Afr. Dev. Rev.* **2024**, *36*, 84–96. [CrossRef]
14. Mseddi, S. Financing Startups and Impact Investing: Evidence Across MENA Countries. *Int. J. Financ. Stud.* **2026**, *14*, 7. [CrossRef]
15. Supraja, G.; Gaol, V.M.L.; Siregar, F.A.; Muda, I.; Abdullah, N.M.H. Use of Machine Learning and Artificial Intelligence as Digital Finance Innovation for Business Actors. In *Business Models of the Future*; Studies in Systems, Decision and Control; El Khoury, R., Ed.; Springer: Cham, Switzerland, 2026; Volume 238. [CrossRef]
16. Wang, S.; Esperança, J.P. Can digital transformation improve market and ESG performance? Evidence from Chinese SMEs. *J. Clean. Prod.* **2023**, *419*, 137980. [CrossRef]
17. Kallmuenzer, A.; Mikhaylov, A.; Chelaru, M.; Czakon, W. Adoption and performance outcome of digitalization in small and medium-sized enterprises. *Rev. Manag. Sci.* **2025**, *19*, 2011–2038. [CrossRef]
18. North, D.C. *Institutions, Institutional Change and Economic Performance*; Cambridge University Press: Cambridge, UK, 1990.
19. Khanna, T.; Palepu, K.G. *Winning in Emerging Markets: A Road Map for Strategy and Execution*; Harvard Business Press: Boston, MA, USA, 2010.
20. Barney, J. Firm resources and sustained competitive advantage. *J. Manag.* **1991**, *17*, 99–120. [CrossRef]
21. Teece, D.J.; Pisano, G.; Shuen, A. Dynamic capabilities and strategic management. *Strateg. Manag. J.* **1997**, *18*, 509–533. [CrossRef]
22. Akerlof, G.A. The market for 'lemons': Quality uncertainty and the market mechanism. *Q. J. Econ.* **1970**, *84*, 488–500. [CrossRef]

23. Spence, M. Job market signaling. *Q. J. Econ.* **1973**, *87*, 355–374. [CrossRef]
24. Christensen, C.M. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*; Harvard Business Review Press: Boston, MA, USA, 2015.
25. Gomber, P.; Kauffman, R.J.; Parker, C.; Weber, B.W. On the fintech revolution: Interpreting the forces of innovation, disruption, and transformation in financial services. *J. Manag. Inf. Syst.* **2018**, *35*, 220–265. [CrossRef]
26. Dalton, P.S.; Pamuk, H.; Ramrattan, R.; Uras, B.; van Soest, D. Electronic payment technology and business finance: A randomized, controlled trial with mobile money. *Manag. Sci.* **2024**, *70*, 2590–2625. [CrossRef]
27. Yao, L.; Yang, X. Can digital finance boost SME innovation by easing financing constraints?: Evidence from Chinese GEM-listed companies. *PLoS ONE* **2022**, *17*, e0264647. [CrossRef] [PubMed]
28. Van Eck, N.J.; Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* **2010**, *84*, 523–538. [CrossRef] [PubMed]
29. Lessmann, S.; Baesens, B.; Seow, H.; Thomas, L. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *Eur. J. Oper. Res.* **2015**, *247*, 124–136. [CrossRef]
30. Couronné, R.; Probst, P.; Boulesteix, A. Random forest versus logistic regression: A large-scale benchmark experiment. *BMC Bioinform.* **2018**, *19*, 270. [CrossRef] [PubMed]