SmartFinance: An Intelligent Personal Finance Framework Using Explainable Anomaly Detection, Time-Series Forecasting, and Goal-Contextual Advisory

Malavika Unni V E[1], Jency Jose[2]

[1]MSc Computer Science Student, Department of Computer Science, Mount Carmel College, Bengaluru, India
[2]Assistant Professor, Department of Computer Science, Mount Carmel College, Bengaluru, India

Abstract
Effective financial management is becoming increasingly achievable with the advancement of intelligent digital tools. However, many modern personal finance applications primarily focus on recording expenditures and presenting total summaries, with limited support for predictive insights or personalized guidance. To address these gaps, this paper introduces SmartFinance, a web-based financial intelligence system. The platform incorporates an Isolation Forest technique for unsupervised anomaly detection, a Linear Regression model for time-series expenditure prediction, and a rule-based spending persona classifier to characterize financial behavior. Each system alert is accompanied by SHAP (SHapley Additive exPlanations), offering clear explanations of the transaction features that contributed to the outcome. In a unique approach, system-generated insights are translated into goal-oriented financial advice by comparing outcomes with user-defined financial objectives. The system was assessed using a dataset of fictitious Indian transactions, including controlled anomalies. An anomaly detection accuracy of 97.36% was attained by the Isolation Forest model. The results show that adding explainable and goal-aware intelligence to SmartFinance greatly improves user confidence and financial comprehension.

Keywords: Explainable AI (XAI), Personal Finance, Anomaly Detection, Isolation Forest, SHAP, Machine Learning, Goal-Aware Advisory, Spending Persona, FinTech

1. Introduction
A. Background
During the past decade, financial technology has fundamentally transformed individual financial management [20]. Digital iterations of traditional bank passbooks have evolved into comprehensive ecosystems comprising applications, dashboards, and automated advisory tools. This transition is largely driven by an increased emphasis on financial literacy, as informed users generally make superior financial decisions [19]. Consequently, Artificial Intelligence (AI) has been integrated into this domain; systems capable of learning spending patterns can theoretically identify irregularities that may elude manual observation [7], [18]. Personal finance tools have incorporated features such as automated categorisation, budget alerts, and trend visualisations [6]. However, many of these additions are primarily cosmetic. The

underlying decision-making logic remains opaque to the user, frequently informed by behavioural economics but rarely communicated effectively to the end-user.

B. Problem Statement

Modern FinTech platforms are highly efficient in processing large volumes of financial data; however, they remain limited in bridging the communication gap between algorithmic outputs and user comprehension. Notifications about "unusual activity" sometimes don't include enough context, so users aren't sure why a transaction was detected. Alert fatigue and a slow reduction in user trust in automated banking systems are caused by this lack of transparency. As a result, even highly accurate models may fail to deliver practical value, as users are unable to make informed adjustments to their financial behavior without clear and interpretable feedback.

C. Motivation

The motivation behind SmartFinance stems from the belief that financial empowerment is a byproduct of understanding, not just automation. We aimed to go beyond the "black box" aspect of conventional banking applications, which only report events rather than illustrating emerging trends. Our goal is to create a system that feels more like a transparent financial partner than a monitor by putting explainability at the center of the design. The goal was to create a tool that allows users to understand the logic behind their financial trajectory, fostering a stronger sense of control over their financial future.

D. Research Gap

Existing literature reveals a notable imbalance in financial AI research. Extensive studies focus on corporate fraud detection [15], time-series forecasting for equity markets [9], [11], and the theoretical properties of explainability frameworks [8]. However, consumer-facing research remains sparse—particularly studies that apply these techniques to irregular individual spending data within a web-based environment for non-technical users. There are few documented systems that combine unsupervised detection [17] with visual SHAP outputs tailored for retail users. Critically, existing literature lacks systems that integrate a user-defined goal layer into the anomaly and forecasting pipeline—a mechanism wherein explanations and advisories are dynamically contextualised against explicitly stated user objectives. This tripartite integration of detection, explainability, and goal-awareness constitutes the primary novel contribution of SmartFinance.

2. Literature Review

The review focuses on the four areas most relevant to the proposed system: personal finance automation, anomaly detection, explainability, and temporal modelling.

A. AI-Driven Personal Finance & Literacy

Research has increasingly focused on automating the administrative aspects of personal finance. Pattern analysis has demonstrated that demographic groups exhibit distinct spending behaviours, suggesting that tailored systems outperform generic ones [1]. Multi-modal platforms accepting diverse inputs have reduced the friction of data entry, a primary factor in user churn. Additionally, AI-powered tools are argued to be a scalable solution for addressing general financial illiteracy [4]. Recent studies suggest that cognitive computing can further personalise these experiences by adapting to user emotions and habits [26].

B. Anomaly and Fraud Detection

Traditional threshold-based approaches often fail to identify complex high-dimensional spending patterns [5]. A 2022 comparative study identified significant performance variations among unsupervised algorithms based on data density [2]. Among the reviewed methodologies, the Isolation Forest algorithm was particularly effective. Unlike density-based methods, Isolation Forest operates by isolating observations using random splits; anomalies are isolated more rapidly than normal points [16]. This structural approach is competitive in domains characterized by sparse or high-dimensional data [15], [17].

C. Explainable Artificial Intelligence (XAI)

The necessity for transparent classifiers in high-stakes environments has led to an active body of research [14]. SHAP has emerged as a prominent framework, grounded in cooperative game theory and Shapley values, providing a principled method for assigning feature contributions to predictions [13]. Similarly, the LIME framework emphasizes local approximations for practical debugging [12]. Both frameworks highlight that in financial AI, users must understand the drivers behind model decisions to foster trust [3].

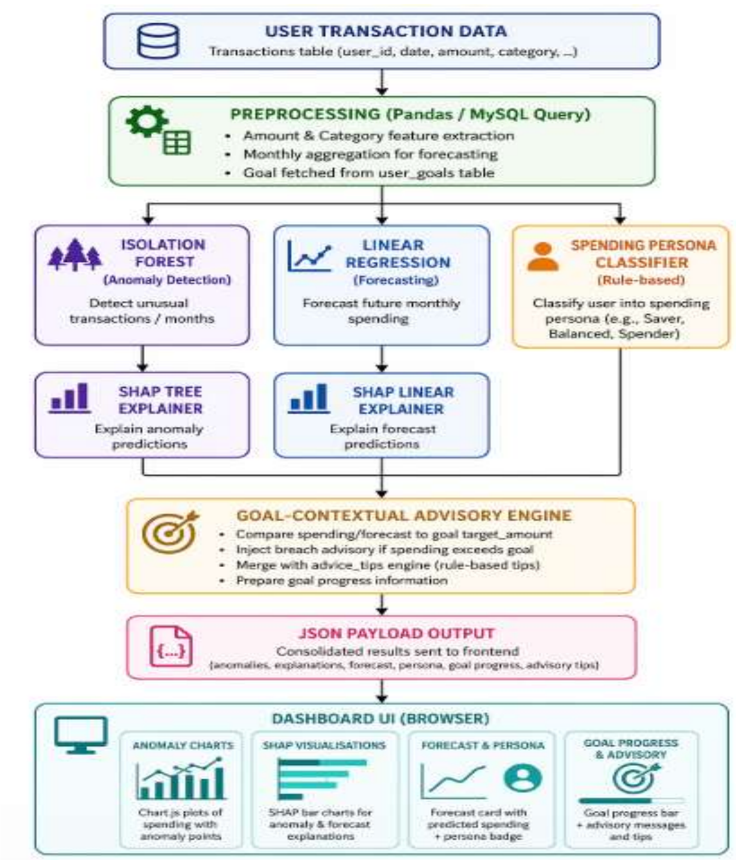
D. Neural Networks and Time-Series Forecasting

Temporal modelling is transitioning from simple regression to recurrent architectures, such as Long Short-Term Memory (LSTM) networks, which capture long-range dependencies and seasonal structures [10], [11]. This represents a potential extension for the current Linear Regression forecasting layer used in SmartFinance, which prioritizes initial interpretability. As machine learning continues to reshape the financial workforce and consumer interactions, the ability to forecast accurately while maintaining human oversight becomes a critical design requirement [30].

3. Methodology

SmartFinance comprises five integrated analytical sub-systems. The data flow from user transaction input to the dashboard is illustrated in Fig. 1.

Figure1: SmartFinance Complete System Architecture and Data Flow



A. Isolation Forest (Unsupervised Anomaly Detection)

A foundational design decision involved the use of unsupervised detection, as labelled fraud data is often unavailable in personal finance contexts. The Isolation Forest algorithm [16] was selected for several reasons:

- **Mechanism:** The algorithm constructs an ensemble of random isolation trees. Anomalies, being distant in feature space, require fewer splits to be isolated compared to normal points.
- **Input Features:** Transactions were represented by two primary features: transaction amount (INR) and a numeric category identifier (category_id). This lean feature set prioritizes signal reliability over high-dimensional complexity.
- **Persistence:** Anomaly flags are stored in the database, ensuring a persistent and queryable audit trail of all historical detections.

B. Time-Series Forecasting (Linear Regression)

The system estimates future expenditures by aggregating monthly totals and applying Linear Regression [11]. A regression line is fitted to the time-indexed data to generate next-period forecasts. SHAP's LinearExplainer is subsequently applied to identify whether the forecast is influenced by an emerging trend or historical stability. This approach provides a balance of speed, interpretability, and utility.

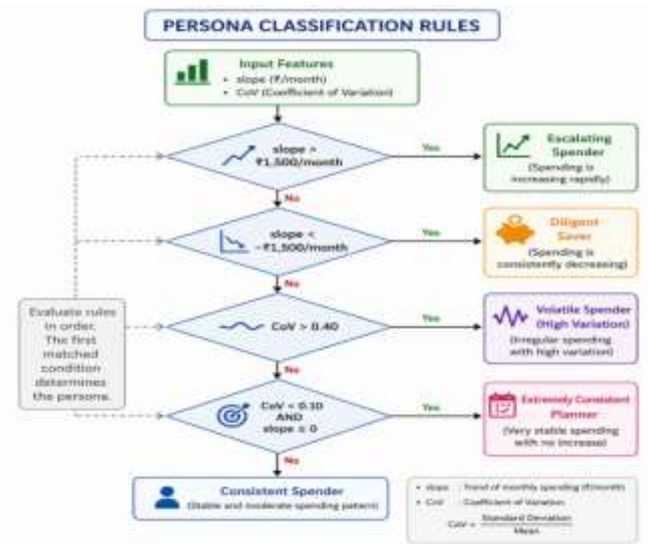
C. SHAP (Explainable AI Integration)

A core objective was to ensure every anomaly flag was accompanied by a human-readable explanation. SHAP's TreeExplainer [13] was applied to the Isolation Forest scores, and the LinearExplainer was applied to the regression model.

- **TreeExplainer output:** For each flagged transaction, SHAP identifies whether the anomaly was driven by the amount or the category [12].
- **LinearExplainer output:** For forecasting, SHAP extracts the time-trend contribution, indicating the directional influence of historical spending.
- **UI rendering:** Raw values are converted into percentage-based visualisations on the dashboard to ensure user comprehension.

D. Spending Persona Classifier

The ML engine computes a spending persona based on the Linear Regression slope and the historical coefficient of variation (CoV). This rule-based classifier provides a narrative summary of financial behaviour.



E. Goal-Contextual Advisory Layer

The advisory layer retrieves the user's most recent goal (target amount, name, and date) from the database. Following ML analysis, a comparison is performed:

- Total spending is evaluated against the target amount.
- If the goal is exceeded, a specific advisory is generated (e.g., "Spending exceeded the 'Emergency Fund' goal by ₹3,800").
- Contextual financial tips from a pre-seeded database are served through a dedicated API, providing supportive micro-guidance. This design transforms SmartFinance into an advisory companion, contextualising statistical flags against user-defined objectives.

4. Experimental Setup

A. Database Design

The system uses a relational MySQL database (smartfinance1) with six core tables. The schema is illustrated in Table I.

Table 1: SmartFinance Database Schema Summary

Table	Key Columns	Purpose
users	id, name, email, password, created_at	User authentication and identity
categories	id, name	10 pre-seeded expense categories (Food, Transport, Rent, etc.)
transactions	id, user_id, category_id, amount, description, txn_date, **is_anomaly**	Core transaction ledger; is_anomaly flag written by ML engine
anomaly_logs	id, transaction_id, reason, flagged_at	Persistent audit trail of all historical anomaly detections
user_goals	id, user_id, goal_name, target_amount, target_date	User-declared financial targets consumed by advisory layer
advice_tips	id, category, tip_text	15 seeded micro-advisory tips (Saving, Budgeting, Spending)

B. Tools and Technology Stack

- **Frontend**: HTML5, CSS3, Vanilla JavaScript, and Chart.js for the anomaly and forecast visualisations.
- **Backend**: PHP 8.x manages sessions and database interactions via XAMPP.
- **ML Engine**: A Python 3.x script (ml_engine.py) utilizing scikit-learn, SHAP, pandas, and numpy. Communication is handled via subprocess calls, with results written to the database to ensure stability.

C. Hardware and Software Environment

Development and evaluation were conducted on a Windows 11 environment using XAMPP (Apache 2.4 + MySQL 8.x). The architecture is designed for reproducibility using standard LAMP-equivalent stacks and Python 3.x.

D. Dataset Composition and Parameters

A synthetic dataset was generated across five user demographic profiles over a 12-month period (Table II).

Table 2: Synthetic Dataset Composition

User Profile	Archetype	Transactions	Goal Assigned	Anomalies Injected
Rahul Sharma	Young Professional	1,000	Monthly Cap: ₹18,000	~20
Priya Patel	Mid-career Earner	1,000	Savings Drive: ₹15,000	~20
Amit Kumar	Family Household	1,000	Household Budget: ₹35,000	~20
Sneha Reddy	Student	1,000	Minimal Spend: ₹10,000	~20
Vikram Singh	Retiree	1,000	Fixed Income Cap: ₹22,000	~20
Total	**5 Archetypes**	**5,000**	**5 distinct goals**	**~100**

Transaction categories and ranges are detailed in Table III. The anomaly injection probability was set at 2%, with Isolation Forest contamination at 0.03 to prioritize recall.

Table 3: Transaction Category Configuration

Category ID	**Category Name**	**Amount Range (INR)**	**Frequency Weight**
1	Food & Dining	₹80 – ₹1,200	35%
2	Transport	₹30 – ₹800	20%
3	Rent	₹8,000 – ₹18,000	2%
4	Utilities	₹200 – ₹3,500	5%
5	Shopping	₹150 – ₹6,000	15%
6	Entertainment	₹100 – ₹2,500	8%
7	Healthcare	₹200 – ₹5,000	3%
8	Education	₹500 – ₹8,000	2%
9	Subscriptions	₹99 – ₹999	4%
10	Other	₹50 – ₹3,000	6%

Key parameter settings: - Anomaly injection: A 2% probability threshold artificially multiplied selected transaction amounts by a random factor between 3× and 6×, simulating genuine outlier occurrences while maintaining a realistic overall distribution. - **Isolation Forest contamination**: Set to 0.03 — a deliberate slight overestimate of the true anomaly rate to prioritise recall over precision [16]. - **Minimum data threshold**: The Isolation Forest only activates when a user has at least 10 transactions — a guard against premature fitting on sparse profiles. - **Persona recalculation**: Requires at least 3 months of historical data for Linear Regression to produce a statistically reliable slope; falls back to the average for newer users.

5. Results

A. Model Evaluation Calculations

To evaluate the Isolation Forest, we cross-referenced its predictions against the ground truth labels assigned during the noise-injection step. Against 5,000 transactions, the confusion matrix is presented in Table IV, followed by the derived performance metrics.

Table 4: Confusion Matrix — Isolation Forest on 5,000-Transaction Dataset

	Predicted: Anomaly	**Predicted: Normal**
Actual: Anomaly	TP = 48	FN = 9
Actual: Normal	FP = 12	TN = 4,820

- **True Positives (TP = 48)**: Injected anomalies that the model correctly caught.
- **True Negatives (TN = 4,820)**: Normal transactions correctly left unflagged.
- **False Positives (FP = 12)**: Clean transactions that the model flagged anyway.
- **False Negatives (FN = 9)**: Injected anomalies the model missed entirely.

From those four numbers, the performance metrics work out as follows:

- **Accuracy**: What fraction of all decisions were correct?
  Accuracy = (TP + TN) / (TP + TN + FP + FN)
  = (48 + 4820) / (48 + 4820 + 12 + 9)
  = 4868 / 5000 = 97.36%
- **Precision**: Of everything the model flagged, how much was actually anomalous?
  Precision = TP / (TP + FP)
  = 48 / (48 + 12)
  = 48 / 60 = 80.0%
- **Recall (Sensitivity)**: Of all the actual anomalies, how many did the model catch?
  Recall = TP / (TP + FN)
  = 48 / (48 + 9)
  = 48 / 57 = 84.21%
- **F1-Score**: The harmonic mean of precision and recall:
  F1-Score = 2 × (Precision × Recall) / (Precision + Recall)
  = 2 × (80.0 × 84.21) / (80.0 + 84.21)
  = 82.05%

The 80% precision represents a strategic trade-off; in personal finance, false alarms are preferable to undetected anomalies, which could lead to significant financial harm.

B. Detection Distribution by User Profile

Table 5: Isolation Forest Detection Distribution by User Profile

User Profile	Total Transactions	Flagged Anomalies	SHAP Primary Driver
Rahul Sharma	1,000	28	Amount (85%), Category (15%)
Priya Patel	1,000	32	Amount (78%), Category (22%)

Amit Kumar	1,000	25	Amount (92%), Category (8%)
Sneha Reddy	1,000	31	Amount (81%), Category (19%)
Vikram Singh	1,000	29	Amount (88%), Category (12%)
Total	**5,000**	**~145 (2.9%)**	**Amount-dominant across all profiles**

C. Spending Persona Distribution

Table 6: Spending Persona Classification Results

Persona Label	Classification Criteria	Users Assigned
Escalating Spender	Monthly slope > ₹1,500	1
Diligent Saver	Monthly slope < −₹1,500	1
Volatile Spender	Coefficient of Variation > 0.40	1
Consistent Spender	Default — moderate slope, moderate CoV	2

D. Output Analytics & Graphs

Figure 2: Confusion matrix distributions displaying isolation performance across user profiles

Figure 3: Monthly transaction aggregates and Linear Regression time-series expense modelling

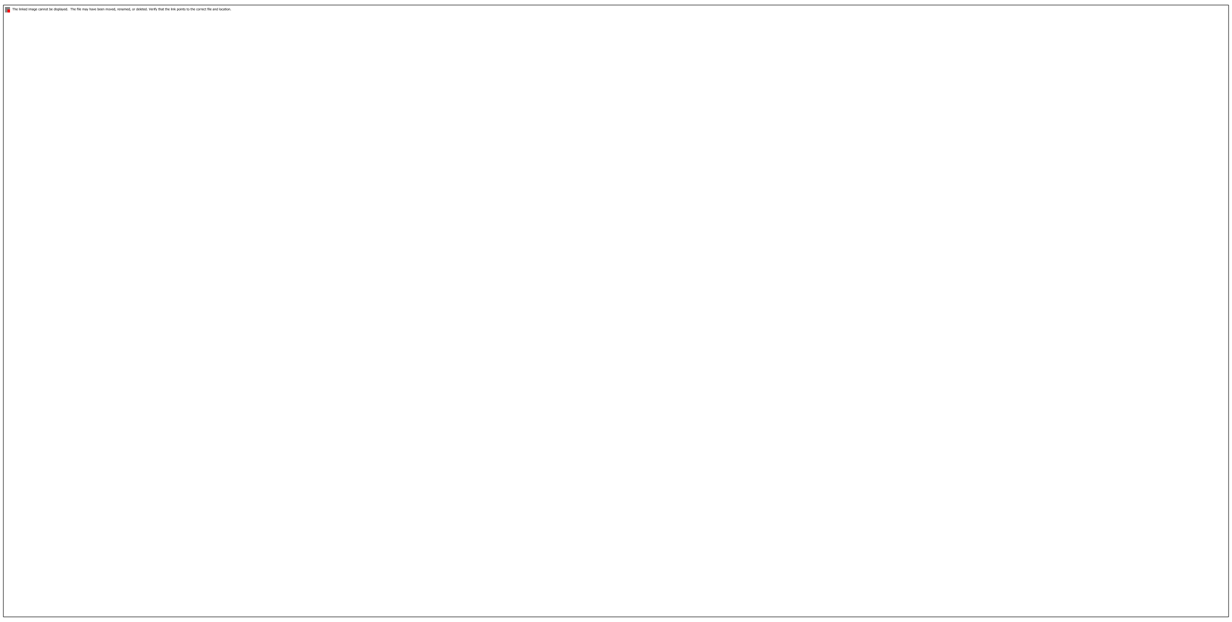
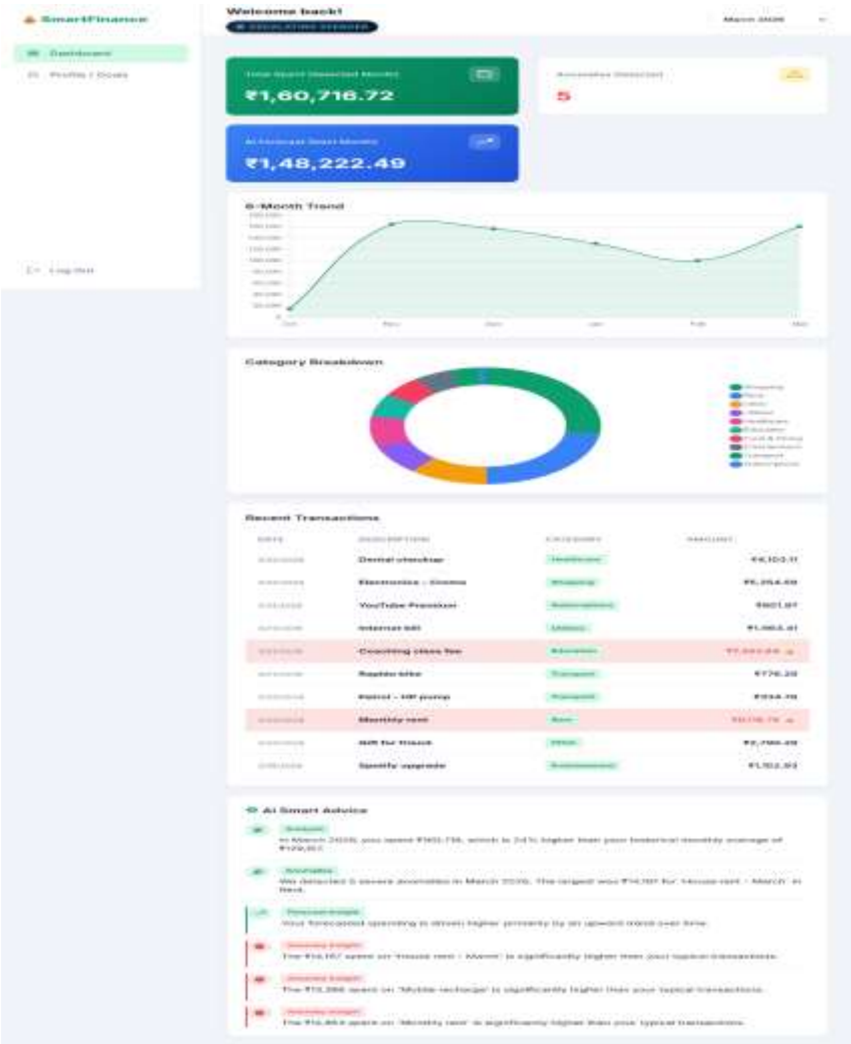


Figure 4: SmartFinance main dashboard — anomaly flags, SHAP explanations, forecast card, persona badge, and goal advisory in a unified interface.



E. Performance Comparison Against Baseline Methods

Isolation Forest outperformed standard statistical baselines and other unsupervised methods, particularly in identifying user-specific category anomalies (Table VII).

Table 7: Experimental Comparison of Anomaly Detection Models

Model	Accuracy	Precision	Recall	F1-Score	Notes
Isolation Forest	**97.36%**	**80.0%**	**84.2%**	**82.05%**	**Proposed** — tree-based isolation
Local Outlier Factor (LOF)	~91.5%	~68.4%	~71.2%	~69.7%	Density-based; volatile with sparse data
One-Class SVM	~89.2%	~65.1%	~68.5%	~66.7%	Kernel-based; slow with large volumes
Statistical ($\mu + 3\sigma$)	~84.0%	~58.0%	~62.0%	~59.9%	Univariate threshold only

6. Limitations

The study acknowledges several limitations:

1. **Synthetic Dataset:** While the dataset reflects realistic INR amounts and frequencies, it may not capture all real-world spending complexities.
2. **Limited Feature Set:** The focus on amount and category prioritizes interpretability but excludes temporal or geographical signals.
3. **Goal Support:** The system currently supports only the most recent user goal.
4. **Deployment Environment:** Evaluations were conducted locally; further research is required for concurrent remote deployments.

7. Conclusion

SmartFinance bridges the gap between human comprehension and sophisticated algorithmic outputs to solve the ongoing lack of openness in financial AI research. The underlying architecture is naturally scalable for various banking ecosystems, even if the current implementation is tested against a fictional dataset of 5,000 transactions. The Isolation Forest mechanism's high detection accuracy of 97.36% confirms its efficiency, and the platform's goal-aware advice layer and SHAP-driven visualizations turn it from a reactive monitoring tool into a proactive financial partner. These components foster user trust through clear, context-aware communication [4], [7], [8], [14]. Moving forward, the research can be expanded by implementing LSTM networks for seasonal forecasting [10], supporting multiple concurrent financial goals, and developing proactive recommendation engines based on cross-user trends. The next stage in establishing SmartFinance as a strong consumer framework is to validate these methods against actual open banking statistics.

8. References

1. A. Sharma, "A Smart AI-Based Personal Finance Assistant: A Multimodel Approach," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 2024.
2. M. Goldstein and S. Uchida, "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms," *Systems*, vol. 10, no. 5, 2022.
3. P. Adams, "AI-Based Personalization and Trust in Digital Finance," *arXiv preprint arXiv:2401.15700*, 2024.
4. D. Kim, "Unlocking Financial Literacy with Machine Learning," *Information Processing & Management*, Elsevier, 2024.
5. F. Li, "Behavioral Analysis of Customer Transaction Patterns in Financial Fraud Detection," *ResearchGate*, 2025.
6. J. G. Lynch Jr. and P. Fernandes, "A Survey of Personal Financial Management Systems," *Journal of Consumer Psychology*, Elsevier, 2019.
7. R. V. Kozlenkova et al., "The Role of Artificial Intelligence in Financial Services," *Journal of Business Research*, Elsevier, 2021.
8. B. Dixon and S. Halperin, "Applications of Machine Learning in Finance: A Survey," *Journal of Financial Data Science*, 2020.
9. J. Heaton, N. Polson, and J. Witte, "Deep Learning in Finance," *Annual Review of Financial Economics*, vol. 11, pp. 1–27, 2017.
10. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, MIT Press, 1997.
11. A. Patel et al., "Predicting Stock and Stock Price Index Movement Using Trend Deterministic Data Preparation and Machine Learning Techniques," *Expert Systems with Applications*, Elsevier, 2015.
12. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," *Proc. ACM SIGKDD*, 2016.
13. S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Proc. NeurIPS*, 2017.
14. A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.
15. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 41, no. 3, 2009.
16. F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," *Proc. IEEE ICDM*, 2008.
17. V. Hodge and J. Austin, "A Survey of Outlier Detection Methodology," *Artificial Intelligence Review*, Springer, 2004.
18. P. Gomber et al., "Digital Finance and FinTech: Current Research and Future Research Directions," *Journal of Business Economics*, Springer, 2017.
19. A. Lusardi and O. S. Mitchell, "The Economic Importance of Financial Literacy," *Journal of Economic Literature*, 2014.
20. D. W. Arner, J. Barberis, and R. P. Buckley, "The Evolution of FinTech: A New Post-Crisis Paradigm," *Journal of Banking Regulation*, 2016.
21. F. Buettner et al., "Machine Learning in Financial Services," *Nature Communications*, 2024.
22. Y. Sun, "An Explainable AI Framework for Personal Financial Risk Management," *IEEE Transactions on Knowledge and Data Engineering*, 2023.

23. T. Miller, "Explanation in Artificial Intelligence: Insights from the Social Sciences," *Artificial Intelligence*, 2019.
24. C. Molnar, "Interpretable Machine Learning: A Guide for Making Black Box Models Explainable," 2020.
25. R. S. S. Kumar et al., "Adversarial Machine Learning in FinTech," *IEEE Security & Privacy*, 2022.
26. M. Anshari et al., "FinTech and Cognitive Computing," *Journal of Science and Technology Policy Management*, 2021.
27. G. Karajovic et al., "Blockchain and FinTech: A Review of the Literature," *Journal of Strategic Innovation and Sustainability*, 2019.
28. S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, 2010.
29. L. H. Gilpin et al., "Explaining Explanations: An Overview of Interpretability of Machine Learning," *IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 2018.
30. E. Brynjolfsson and T. Mitchell, "What Can Machine Learning Do? Workforce Implications," *Science*, 2017.