IJVAST | Vol. 1, Issue 1 | Jan 2026 | Open Access — Peer Reviewed Research Paper

AI-Based Personal Finance and Expense Analyzer

Aditya R. Pawar[1]

Department of Computer Science and Engineering

Institute of Engineering and Technology, Latur, Maharashtra, India

Abstract—The rapid expansion of digital payment ecosystems has made it easier for individuals to spend money and considerably harder for them to understand where that money goes. This paper presents the design, implementation, and evaluation of an AI-Based Personal Finance and Expense Analyzer: a web-based system that consolidates income and expenditure records, classifies every transaction into a meaningful spending category, and converts the resulting data into actionable financial guidance. The proposed system integrates four intelligent components — an automated expense categorization engine built on a TF-IDF and Linear Support Vector Machine pipeline, a budget monitoring module with threshold-based alerting, a short-horizon expenditure forecasting module based on seasonal time-series decomposition, and a rule-guided savings recommendation engine that translates analytical output into personalised advice. A corpus of 24,500 anonymised transaction records was assembled from public financial datasets and synthetically generated Indian retail transactions to train and validate the models. The categorization engine achieved an overall accuracy of 93.4% and a macro F1-score of 0.91 across twelve spending categories, while the forecasting module recorded a mean absolute percentage error of 8.7% for month-ahead category-wise prediction. A four-week pilot with 40 student users showed a measurable improvement in budget adherence and a reduction in unplanned discretionary spending. The paper also documents the system architecture, module design, evaluation methodology, observed limitations relating to data sparsity and cold-start users, and the future scope of the work.

Index Terms—Personal Finance Management, Expense Tracking, Machine Learning, Transaction Classification, Budget Analytics, Expenditure Forecasting, Savings Recommendation, Natural Language Processing, Support Vector Machine, Time-Series Prediction, Financial Data Visualization, Fintech, Decision Support System.

I. INTRODUCTION

Personal financial management has become both more necessary and more difficult in the digital era. The widespread adoption of unified payment interfaces, mobile wallets, credit instruments, and subscription-based services has dramatically increased the number of small, frequent, and dispersed financial transactions that an average individual performs each month. While each of these payments is individually trivial, their cumulative effect on monthly savings is substantial and frequently invisible to the person making them.

The difficulty is not a shortage of financial data but a shortage of interpretation. Bank statements, wallet histories, and card summaries are maintained in isolated silos, expressed in terse merchant codes, and offered without any form of behavioural analysis. Consequently, a user may be able to state exactly how much was spent in a month while remaining unable to explain why the amount was higher than intended, which category caused the deviation, or what corrective action would restore the intended saving rate [1], [2].

Conventional expense-tracking applications address only the recording problem. They rely on manual entry or on static keyword rules that map a merchant string to a fixed category. Such rules degrade quickly because merchant descriptors are

C. Preprocessing and Feature Extraction

Raw descriptors are lower-cased, stripped of payment-gateway prefixes, reference numbers, and terminal identifiers, and tokenised. The cleaned string is converted into a combined word-level and character-level TF-IDF vector. Three engineered features are appended: the transaction amount bucket, the day-of-week index, and a recurrence flag indicating whether a similar descriptor and amount pair has appeared in previous billing cycles.

D. Expense Categorization Engine

The categorization engine assigns each expense to one of twelve categories: food and dining, groceries, transport, utilities, rent and housing, education, healthcare, shopping, entertainment, subscriptions, transfers and savings, and miscellaneous. A Linear Support Vector Machine is used as the primary classifier. Predictions falling below a calibrated confidence threshold are routed to a review queue where the user confirms or corrects the label, and the correction is retained for periodic incremental retraining.

E. Budget Management Module

Users define monthly budget ceilings either manually or by accepting a system-suggested allocation derived from the trimmed mean of the preceding three months. The module noisy, abbreviated, regionally inconsistent, and continuously changing. They also offer no predictive capability: the user learns about an overspend only after it has occurred [3].

This paper proposes an AI-Based Personal Finance and Expense Analyzer that treats the transaction log as a learnable signal rather than a passive ledger. The system automates categorization through supervised text classification, monitors category-wise budgets continuously, forecasts near-term expenditure using time-series modelling, and generates savings recommendations grounded in the user's own historical behaviour.

The principal contributions of this work are: (1) a modular architecture for an end-to-end personal finance analyzer suitable for deployment as a lightweight web application; (2) an automated expense categorization pipeline trained on Indian retail transaction descriptors; (3) a category-wise expenditure forecasting method for short planning horizons; (4) a transparent savings recommendation engine that links each suggestion to an observable spending pattern; and (5) an empirical evaluation covering model accuracy and a four-week user pilot.

II. RELATED WORK

A. Rule-Based Expense Trackers

Early personal finance tools were essentially digital ledgers with reporting features. Categorization was performed by regular-expression matching over merchant names, with the user correcting errors manually. Studies of such systems report high abandonment rates, attributing them primarily to the sustained manual effort required for data entry and correction [4].

B. Machine Learning for Transaction Classification

Transaction descriptors are short, abbreviation-heavy text strings, which makes their classification a constrained natural language processing problem. Character n-gram and TF-IDF representations combined with linear classifiers have been shown to outperform both keyword rules and several deep architectures on short-text banking data, largely because the vocabulary is small and highly discriminative [5], [6]. More recent work applies transformer-based encoders, achieving marginal accuracy gains at a considerable computational cost that is difficult to justify for a student-scale deployment [7].

C. Expenditure Forecasting

Household spending exhibits strong weekly and monthly seasonality driven by salary credit dates, rent cycles, and recurring bills. Classical approaches such as ARIMA and Holt-Winters exponential smoothing remain competitive for short horizons and limited history, whereas recurrent neural models evaluates consumption against each ceiling after every transaction and raises graded alerts at 70%, 90%, and 100% utilisation, together with a pace indicator that compares elapsed budget against elapsed days in the cycle.

F. Financial Analytics Module

The analytics module computes the savings rate, the ratio of fixed to discretionary expenditure, category-wise month-on-month variation, merchant concentration, and the recurring-payment inventory. Results are rendered as category distribution charts, cumulative spending curves, and a cash-flow summary that contrasts credited income against total outflow.

G. Expenditure Prediction Module

The prediction module produces a month-ahead expenditure estimate for each category. Historical series are decomposed into trend, seasonal, and residual components, after which Holt-Winters exponential smoothing is applied to the deseasonalised series. Categories with fewer than three observed cycles fall back to a weighted moving average, ensuring that new users still receive a usable projection.

H. Recommendation Engine

Recommendations are produced by evaluating analytical indicators against a curated rule set. Each generated suggestion is accompanied by the evidence that triggered it and the estimated monthly saving if it is adopted, so that the user can assess the advice rather than merely receive it.

IV. METHODOLOGY

A. Dataset Preparation

A corpus of 24,500 anonymised expense records was assembled. Approximately 60% of the records were derived from publicly available personal-finance datasets, while the remainder were synthesised to reflect Indian merchant naming conventions, unified payment interface descriptor formats, and typical student expenditure patterns. Each record was labelled with one of the twelve target categories by two independent annotators; disagreements, which affected 4.1% of the corpus, were resolved by discussion.

B. Experimental Protocol

The labelled corpus was partitioned into training, validation, and test subsets in a 70:15:15 ratio using stratified sampling to preserve category proportions. Hyperparameters were tuned on the validation subset through grid search with five-fold cross-validation, and all reported metrics were computed on the held-out test subset, which was used exactly once.

C. Baseline Models

require training data that individual users rarely possess [8], [9].

D. Behavioural Recommendation

Research in behavioural finance indicates that generic advice produces little change in spending behaviour, while specific, quantified, and personally referenced feedback measurably improves saving outcomes. This finding motivates the explainable, evidence-linked recommendation strategy adopted in the present work [10], [11].

III. SYSTEM ARCHITECTURE

A. Architectural Overview

The proposed system follows a four-layer architecture consisting of a data layer, an intelligence layer, an application layer, and a presentation layer, as summarised in Fig. 1. Layer separation permits each analytical component to be retrained or replaced without disturbing the user interface or the persistence schema.

B. Data Acquisition Module

The data acquisition module accepts financial records through three channels: manual entry of individual income and expense items, bulk import of bank or wallet statements in CSV and Excel formats, and optical character recognition of uploaded receipt images. Every ingested record is normalised into a common schema containing the transaction date, amount, direction, raw descriptor, payment mode, and an optional user note.

Four classifiers were compared: a keyword rule baseline representing conventional trackers, Multinomial Naive Bayes, Linear Support Vector Machine, and Random Forest. All learned models shared the same feature representation to ensure that observed differences were attributable to the classifier rather than to feature engineering.

D. Evaluation Metrics

Classification quality was measured using accuracy, macro-averaged precision, recall, and F1-score, the macro average being preferred because category frequencies are strongly imbalanced. Forecasting quality was assessed using mean absolute percentage error and root mean squared error over a three-cycle rolling-origin evaluation.

E. Implementation Environment

The prototype was implemented in Python 3.11 using scikit-learn for modelling, pandas and NumPy for data handling, statsmodels for time-series decomposition, and Flask for the application layer. Transaction data were persisted in a relational database, and the interface was built with HTML5, CSS3, and Chart.js. All experiments were executed on a machine with an Intel Core i5 processor and 8 GB of memory, confirming that the system does not depend on specialised hardware.

F. Pilot Study Design

A four-week pilot was conducted with 40 undergraduate volunteers. Participants recorded expenditure through the system for the entire period, with the first two weeks serving as an observation baseline during which analytics and recommendations were withheld. The two subsequent weeks enabled the full analytical feature set, allowing within-subject comparison of budget adherence and discretionary spending.

Fig. 1 — Layered System Architecture of the Proposed Personal Finance and Expense Analyzer

Layer	Principal Components and Responsibilities
Presentation Layer	Dashboard, category charts, budget meters, alert panel, recommendation cards
Application Layer	Authentication, transaction CRUD services, budget controller, report generator
Intelligence Layer	Categorization engine, analytics engine, forecasting module, recommendation engine
Data Layer	Statement importer, OCR receipt parser, normalisation pipeline, transaction database

TABLE I — Comparative Analysis: Manual Tracking vs. Rule-Based Applications vs. Proposed System

Feature	Manual Spreadsheet	Rule-Based Tracker App	Proposed AI Analyzer	Benefit to User
Data Entry	Fully manual	Manual with partial import	Manual, bulk import and OCR	Lower effort per record
Categorization	User defined	Static keyword rules	Supervised ML classifier	Adapts to new merchants
Budget Alerts	None	Fixed limit alert	Graded pace-aware alerts	Warning before overspend

Feature	Manual Spreadsheet	Rule-Based Tracker App	Proposed AI Analyzer	Benefit to User
Analytics	Basic totals	Category pie chart	Trend, ratio and concentration	Diagnostic insight
Prediction	Not available	Not available	Category-wise forecast	Forward planning
Advice	Not available	Generic tips	Evidence-linked suggestions	Actionable guidance
Learning	None	None	Incremental retraining	Accuracy improves with use

TABLE II — Model Performance on the Held-Out Test Set (24,500 Transaction Records)

#	Model / Module	Accuracy	Macro F1	Observation
1	Keyword rule baseline	71.2%	0.64	Fails on unseen merchants
2	Multinomial Naive Bayes	86.5%	0.83	Fast but weak on rare classes
3	Random Forest	90.1%	0.88	Higher inference latency
4	Linear SVM (selected)	93.4%	0.91	Best accuracy and latency balance
5	Forecasting — Holt-Winters	MAPE 8.7%	RMSE 412	Stable over three cycles
6	Forecasting — moving average	MAPE 14.3%	RMSE 689	Fallback for cold-start users

V. RESULTS AND DISCUSSION

A. Categorization Performance

As reported in Table II, the Linear Support Vector Machine achieved the highest test accuracy of 93.4% with a macro F1-score of 0.91, exceeding the keyword rule baseline by 22.2 percentage points. The improvement was most pronounced for merchant strings absent from the training vocabulary, where character n-gram features allowed partial lexical matching that rule-based mapping could not achieve.

Residual errors were concentrated in three category pairs: groceries against food and dining, shopping against miscellaneous, and subscriptions against entertainment. These confusions are semantically reasonable, since a single merchant may legitimately serve more than one purpose, and they are largely corrected by the user review queue during the first weeks of use.

B. Forecasting Accuracy

Month-ahead category forecasts produced a mean absolute percentage error of 8.7% for categories with at least three cycles of history. Accuracy was highest for fixed obligations such as rent and utilities and lowest for entertainment and shopping, which are inherently irregular. The moving-average fallback used for sparse categories recorded a higher error of 14.3%, which remains acceptable as an initial estimate for new users.

C. Pilot Study Outcomes

VII. FUTURE SCOPE

Future development will proceed along five directions. Secure account aggregation through the Account Aggregator framework would eliminate manual statement import and improve data completeness. Replacing the curated rule set with a reinforcement-learning recommender trained on accepted and rejected suggestions would allow the advisory component to personalise itself over time.

Anomaly detection over the transaction stream would permit early identification of duplicate charges, silent subscription renewals, and potentially fraudulent activity. A goal-planning module could translate a stated savings target into a category-wise monthly allocation and track progress against it. Finally, deploying the analyzer as a mobile application with on-device inference and end-to-end encryption would address both the accessibility and the privacy limitations identified above.

VIII. CONCLUSION

This paper presented an AI-Based Personal Finance and Expense Analyzer that integrates automated expense categorization, budget monitoring, financial analytics, expenditure forecasting, and savings recommendation within a single modular system. The categorization engine attained 93.4% accuracy with a macro F1-score of 0.91 across twelve categories, and the forecasting module achieved a mean absolute percentage error of 8.7% for month-ahead prediction.

During the pilot, budget adherence improved from 58% of participants remaining within their declared monthly ceilings in the baseline fortnight to 79% in the analytics-enabled fortnight. Average discretionary expenditure declined by 11.6%, while expenditure on fixed categories remained essentially unchanged, indicating that the observed reduction arose from adjustable rather than committed spending. Participants identified the pace indicator and the recurring-payment inventory as the two most influential features.

D. Discussion

The results support the central premise of this work: the value of a personal finance system lies less in recording transactions than in interpreting them. Automating categorization removed the principal source of user effort, and converting the resulting data into anticipatory alerts and evidence-linked advice produced measurable behavioural change within a short observation window. It should nonetheless be noted that the pilot cohort was small and homogeneous, so the effect sizes reported here are indicative rather than conclusive.

VI. LIMITATIONS

Several limitations qualify these findings. First, the system depends on the completeness of the transaction record; cash expenditure that the user does not enter manually remains invisible to every downstream analysis. Second, a portion of the training corpus was synthetically generated, which may not fully reproduce the noise characteristics of production banking feeds.

Third, cold-start users receive materially weaker forecasts until three billing cycles have been observed. Fourth, the recommendation engine is rule-guided rather than learned, so its coverage is bounded by the curated rule set. Fifth, the pilot involved 40 students over four weeks and therefore cannot support claims about long-term behaviour or about households with more complex financial structures. Finally, the prototype stores financial data locally without the encryption and consent governance that a production deployment would require.

A four-week pilot demonstrated improved budget adherence and a reduction in discretionary expenditure, indicating that interpretable, evidence-linked financial feedback can influence behaviour even over short periods. The limitations relating to data completeness, cold-start users, and the scale of the evaluation define a clear agenda for the extensions outlined above. The work shows that an accessible, low-cost analytical system can convert an ordinary transaction log into practical financial guidance for individual users.

IX. REFERENCES

[1] Reserve Bank of India, Annual Report on Payment and Settlement Systems, RBI, Mumbai, 2024.

[2] National Payments Corporation of India, UPI Ecosystem Statistics, NPCI, Mumbai, 2024.

[3] Kumar, A. and Rao, S., A Survey of Personal Finance Management Applications, International Journal of Computer Applications, vol. 183, pp. 21–29, 2023.

[4] Sharma, P., Usability and Abandonment in Expense Tracking Software, Journal of Human-Computer Interaction Studies, vol. 12, pp. 88–104, 2022.

[5] Joulin, A., Grave, E., Bojanowski, P. and Mikolov, T., Bag of Tricks for Efficient Text Classification, Proceedings of EACL, pp. 427–431, 2017.

[6] Ramos, J., Using TF-IDF to Determine Word Relevance in Document Queries, Proceedings of the First Instructional Conference on Machine Learning, 2003.

[7] Devlin, J., Chang, M., Lee, K. and Toutanova, K., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of NAACL-HLT, pp. 4171–4186, 2019.

[8] Hyndman, R. and Athanasopoulos, G., Forecasting: Principles and Practice, 3rd ed., OTexts, Melbourne, 2021.

[9] Box, G., Jenkins, G. and Reinsel, G., Time Series Analysis: Forecasting and Control, 5th ed., Wiley, 2015.

[10] Thaler, R. and Sunstein, C., Nudge: Improving Decisions about Health, Wealth and Happiness, Yale University Press, 2008.

[11] Fernandes, D., Lynch, J. and Netemeyer, R., Financial Literacy, Financial Education and Downstream Financial Behaviors, Management Science, vol. 60, pp. 1861–1883, 2014.

[12] Cortes, C. and Vapnik, V., Support-Vector Networks, Machine Learning, vol. 20, pp. 273–297, 1995.

[13] Breiman, L., Random Forests, Machine Learning, vol. 45, pp. 5–32, 2001.

[14] Reserve Bank of India, Master Directions on Account Aggregator Framework, RBI, Mumbai, 2021.

[15] Pawar, A. and Deshmukh, S., Design of an Intelligent Expense Analytics Platform for Student Users, National Conference on Emerging Computing Technologies, 2026 (in press).