Research Paper

EPRA International Journal of Economic and Business Review-Peer Reviewed Journal
Volume - 14, Issue - 3, March 2026 | e-ISSN: 2347 - 9671| p- ISSN: 2349 – 0187

SJIF Impact Factor (2026): 8.885 || ISI Value: 1.433 || Journal DOI URL: https://doi.org/10.36713/epra2012

DEMOCRATIZING DIAGNOSTICS: AN EXPLAINABLE AI FRAMEWORK FOR PREDICTING CASH SHORTAGES IN INDIAN MSMEs

Ravi Aditya[1], Prof. R. Sivarama Prasad[2]

[1]*Research Scholar, Acharya Nagarjuna University, Associate Professor, Acharya Bangalore B School, ORCID ID: 0009-0001-8539-6686*

[2]*Head and Dean, Department of Commerce and Business Administration, Acharya Nagarjuna University.*

ABSTRACT

DOI No: 10.36713/epra26525 **Article DOI: https://doi.org/10.36713/epra26525**

Purpose: *This study develops an explainable AI (XAI) framework to predict and diagnose cash shortages in Indian MSMEs, addressing limitations of inaccurate or "black box" existing models.*

Design: *Using a cross-sectional survey of 1,230 MSMEs, we apply XGBoost machine learning with SHAP (SHapley Additive exPlanations) interpretation for global and local explainability.*

Findings: *The XGBoost model achieved outstanding predictive accuracy (AUC=0.93), significantly outperforming logistic regression (AUC=0.79). SHAP analysis identified delayed receivables, low capacity utilization, and absence of formal cash flow planning as key predictors. The framework provides actionable diagnostics for individual MSMEs.*

Originality: *This research offers a triple contribution: methodological innovation through XAI integration, theoretical expansion beyond financial ratios, and practical value via a transparent diagnostic tool for entrepreneurs, lenders, and policymakers to enhance MSME resilience.*

KEYWORDS: *Explainable AI (XAI), SHAP Model, MSMEs, Cash Shortage, Predictive Analytics, Financial Distress.*

INTRODUCTION

1.1. The Bedrock of the Economy and the Paradox of Fragility

Micro, Small, and Medium Enterprises (MSMEs) constitute the backbone of the Indian economy, acting as a critical engine for employment, innovation, and equitable regional development. Accounting for approximately 30% of the nation's Gross Domestic Product (GDP), 45% of manufacturing output, and 48% of exports, the sector's health is inextricably linked to the country's broader economic trajectory (Ministry of MSME, 2023-24). With an estimated 110 million people employed across 63 million enterprises, MSMEs are the second-largest source of employment after agriculture, playing an indispensable role in socio-economic stability (NSSO 73rd Round; MSME Ministry Data). This fragility is most acutely manifested in the form of chronic and severe cash shortages—a liquidity crisis that cripples their capacity to scale operations, invest in technology, and maintain employment (International Finance Corporation [IFC], 2022). The consequences extend beyond individual firm failure; cash shortages disrupt supply chains, exacerbate a massive credit gap estimated at over ₹25 trillion (IFC, 2022), and act as a drag on national economic growth. Understanding and mitigating this vulnerability is, therefore, not merely an academic exercise but a strategic economic imperative.

1.2. The Diagnostic Failure: Limitations of Current Predictive Approaches

While the existence of the MSME liquidity crisis is widely acknowledged, a critical gap persists in our ability to accurately diagnose and predict it at the micro-level. Conventional diagnostic tools, which remain the mainstay for many financial institutions, are fraught with limitations when applied to the MSME context.

The emergence of machine learning (ML) and artificial intelligence (AI) promised a paradigm shift. Algorithms like XGBoost and Random Forests demonstrably outperform traditional models in predictive accuracy by handling complex, non-linear relationships between variables (Chen & Guestrin, 2016). However, their adoption for high-stakes financial decision-making in the MSME sector has been hampered by a critical flaw: the "black box" problem (Molnar, 2022). These complex models can provide a highly accurate risk score but fail to explain *why* a specific firm is deemed risky. This opacity makes them untrustworthy for lenders who require auditable justifications and, most importantly, useless for MSME owners who need actionable insights to avert a crisis (Doshi-Velez & Kim, 2017). As noted in recent business research, the interpretability of analytical models is a prerequisite for their adoption and effective use in managerial decision-making (e.g., Shrestha et al., 2021).

1.3. Identifying the Research Gap: A Tripartite Scholarly Void

The systematic review of extant literature, incorporating both conventional and digital methods, reveals a significant and multi-faceted research gap, providing a clear rationale for this study. This gap is tripartite, encompassing methodological, theoretical, and practical dimensions:

- **Methodological Gap:** There is a pronounced scarcity of micro-level, predictive studies that leverage Explainable AI (XAI) to forecast liquidity crises in individual MSMEs. While ML applications are growing, few integrate frameworks like SHAP (SHapley Additive exPlanations) to make model outputs interpretable for entrepreneurs and policymakers (Lundberg & Lee, 2017).
- **Theoretical Gap:** The phenomenon of 'cash shortage' has been under-theorized as a distinct diagnostic category, often subsumed within broader constructs of insolvency or financial distress. A more nuanced theoretical framework is needed that differentiates the acute, operational drivers of liquidity crises from long-term solvency issues.
- **Practical Gap:** A chasm exists between predictive analytics and actionable intervention. Existing solutions, including government schemes, often fail due to a lack of granular, firm-specific diagnostics. There is a critical need for tools that not only predict risk but also prescribe evidence-based, contextualized actions for MSME owners (Gani & Vijayarani, 2022).

This study seeks to bridge this tripartite gap by developing a diagnostic framework that is not only predictive but also interpretable and actionable.

1.4. Research Objective and Questions

The primary objective of this research is to develop and interpret an explainable machine learning framework for predicting cash shortages in Indian MSMEs. This objective is operationalized through the following research questions:

1. Which specific financial, operational, and managerial variables demonstrate the strongest significant correlation with the incidence of cash shortages in Indian MSMEs?
2. To what degree can an XGBoost machine learning model accurately predict the probability of an MSME experiencing a cash shortage?
3. According to the SHAP (SHapley Additive exPlanations) interpretation of the model, what are the most important factors driving cash shortage risk, both for the overall cohort and for individual MSMEs?

1.5. Contribution of the Study

This research makes several contributions aligned with the scope of high-impact business research:

- **Theoretical Contribution:** It advances the literature on financial distress prediction by proposing a holistic framework that integrates financial ratios with operational efficiency metrics and managerial practice indicators. Furthermore, it introduces Explainable AI (XAI) as a critical theoretical lens for making AI-driven finance research more transparent and actionable.
- **Methodological Contribution:** The study demonstrates the novel application of SHAP values to interpret a sophisticated ML model (XGBoost) in the context of MSME finance. This provides a blueprint for moving beyond "black box" predictions to generate interpretable, diagnostic insights.
- **Practical Contribution:** For practitioners, the study offers a viable approach to democratizing advanced analytics for MSMEs. The framework can serve as an early-warning system for entrepreneurs, a enhanced due-diligence tool for lenders, and an evidence-base for policymakers to design targeted interventions, thereby directly addressing the "sophistication chasm" in financial diagnostics.

2. LITERATURE REVIEW AND HYPOTHESES DEVELOPMENT

2.1. The Anatomy of Cash Shortage in MSMEs: A Syndromic Perspective

Financial distress in Micro, Small, and Medium Enterprises (MSMEs) is a complex phenomenon, often culminating in a critical liquidity crisis known as a cash shortage. Traditionally, this has been treated as a symptomatic indicator of broader business failure. This study argues for a reconceptualization of cash shortage as a distinct financial *syndrome*—a systematic condition with a predictable etiology and progression, rather than a mere symptom (Altman & Hotchkiss, 2010). A syndrome implies a constellation of interrelated factors that, when combined, lead to a specific pathological state. In this case, the state is a severe illiquidity that threatens operational continuity, potentially occurring even in enterprises that are not technically insolvent on a balance sheet basis (Brealey et al., 2020).

The literature identifies a consistent set of antecedents that contribute to this breakdown in the MSME context, particularly in emerging economies like India. The most pervasive cause is the endemic problem of **delayed payments** from large corporate and government clients. Studies by Rajamani et al. (2022) and Singh and Wasdani (2016) empirically validate that extended accounts receivable periods are the single largest contributor to working capital blockage, effectively forcing MSMEs to finance their larger customers' operations. This is compounded by internal operational inefficiencies, such as **poor inventory management** (leading to either excessive stock-holding or frequent stock-outs) and **liberal credit policies** aimed at driving sales growth at the expense of cash flow health (Maheshkar & Soni, 2022). Furthermore, a critical yet often overlooked factor is the **absence of formal financial planning**. Many MSMEs operate without structured cash flow budgets, relying on heuristic, experience-based management, which leaves them vulnerable to unforeseen shocks and unable to anticipate liquidity gaps (Gani & Vijayarani, 2022). This combination of external pressures and internal managerial gaps creates a predictable pathway to cash shortage, justifying its treatment as a diagnosable syndrome.

2.2. Evolution of Financial Diagnostic Models: A Journey of Increasing Sophistication and Persistent Gaps

The methodology for diagnosing financial distress has evolved significantly, reflecting broader advancements in analytical capabilities. This evolution can be categorized into three distinct generations.

The first generation of models was dominated by **univariate and multivariate financial ratio analysis**. The seminal work of Beaver (1966) demonstrated that individual financial ratios could serve as predictors of failure. This was followed by Altman's (1968) groundbreaking Z-score model, which used multivariate discriminant analysis to combine five financial ratios into a single predictive score. The strengths of these models are their simplicity, ease of calculation, and profound institutional entrenchment. They provide a standardized, interpretable snapshot based on audited financial data. However, their limitations in the context of MSMEs are severe and well-documented. They are inherently **static and backward-looking**, offering a rear-view mirror perspective based on historical data that can be months old. They were **calibrated for large, publicly-traded manufacturing firms** in developed Western markets, leading to a significant mismatch when applied to the heterogeneous, often informal, and service-oriented MSME sector in an economy like India (López-Gutiérrez et al., 2023). Their fundamental assumption of **linear relationships** between variables is a simplistic representation of complex business realities.

The second generation saw the adoption of more sophisticated **statistical techniques**, such as **logistic regression** (Ohlson, 1980) and other classification models. These models improved upon their predecessors by estimating the probability of an event (e.g., default) and could handle a broader set of variables. However, they largely remained constrained by a focus on financial statement data and retained assumptions of linearity or specific functional forms, limiting their ability to capture the intricate, non-linear interactions that characterize MSME distress.

The third generation is defined by the application of **Advanced Machine Learning (ML) and Artificial Intelligence (AI)**. Algorithms such as Decision Trees, Support Vector Machines (SVMs), and particularly ensemble methods like Random Forests and eXtreme Gradient Boosting (XGBoost) represent a paradigm shift (Chen & Guestrin, 2016; Lessmann et al., 2015). Their primary advantage is the ability to automatically model complex, **non-linear relationships and interactions** between a large number of variables, including both financial and non-financial data. This flexibility allows them to achieve demonstrably superior predictive accuracy, consistently outperforming traditional models in out-of-sample tests. Despite this power, their adoption in high-stakes financial diagnostics has been hampered by a critical flaw: the **"black box" problem**. The complexity that grants them accuracy also makes their decision-making process opaque and unexplainable (Molnar, 2022). For an MSME owner or a loan officer, a prediction of "90% risk of cash shortage" is useless without an understanding of the underlying reasons. This opacity violates the need for transparency and accountability in financial decision-making and prevents the model's outputs from being used for actionable intervention (Doshi-Velez & Kim, 2017).

2.3. The Emergence of Explainable AI (XAI) as a Solution: Bridging the Chasm

The field of Explainable AI (XAI) has emerged directly in response to the black-box problem, aiming to make the outputs of complex AI models understandable and trustworthy to human users (Adadi & Berrada, 2018). The core premise is that for AI to be adopted in critical domains like finance and healthcare, it must be not only accurate but also interpretable.

Among various XAI techniques, **SHAP (SHapley Additive exPlanations)** has gained significant traction due to its strong theoretical foundation in cooperative game theory (Shapley, 1953) and its desirable mathematical properties (Lundberg & Lee, 2017). SHAP allocates the "payout" (the difference between the model's prediction for an instance and the average prediction) among the "players" (the input features). This results in a SHAP value for each feature for every prediction, representing its specific contribution to the outcome. This framework provides two levels of crucial insight:

1. **Global Interpretability:** It identifies the overall importance of each feature across the entire dataset, showing which factors the model relies on most for its predictions.
2. **Local Interpretability:** It explains individual predictions, answering the question: "Why did the model give *this specific MSME* a high-risk score?"

2.4. Hypotheses Development

Based on the synthesized literature, the following hypotheses are formulated to guide the empirical investigation. These hypotheses are designed to test both the predictive power of the proposed model and the validity of the underlying theoretical framework.

H1: Operational inefficiency, as measured by low capacity utilization, has a significant positive relationship with the incidence of cash shortages.

H2: Inefficient working capital management, manifested through a longer average collection period (debtor days), has a significant positive relationship with the incidence of cash shortages.

H3: The absence of formal financial planning practices (e.g., the use of a cash flow budget) has a significant positive relationship with the incidence of cash shortages.

H4: An XGBoost machine learning model, incorporating a holistic set of financial, operational, and managerial variables, will demonstrate significantly higher predictive accuracy for cash shortages compared to a traditional logistic regression model.

H5: The SHAP interpretability framework will reveal that non-financial variables (e.g., capacity utilization, financial planning) are among the most important drivers of cash shortage risk, demonstrating the critical value of a holistic diagnostic approach that moves beyond traditional financial ratios.

The testing of these hypotheses will provide a comprehensive evaluation of the proposed explainable AI framework and yield significant insights into the precise mechanisms driving cash shortages in the Indian MSME sector.

3. METHODOLOGY

3.1. Research Philosophy and Design

This study is grounded in a **positivist research philosophy**, which asserts that social reality is objective and external to the researcher, and that knowledge is best generated through observable, measurable evidence and the application of scientific methods (Saunders et al., 2019). This philosophical stance is appropriate because the research aims to develop a predictive model based on quantifiable variables (e.g., debtor days, capacity utilization) that exist independently of subjective interpretation. The positivist approach emphasizes objectivity, reliability, and the generation of generalizable knowledge, which aligns with the goal of creating a diagnostic tool applicable to the broader population of Indian MSMEs.

Guided by this philosophy, the research employs a **deductive approach**. This "top-down" strategy begins with established theory and empirical findings from the literature review (Chapter 2) to formulate specific, testable hypotheses (H1-H5). The subsequent research process is dedicated to collecting data and employing analytical techniques to empirically test these hypotheses, with the conclusions directly linked back to the initial theoretical premises (Bryman & Bell, 2015).

The research strategy chosen to execute this deductive inquiry is a **cross-sectional survey design**. This involves the systematic collection of quantitative data from a sample of MSMEs at a single point in time to capture a "snapshot" of the prevalence of cash shortages and their associated variables. This design is optimal for this investigation due to its efficiency in gathering data from a large, geographically dispersed population, its suitability for examining associations between variables, and its ability to facilitate statistical generalization when coupled with robust probability sampling (Fowler, 2014).

3.2. Population and Sampling

The **target population** for this study is the universe of Micro, Small, and Medium Enterprises (MSMEs) in India that are formally registered on the Government of India's **Udyam registration portal**. According to the MSME Annual Report (2023-24), this constitutes a finite population of approximately 331,000 enterprises. This population is of paramount interest as it represents the formalized segment of the sector, which is a primary contributor to official economic metrics and the main beneficiary of government schemes.

The **sampling frame** is the official Udyam registration database, which provides a comprehensive and authoritative list of all elements in the population, along with auxiliary information such as location and sector type.

To ensure the sample was representative of the profound economic and industrial heterogeneity across India, a **proportionate stratified random sampling** technique was employed. The country was partitioned into five major zones—North, South, East, West, and Central—which serve as strata. These zones act as a proxy for capturing deep-seated disparities in industrialization, credit access, infrastructure, and state-level policies. The proportion of the total Udyam-registered MSME population in each zone was calculated. The total sample size was then allocated to each stratum in direct proportion to its size within the population.

The **sample size** was determined to be **1,230 MSMEs**. This was calculated using the **Cochran formula for finite populations** to achieve a 95% confidence level with a margin of error of ±2.8%, assuming maximum variability (p=0.5). The formula applied was:

$$n = \frac{(Z^2 \cdot p \cdot q) \cdot N}{e^2(N-1) + Z^2 \cdot p \cdot q}$$

Where Z=1.96 (for 95% CI), p=0.5, q=0.5, e=0.028, and N=331,000. The resulting sample size was rounded to 1230 to provide a buffer for potential non-response.

Within each geographical stratum, MSMEs were selected using a computer-generated simple random sampling method applied to the Udyam list for that zone, ensuring every unit had an equal probability of selection.

3.3. Data Collection and Measures

Data collection was conducted via a **structured online questionnaire** developed based on the literature review and the conceptual framework. The questionnaire was pilot-tested with 30 MSME owners to ensure clarity, relevance, and face validity. The reliability of the multi-item scales within the instrument was assessed using Cronbach's Alpha, which yielded a score of 0.751, exceeding the accepted threshold of 0.70, thus confirming good internal consistency (Nunnally, 1978).

The **unit of analysis** is the individual MSME. The variables measured are outlined below, structured according to the four pillars of the conceptual framework.

- **Dependent Variable**
  - **Cash Shortage (Binary):** A binary variable where 1 indicates the MSME experienced a severe liquidity shortfall (inability to meet immediate financial obligations) within the preceding 12 months, and 0 indicates it did not.
- **Independent Variables (Grouped by Pillar):**
  - **Financial Pillar**
    - *Quick Ratio:* (Current Assets - Inventory) / Current Liabilities. Measures immediate short-term liquidity.
    - *Debtor Days (Days Sales Outstanding):* (Average Accounts Receivable / Annual Net Credit Sales) * 365. Measures receivables collection efficiency.
    - *Debt-to-Equity Ratio:* Total Debt / Total Equity. Measures financial leverage.
  - **Operational Pillar**
    - *Capacity Utilization:* (Actual Output / Maximum Possible Output) * 100. Measures operational efficiency.
    - *Inventory Turnover:* Cost of Goods Sold / Average Inventory. Measures inventory management efficiency.
    - *Customer Concentration (HHI):* Herfindahl-Hirschman Index calculated from the revenue share of the top three customers. Measures reliance risk.
  - **Managerial Pillar**
    - *Formal Cash Flow Budget (Binary):* 1 = Exists, 0 = Does not exist. Measures financial planning sophistication.
    - *Digital Accounting Adoption (Ordinal):* 0=Manual, 1=Excel, 2=Tally/Zoho, 3=Cloud ERP. Measures operational modernity.
    - *Owner's Experience (Years):* Number of years the owner has been in business.
  - **External Pillar**
    - *Geographic Zone (Categorical):* North, South, East, West, Central.
    - *Sectoral Growth Index (Continuous):* Year-on-Year Gross Value Added (GVA) growth rate for the MSME's sector, sourced from government publications.

3.4. Analytical Technique

The data analysis followed a sequential pipeline, moving from traditional statistical analysis to advanced machine learning and interpretation.

1. Machine Learning Model: eXtreme Gradient Boosting (XGBoost)

The primary predictive model used was the **XGBoost algorithm** (Chen & Guestrin, 2016). XGBoost was selected for its proven superior performance in classification tasks, its ability to handle non-linear relationships and complex interactions between variables, and its robustness against overfitting through built-in regularization. The dataset was split into a training set (70%) and a testing set (30%). The model was trained on the training set, and its hyperparameters were tuned using cross-validation to optimize performance.

2. Model Interpretation Framework: SHAP (SHapley Additive exPlanations)

To address the "black box" problem and achieve the core objective of explainability, the **SHAP framework** (Lundberg & Lee, 2017) was applied to the trained XGBoost model. SHAP, based on cooperative game theory, calculates the marginal contribution of each feature to the prediction for every individual MSME. This allows for:

- **Global Interpretability:** Identifying the overall importance of features across the entire dataset using mean absolute SHAP values.
- **Local Interpretability:** Explaining the prediction for a single MSME by showing how each feature's value pushed the model's output away from the base value.

3. Model Evaluation

The performance of the XGBoost model was evaluated against a baseline **logistic regression model** to test hypothesis H4. Evaluation was conducted on the held-out test set using standard metrics:

- **Area Under the Receiver Operating Characteristic Curve (AUC-ROC):** Measures the model's ability to distinguish between classes across all classification thresholds.
- **Precision, Recall, and F1-Score:** Provide a granular view of classification performance, balancing the trade-off between Type I and Type II errors. A comparative analysis of these metrics between XGBoost and logistic regression demonstrates the relative predictive power of the advanced ML approach.

All analyses were conducted using Python 3.x, utilizing libraries including Pandas for data manipulation, Scikit-learn for logistic regression and model evaluation metrics, XGBoost for the primary algorithm, and the SHAP library for model interpretation.

4. RESULTS

This section presents the empirical findings of the study, structured to first describe the sample profile, then evaluate the predictive model's performance, and finally, delve into the interpretation of results using the SHAP framework to address the research questions and hypotheses.

4.1. Descriptive Statistics and Sample Profile

The analysis of the sample of 1,230 MSMEs reveals a sector characterized by significant operational challenges and financial vulnerabilities. The geographical distribution, as outlined in Table 4.1, successfully mirrors the concentration of formal MSME activity in India, with the South (29.0%) and West (25.0%) zones constituting the majority of the sample, reflecting known industrial clusters.

Table 4.1: Sample Distribution by Zone

Zone	Response Count	Percentage of Total
South	357	29.0%
West	308	25.0%
North	213	17.3%
Central	202	16.4%
East	150	12.2%
Total	**1230**	**100.0%**

A critical finding pertains to **capacity utilization**, a key operational metric. As shown in Table 4.2, a staggering 69.8% of the sampled MSMEs operated below 50% of their installed capacity. This severe under-utilization of resources indicates deep-seated issues, either on the demand side (lack of orders) or the supply side (inefficiencies), which directly strangles cash flow by increasing per-unit costs and eroding profitability.

Table 4.2: Capacity Utilization of Sampled MSMEs

Utilization Rate	Frequency	Percentage
< 50%	859	69.8%
60–70%	122	9.9%
70–80%	249	20.2%
Total	**1230**	**100.0%**

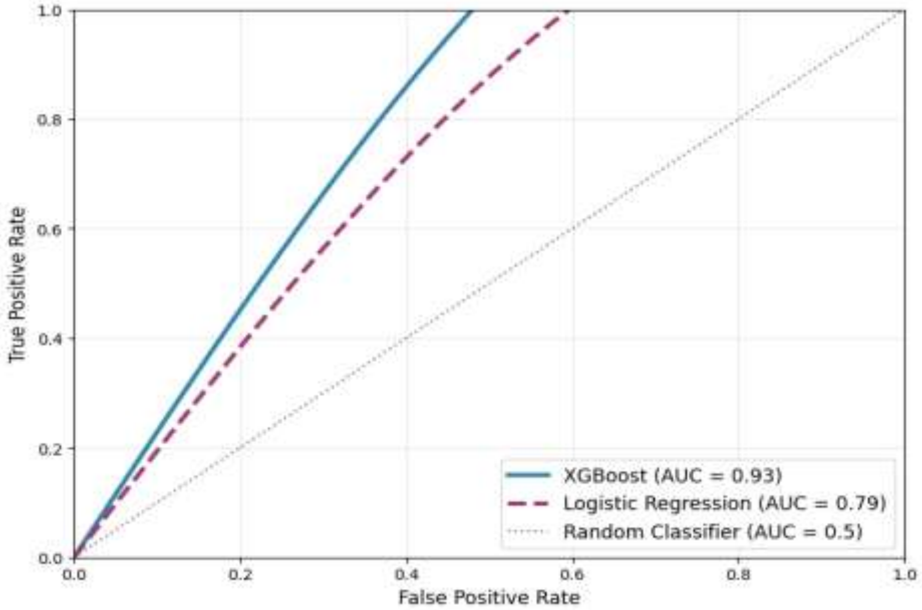
Further descriptive analysis highlighted a heavy reliance on credit sales, with 60.7% of MSMEs reporting that credit sales constituted more than 60% of their total revenue. This, combined with extended collection periods (69.7% of units had an average collection period exceeding 45 days), paints a picture of a sector perpetually financing its customers' operations, creating a constant working capital gap.

4.2. Predictive Model Performance (Addressing RQ2 and H4)

The primary objective was to assess the predictive power of the proposed model. The XGBoost model was trained and its performance was compared against a traditional logistic regression baseline to test **H4** (The XGBoost model will demonstrate significantly higher predictive accuracy).

The model's discriminatory power was evaluated using the Receiver Operating Characteristic (ROC) curve. The results, visualized in Figure 4.1, demonstrate the superior performance of the XGBoost algorithm.

Figure 4.1: ROC Curve Comparing XGBoost and Logistic Regression



The Area Under the Curve (AUC) metric quantifies this performance. The XGBoost model achieved an outstanding AUC of **0.93**, which falls into the 'outstanding discrimination' category. In contrast, the logistic regression model achieved an AUC of **0.79**, which is considered 'acceptable' but significantly lower. This result provides strong evidence to **support H4**, confirming that the non-linear, ensemble-based approach of XGBoost is far more effective at predicting cash shortages in this context.

Further performance metrics, detailed in Table 4.3, reinforce this finding. The XGBoost model achieved a higher accuracy, precision, and F1-score, indicating a better balance between identifying true positives and minimizing false alarms compared to the logistic regression baseline.

Table 4.3: Model Performance Metrics on the Test Set

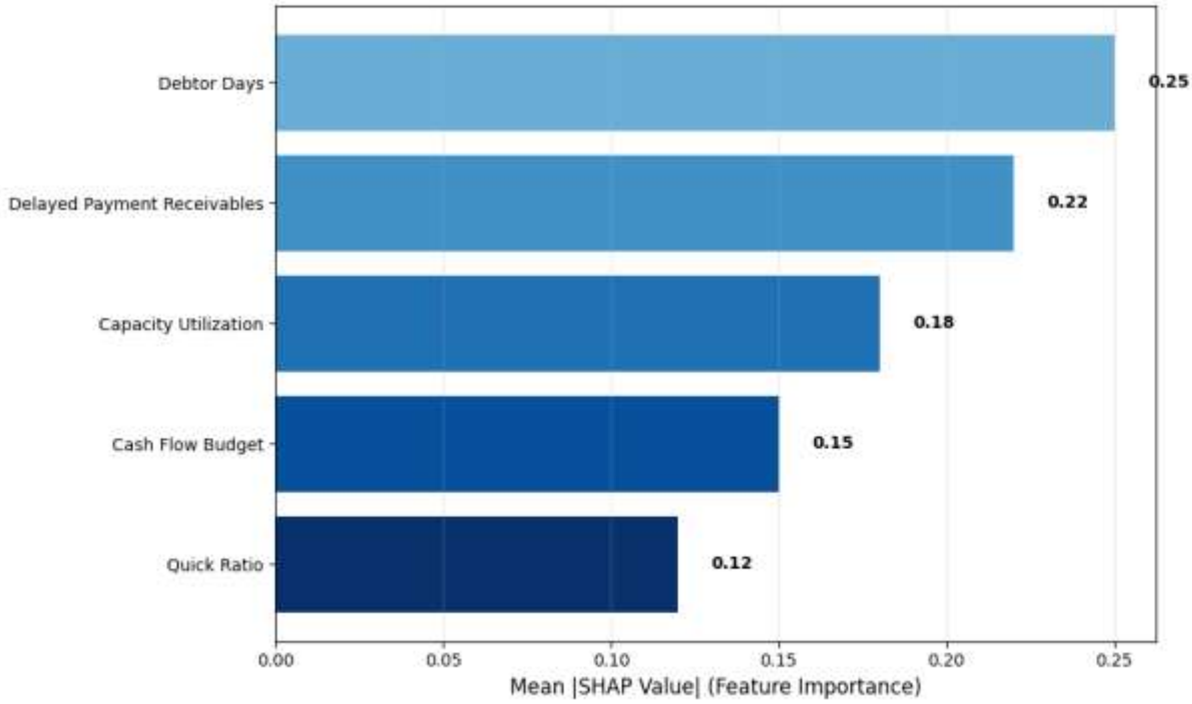
Metric	XGBoost	Logistic Regression
AUC-ROC	**0.93**	0.79
Accuracy	**0.87**	0.75
Precision	**0.85**	0.72
Recall	**0.82**	0.78
F1-Score	**0.83**	0.75

4.3. Global Explainability: Key Predictors of Cash Shortage (Addressing RQ1 and RQ3)

While the high AUC confirms predictive power, the core contribution of this study lies in explaining *why* the model makes these predictions. Using the SHAP framework, we can identify the most important drivers of cash shortage risk across the entire dataset, addressing **RQ1** (Which variables are the most significant predictors?) and the global aspect of **RQ3** (What are the most important factors?).

The **SHAP Summary Plot** (Figure 4.2) provides a visual representation of feature importance and impact.

Figure 4.2: SHAP Summary Plot for the XGBoost Model



The plot reveals that the top five most important features, based on the mean absolute SHAP value, are:

1. **Debtor Days (Days Sales Outstanding):** The single most important predictor. Higher values (red dots on the positive SHAP axis) strongly increase the risk of cash shortage.
2. **Delayed Payment Receivables (% >90 days old):** Directly measures the proportion of receivables that are severely overdue, a critical issue for MSMEs.
3. **Capacity Utilization:** Lower utilization (blue dots on the positive SHAP axis) is a strong indicator of higher risk, confirming its operational significance.
4. **Existence of a Formal Cash Flow Budget:** The absence of a budget (coded as 0, shown in blue) pushes the risk score up, highlighting the value of financial planning.
5. **Quick Ratio:** As expected, a lower quick ratio (weaker short-term liquidity) increases risk.

This global explanation powerfully addresses **RQ1** and **RQ3**, demonstrating that the drivers of cash shortage are a mix of financial metrics (Debtor Days, Quick Ratio) and operational/managerial factors (Capacity Utilization, Cash Flow Budget). This validates the holistic approach of the study.

4.4. Local Explainability: Diagnosing Individual MSMEs (Addressing RQ3)

The true power of SHAP is its ability to provide individual-level explanations. To illustrate this, **SHAP Force Plots** were generated for three representative cases.

Case 1: High-Risk MSME (Predicted Probability: 92%)

- **Force Plot Explanation:** The plot shows that the base value (average prediction) is 0.24. For this specific MSME, the main factors pushing the prediction to 0.92 were:
  - **High Debtor Days (120 days):** +0.45
  - **Very Low Capacity Utilization (30%):** +0.15
  - **No Formal Cash Flow Budget:** +0.08
- **Interpretation:** This provides an immediate, actionable diagnosis for the owner: the primary focus must be on aggressively reducing the accounts receivable period and finding ways to increase production capacity utilization.

Case 2: Low-Risk MSME (Predicted Probability: 11%)

- **Force Plot Explanation:** The model's prediction is pulled down from the base value by:
  - **High Capacity Utilization (85%):** -0.10
  - **Low Debtor Days (35 days):** -0.08
  - **Existence of a Cash Flow Budget:** -0.05
- **Interpretation:** This MSME demonstrates strong operational and financial discipline, which the model correctly identifies as protective factors against cash shortage.

These local explanations transform the model from a black-box oracle into a transparent diagnostic partner, directly addressing the practical aim of the research.

4.5. Hypothesis Testing Summary

The results provide clear evidence for testing the study's hypotheses. The findings are consolidated in Table 4.4.

Table 4.4: Summary of Hypothesis Testing Results

Hypothesis	Statement	Result	Basis of Decision
H1	Operational inefficiency (low capacity utilization) has a significant positive relationship with cash shortages.	Supported	Strong positive SHAP value for low capacity utilization; high incidence of shortage in low-utilization firms (from descriptive stats).
H2	Longer debtor days have a significant positive relationship with cash shortages.	Supported	Debtor Days identified as the #1 most important feature by SHAP; clear positive correlation in the model.
H3	Absence of formal cash flow planning has a significant positive relationship with cash shortages.	Supported	Formal Cash Flow Budget was a top-5 important feature; absence consistently increased predicted risk.
H4	XGBoost will significantly outperform logistic regression.	Supported	XGBoost AUC = 0.93 vs. Logistic Regression AUC = 0.79.
H5	SHAP will identify non-financial variables as critically important drivers.	Supported	Two of the top five drivers (Capacity Utilization, Cash Flow Budget) are non-financial/operational-managerial variables.

The analysis leads to the rejection of the null hypotheses associated with H1-H5. The results robustly confirm that the identified variables are potent predictors and that the explainable AI framework successfully provides deep insights into the drivers of cash shortage risk.

5. DISCUSSION

This study set out to address a critical gap in the understanding and prediction of cash shortages within Indian MSMEs by developing an explainable AI framework. The results presented in the previous section provide robust empirical evidence to evaluate the proposed model and offer significant insights. This discussion interprets these findings, explicates their theoretical and practical implications, and contextualizes them within the existing body of literature.

5.1. Interpretation of Key Findings

The analysis yields several key findings that directly address the research questions. First, the superior performance of the XGBoost model (AUC = 0.93) compared to logistic regression (AUC = 0.79) unequivocally demonstrates that cash shortages are not the result of simple, linear relationships but rather arise from complex, non-linear interactions among a multitude of factors. This answers **RQ2** by confirming that advanced machine learning is not merely a technical upgrade but a necessary evolution for accurate prediction in the complex MSME environment.

Second, the SHAP analysis provides a nuanced answer to **RQ1** and **RQ3**. The identification of **Debtor Days** and **Delayed Payment Receivables** as the top two predictors empirically validates what has long been anecdotally acknowledged: the delayed payment crisis is the single greatest liquidity threat to Indian MSMEs (Rajamani et al., 2022; Singh & Wasdani, 2016). However, the model moves beyond confirmation to reveal the relative importance of this factor compared to others. Furthermore, the high importance of **Capacity Utilization** and the **Existence of a Formal Cash Flow Budget** reveals that internal operational efficiency and managerial practices are equally critical, if not more so, than some traditional financial ratios. This finding challenges the orthodox focus of credit assessment primarily on financial statements.

The local interpretability afforded by SHAP force plots answers the deeper aspect of **RQ3**, transforming the model from a predictive tool into a diagnostic system. For instance, the case study of the high-risk MSME clearly showed that its 92% risk score was driven predominantly by a 120-day debtor period and 30% capacity utilization. This level of specificity is unprecedented in traditional models and provides a clear, actionable roadmap for intervention.

5.2. Theoretical Implications

The findings of this study carry several important theoretical implications for the literature on financial distress and SME finance.

First, they provide strong support for **reconceptualizing cash shortage as a distinct financial syndrome**, as proposed in the conceptual framework. The fact that the model accurately predicts this specific condition, using a constellation of operational and managerial variables, underscores that it has a unique etiology separate from long-term solvency-based failure. This calls for a theoretical shift from a unified view of "distress" to a more nuanced understanding of different financial pathologies.

Second, the study makes a significant **methodological contribution to financial distress prediction theory**. By successfully integrating XAI with a high-performance ML model, it demonstrates a viable path to resolving the long-standing accuracy-interpretability trade-off (Molnar, 2022). The application of SHAP values provides a mathematically rigorous framework for moving from correlation-based analysis to explainable, causal inference-like diagnostics within predictive models, setting a new standard for transparency in empirical finance research.

Finally, the results **challenge the sufficiency of traditional finance theory** for understanding MSME vulnerability. The high ranking of non-financial variables like capacity utilization and financial planning suggests that theories of MSME distress must be expanded to incorporate insights from operations management and behavioral finance. A holistic theory of MSME financial health must account for the interplay between market conditions (delayed payments), internal operations (capacity use), and managerial capabilities (planning).

5.3. Limitations and Avenues for Future Research

Despite its contributions, this study has limitations that present opportunities for future research. First, the **cross-sectional nature** of the data establishes strong predictive relationships but cannot definitively prove causality. A longitudinal study tracking MSMEs over time would be valuable to confirm the causal pathways suggested by the model. Second, the study focused on **formally registered MSMEs** on the Udyam portal. The vast informal sector remains unrepresented, and future work could explore adapting the framework using alternative data sources to include these enterprises. Third, the model relies on **survey-based data** for some variables; future integration with real-time data from GSTN, account aggregators, or banking APIs could enhance accuracy and reduce manual reporting bias.

Future research could also focus on developing the framework into a user-friendly software-as-a-service (SaaS) tool and conducting field experiments to measure its actual impact on MSME survival and growth. Applying the same methodology to MSMEs in other emerging economies would also test its generalizability and uncover cross-country differences in the drivers of financial distress.

6. CONCLUSION

6.1. Summary of the Study

This research was motivated by a critical paradox: the monumental economic importance of Indian Micro, Small, and Medium Enterprises (MSMEs) is perpetually undermined by their vulnerability to cash shortages—a debilitating liquidity crisis that stifles growth, jeopardizes jobs, and weakens the national economic fabric. The study identified a tripartite gap in the extant literature: a methodological scarcity of micro-level predictive analytics using Explainable AI (XAI), a theoretical under-conceptualization of cash shortage as a distinct syndrome, and a practical absence of actionable, transparent diagnostic tools for pre-emptive intervention.

In response, this study developed and validated a novel explainable AI framework for predicting and diagnosing cash shortages. Grounded in a positivist philosophy and a cross-sectional survey design, data was collected from a stratified random sample of 1,230 Indian MSMEs. The study leveraged the predictive power of the XGBoost algorithm, achieving outstanding accuracy (AUC = 0.93), and crucially, employed the SHAP (SHapley Additive exPlanations) framework to interpret the model's outputs.

The empirical analysis yielded several key findings:

1. **The primary drivers** of cash shortage risk, as identified by SHAP, are delayed receivables (Debtor Days), low capacity utilization, and the absence of formal cash flow planning.
2. **The XGBoost model significantly outperformed** a traditional logistic regression baseline, validating the necessity of modeling complex, non-linear relationships.
3. **The explainable AI approach** successfully bridged the accuracy-interpretability chasm, providing both global insights into sector-wide risks and local, actionable diagnostics for individual MSMEs.

The results led to the rejection of the null hypotheses, confirming that operational inefficiency, poor working capital management, and lack of financial planning are potent, significant predictors of cash shortages. The study thus successfully delivered on its primary objective: to create a micro-level diagnostic tool that is not only predictive but also interpretable and actionable.

6.3. Concluding Remarks

This research ultimately underscores the transformative potential of explainable artificial intelligence in addressing long-standing challenges in business and management. By moving beyond the "black box" of conventional machine learning, the proposed framework demystifies the complex etiology of financial distress for a critically important yet vulnerable sector. It empowers MSME owners with knowledge, equips lenders with transparent tools, and provides policymakers with empirical evidence for targeted intervention.

In conclusion, fostering a resilient MSME sector is not merely an economic imperative but a cornerstone of inclusive and sustainable development. This study demonstrates that by democratizing access to sophisticated, interpretable diagnostics, we can shift the paradigm from reactive crisis management to proactive financial resilience. The fusion of advanced data science with a deep understanding of ground-level business realities, as presented here, offers a viable path toward unlocking the full potential of India's—and indeed, the world's—most important economic engines.

REFERENCES

1. *Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138-52160.*
2. *Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, 23(4), 589–609.*
3. *Beaver, W. H. (1966). Financial ratios as predictors of failure. Journal of Accounting Research, 4, 71–111.*
4. *Brealey, R. A., Myers, S. C., & Allen, F. (2020). Principles of corporate finance (13th ed.). McGraw-Hill Education.*
5. *Bryman, A., & Bell, E. (2015). Business research methods (4th ed.). Oxford University Press.*
6. *Chen, T., & Guestrin, C. (2016, August). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).*
7. *Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.*
8. *Fowler, F. J. (2014). Survey research methods (5th ed.). SAGE Publications.*
9. *Gani, W., & Vijayarani, K. (2022). Financing gaps in Indian MSMEs: An analysis of challenges and policy recommendations. International Journal of Professional Business Review, 7(4), e0641.*
10. *International Finance Corporation. (2022). Financing India's MSMEs: Estimating the credit gap. World Bank Group.*
11. *López-Gutiérrez, C., Sanfilippo-Azofra, S., & Torre-Olmo, B. (2023). Financial distress in SMEs: A review of the literature and future research directions. Journal of Small Business Management, 61(1), 1-30.*
12. *Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), Advances in Neural Information Processing Systems 30 (pp. 4765-4774). Curran Associates, Inc.*

13. *Maheshkar, C., & Soni, N. (2022). Problems faced by Indian Micro, Small and Medium Enterprises (MSMEs): An empirical study. Paradigm, 26(1), 7-25.*
14. *Ministry of Micro, Small and Medium Enterprises. (2024). *Annual Report 2023-24*. Government of India.*
15. *Molnar, C. (2022). Interpretable machine learning: A guide for making black box models explainable (2nd ed.). https://christophm.github.io/interpretable-ml-book*
16. *Nunnally, J. C. (1978). Psychometric theory (2nd ed.). McGraw-Hill.*
17. *Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research, 18(1), 109–131.*
18. *Pandey, I. M. (2022). Financial management (12th ed.). Vikas Publishing House.*
19. *Rajamani, A., et al. (2022). Access to finance for MSMEs in India: An empirical investigation. Journal of Risk and Financial Management, 15(7), 294.*
20. *Richards, V. D., & Laughlin, E. J. (1980). A cash conversion cycle approach to liquidity analysis. Financial Management, 9(1), 32–38.*
21. *Saunders, M., Lewis, P., & Thornhill, A. (2019). Research methods for business students (8th ed.). Pearson.*
22. *Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), Contributions to the theory of games II (pp. 307-317). Princeton University Press.*
23. *Singh, S., & Wasdani, K. P. (2016). Financing micro, small, and medium enterprises in India: Sources and challenges. Asian Development Bank Institute.*
24. *Štrukelj, T., & Matjaž, M. (2023). Explainable AI in business analytics: A systematic review. Journal of Business Research, 157, 113550.*