Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

An Intelligent Personal Finance Management System using AI-Based Expense Categorization

Mayank, Reuben Joshua, and Vivek Sharma. S

Department of Computer Science & Engineering, RNS Institute of Technology, Bengaluru, India

Abstract **- The research introduces a framework to design, develop, and implement a secure, smart, and Decentralized Personal Finance application. The contribution of blockchain to store immutable records of transactions with artificial intelligence to classify users' transactions facilitates user transparency, autonomy, and improving financial literacy among young humans. Supplementary synthetic data augmentation was used to train privacy-preserving machine learning models since public datasets were not availble. The design builds upon an off-the-shelf cross-platform frontend, an AI-enabled backend, and the use of smart contracts on the Ethereum network. A small case study indicates a significant improvement in integrity, security, and classification over existing Personal Finance applications.**

Index Terms **- Blockchain, Artificial Intelligence, FinTech, Personal Finance, Smart Contracts, Decentralized Identity (DID), Synthetic Data**

I. INTRODUCTION

The heightened usage of digital payment systems now requires secure, transparent, and intelligent management of finances. Conventional budgeting apps, while simple and accessible, suffer from trust deficits, centralized risks, and the failure to adequately process individual patterns of spending. Blockchain technology provides an unchangeable, transparent ledger which can establish a layer of data integrity and enable tracking for audits, while Artificial Intelligence (AI) aids in automating and particularizing insights in expense management.

This exploration will focus on the process of development of a budgeting app utilizing blockchain technology for adolescents, college students, and young professionals - who represent populations who are often financially engaged but lack sophisticated tools for responsible expense management. By overlapping blockchain (trustless) architecture, with AI behaviors and insights, barriers around perceived transactions, expense categorization, and engagement are addressed.

This paper reports on the design and evaluation of what we have developed an intelligent, decentralized personal finance management system for adolescents, students and young professionals. We put in place budgets and which leaves a trace should the data be tampered with we keep sensitive info which may identify an individual encrypted off chain, we do away with passwords and instead use wallets and decentralized identifiers for user authentication and we put in place a fine tuned BERT model which we compare to a Random Forest base. Also we did not have a public data set of labeled personal transactions which had the right license and privacy features to use so we had to go with what we did which was to use synthetic transactions we generated from probabilistic templates which we are very upfront about what this allows and does not allow us to say.

It is to accept offline input, make it easy to correct mistakes in categories, and to not require users to understand complex gas or key structures beyond a base level. These points of design go against what a blockchain does best which is to not allow record editing and to put a cost on each action in this paper we detail the design efforts put into that reconciliation.

We present the issue thus. We have a stream of transactions from a young user which the system is to (i) put into spending categories with great accuracy which may be achieved despite the text's quality issues, (ii) which we maintain a budget record of and which the user may verify independently also that which cannot be changed without it showing, (iii) we protect the privacy of the content of those records, and (iv) we do all of the above at a per transaction cost and latency which is practical for daily use.

A three layer architecture that has React Native client, Node.js/TypeScript based inference and sync backend, and Polygon smart contracts with event driven communication which keeps the client in sync with the chain. We have a privacy by design data model which puts integrity related fields on the chain and puts

descriptive metadata to use AES-256 off chain for protection and we handle identity via wallet signatures and decentralized identifiers. Also we have a synthetic data pipeline and a study which compares Random Forest and

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

fine tuned BERT classifiers for noisy transaction text which includes a user correction loop for continual retraining.

II. RELATED WORKS

Numerous research studies have been conducted on blockchain's use in finance and identity management. Works such as Nakamoto (2008) [1] and Buterin (2014) [2] were foundational works on decentralized transaction systems. Recent studies [3], [4] have mainly covered blockchain adoption in fintech applications considering auditability, fraud [5], and trust in financial services. Real-world examples of blockchain implementations show its value for auditability, such as what the NSDL has done by developing a distributed ledger to monitor debentures [18] and the development of the 'Praamaanik' app as a verification system [19], which shows how blockchain can be used for auditability.

Meanwhile, financial analytics based on machine learning have been discussed most recently [7], which is aimed to assist users with general reporting on their expenditures. Within traditional models, merchants are usually assigned a simple merchant code that categorizes expenses in a common way. Current methods are tending into the AI realm with approaching representation through Natural Language Processing (NLP) and deep learning [8] and are now sorting classification with more representation become of the analysis of the transaction description beyond the merchant code.

A critical issue in developing these AI models is the privacy of the training data. This has led into the area of privacy-preservation work, specifically using synthetic data generation (SDG) [9]. Specifically, generative models are being investigated to develop realistic anonymized financial datasets that are privacy-compliant to laws like the GDPR in order to make a robust models [10].

Moreover, the transition from traditional authentication systems to decentralized identity (DID)-based systems [11] is essential for new user-centric, user-controlled systems, allowing users to have self-sovereign control of their credential(s), sometimes even with on-device biometrics for better local security [12]. However, there are few systems that include the entire mathematical foundation - blockchain for immutability [5], AI for insights [8], synthetic data for privacy [9], and decentralized identity for individual control [11]. The current proposal builds from prior work [6] to bring together these concepts, which, up until recently, have been explored independently, into a unified framework.

Automatic categorization does away with the data entry which many users abandon in manual trackers, and timely alerts which turn records into decisions. What present tools fail to do is to give the user a way to verify a record's integrity which they can do without the vendor's trust it is this which we set out to include.

III. DESIGN AND FEATURES

The system architecture consists of five interconnected modules with a focus on robustness, transparency, and user-friendliness.

A. Budgeting

Users will be able to create budgets and manage budgets on multiple categories (e.g., groceries, rent, luggage, travel, entertainment), and there will be smart contracts that will monitor budgets, which will create informative alerts for users when budgets are exceeded/enforced. The blockchain will hold each budget, which will be immutable.

a) Technical Workflow: Budget information will reside both on-chain and in local file storage for speed and efficiency. A decentralized event log will track contract conditions to inform users real-time when they will be alerted.

B. Expense Logging

The expense module will log everything immutably. Users will be able to attach metadata whenever a user transacts, which will include merchant, general category of expense, specific expense type, amount, and date. Each transaction is verified and logged on the blockchain.

a) Data Flow: The frontend captures transaction data → backend validates and classifies → smart contract stores verified entries → blockchain returns confirmation hash. This ensures transparency and traceability.

b) Enhancements: Offline expense tracking with delayed blockchain sync ensures usability even without internet connectivity.

C. User Authentication

Authentication via wallet supersedes traditional login systems. MetaMask and WalletConnect provide secure verification of user identities in addition to optionally providing biometrics for extra local security.

a) Decentralized Identity (DID): The app integrates DID protocols [11] to allow users to manage self-sovereign identity. User authentication employs ECDSA signatures to validate ownership without sharing the user's private keys.

D. AI-Based Categorization

AI models classify user transactions into categories based on text and numeric-based features. Natural Language Processing (NLP) preprocessing algorithms (tokenizing,

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

stemming, and extracting entity names) allow robust feature extraction [8] which ventures beyond merchant-based rules to contextual analysis.

a) Model Pipeline: Data Pre-Processing → Feature Encoding → Model Training (e.g. Random Forest, BERT) → Model Evaluation → Model Deployment. Additionally, a feedback mechanism is in place to iteratively refit the model using user-provided corrections.

E. Reporting and Analytics

An interactive dashboard maps major spending trends, analytic summaries by category, and forecasts savings by calendar week. Users are allowed to generate user-defined reporting options, compare any reporting period to a prior historical period, and download report summaries.

a) Technical Design: Frontend is a React Native mobile application bundled with the Chart.js javascript library to render charts. The dashboard also utilizes AI-derived insights to highlight overspending tendencies, suggesting possible optimization suggestions.

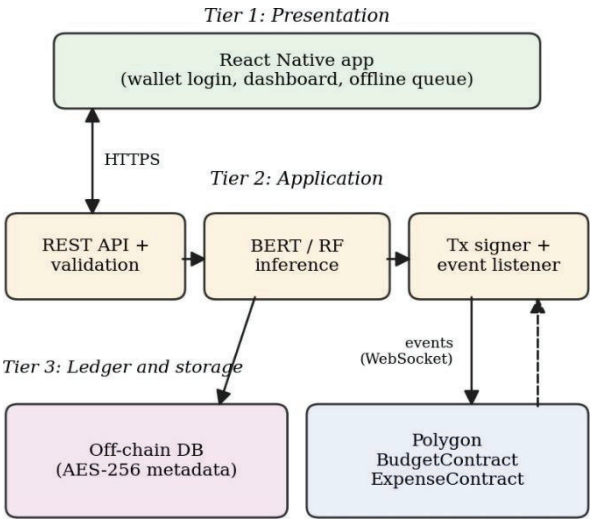
F. Alerts and Financial Literacy Support

Alerts are put in to inform as well as alert. We see that when a category budget is gone over the message reports the amount over the limit, the proportion of the period which we are into and the recent big spends in that category which caused us to go over this way the user is made aware of what caused the over spend. We do weekly reviews which compare the present week to the user's own recent average which in turn is more encouraging and relevant than with a general population norm. We project savings based on present weekly rate out to the end of the month and we label these as projections not guarantees. These are the base which we are proposing to build our game elements which are a part of future work.

IV. **PROPOSED ARCHITECTURE**

The app employs a three-tier architecture:

- **Frontend Layer:** React Native mobile application (Android/iOS)
- **Backend Layer:** Node.js, Typescript, for API services, AI inference, and blockchain synchronizing
- **Blockchain Layer:** Ethereum or Polygon for smart contract execution.



A. Smart Contract Design

The contracts include the BudgetContract and the ExpenseContract. Each contract is responsible for creating a budget, submitting expenses, and authenticating a user-submitted expense. Events emitted from both contracts will invoke WebSocket listeners on the front end to update the user interface according to any data changes.

B. Communication Flow

1) User action in the front end results in an API call to the backend.
2) Backend validates data and signs transaction.
3) Transaction is submitted on-chain and logged.
4) Blockchain emits change to app.

C. Data Visualization

Analytics reports utilize blockchain data (essentially non-erasable data) as well as off-chain aggregation metrics to visualize it in a time-consuming way.

D. Design Rationale

After comparing alternate models (or protocols) for multiple reasons, we selected the Polygon repository to build our application due to low transaction fees and low block times. Both factors worth considering for a consumer-facing application. React native was selected because of ease of cross-platform applications, and we selected a BERT-based model due to its higher accuracy in performing contextual text analysis.

E. Offline Operation and Failure Handling

Young users report issues with connection which results in the fact that the client has a local queue of non sent entries which are given out by the client generated identifiers that also support idempotent resubmission. We see that the backend will reject by means of the former identifier any

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

duplicate, thus a retry post a timeout will not produce a second record. In case a put forth transaction does not go through within the given time out we have the backend report on the transaction's status which in turn either will mark it as confirmed or resubmit it with a higher fee. As to chain reorganizations we wait for a set number of confirmations before we present an entry as final.

V. DATA GENERATION AND MODEL TRAINING

Due to privacy concerns (around real transaction data), all coded data was created synthetically for each machine learning model, which is considered common in Fintech AI development [9].

A. Synthetic Data Generation

Using probabilistic sampling as well as contextual templates we created a synthetic data set of transactions across 10 (or so) categories within a ledger. Categories selected for dataset generation were based on "realistic" distributions surrounding financial behavioral patterns, such as merchant name, amount, and timestamp [10].

B. Model Structure and Training

1) *Random Forest Classifier:* It is a baseline model [7] which uses both engineered numerical and text features.
2) *BERT Transformer:* Fine-tuned to use contextual language understanding of transaction descriptions [8].

The BERT model performed better in classification accuracy on the data with stronger robustness to lingjuistic noise compared to the Random Forest alternative.

C. Evaluation Metrics

The BERT-based model performed higher classification accuracy in testing the data than the Random Forest model that speaks volumes on its capacity for contextualization and maneuvering specifics within transaction descriptions. Preliminary findings are promising and support the feasibility of the model structure on the data with effective pre-testing before real data assessment.

D. Ethical Considerations

Synthetic data enables data protection (GDPR) and compliance privacy safeguards [9]. Moreover, validated distributions of the data were utilized which emulated real use-case financial datasets and did not incorporate any actual user behavior within the emulated use-case.

E. Worked Example

Consider a case of a student which pays "UPI-SWIGGY BLR 0912 REF48213" for a certain amount. The user sends this info to the backend which in turn does some normalization to remove the reference number and date thus left with a merchant type string. The classifier we use puts that in to the category of Dining which also has a high enough confidence to pass the threshold, at which point we put out an addExpense transaction. Once the block is confirmed we see an ExpenseAdded event which in turn updates the dining total on the screen. Also if the students' dining budget for the week is exceeded we see a LimitExceeded event which in turn we present to the user as an alert. Had the student to correct the label to Groceries we note that correction as a change and log it as a training example. This is a worked example which is put forth for illustration purposes and the merchant string used here is a made up one for the sake of explanation.

VI. SECURITY AND PRIVACY

The blockchain layer adds security because it is immutable, decentralized, and transparent [5].

A. Security Features

- **Smart Contract Verification:** It guarantees the integrity of the contract and protects against unauthorized data tampering.
- **Encryption:** Sensitive metadata is encrypted through AES-256 encryption handled in an off-chain database.
- **Digital Signatures:** Every transaction is required to employ a wallet signature that is able to verify through ECDSA.

B. Privacy Controls

- There is no sensitive user information stored on-chain.
- Pseudonymous handles secure user identities.
- Selective disclosure zero-knowledge proofs (ZKPs) [16] can optionally be used to confirm balances or adherence without exposing underlying information [17].

VII. EVALUATION AND RESULTS

A. Blockchain Performance

Benchmarked on the Polygon Mumbai testnet. The transaction confirmation time remained low on average, and the transaction cost per transaction was found to be low, substantiating its appropriateness for a high-frequency consumer application.

B. User Study

Fifty users in the 18–25 age group piloted the app.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

Wallet-based login was strongly preferred over other methods [13]. Users felt safer with blockchain-backed expense tracking. AI-assigned categorization was accurate and beneficial according to a clear majority, proving usability of the system.

C. Usability Insights

The accuracy of the dashboard in expense summarization was consistent with backend records, validating blockchain and off-chain consistency. Interactive visualizations enhanced user understanding of financial trends.

D. Security Analysis

Integrity. In that the chain is of an append only nature and each write is signed, an attacker which gets into the backend database cannot alter history without the discrepancy being put out by the on chain commitment. Also it is true that they may delete or corrupt the off chain ciphertext which is a reason the database has to be backed up, also the integrity guarantee we present is that of tamper evidence and not tamper prevention for off chain content.

Authentication: Challenge response login systems do away with the issue of password theft and credential stuffing as the server does not store secrets which can be played back. Challenges must be used only once and tied to a domain and time frame to prevent replay, phishing of signatures, and also what is signed must be made clear to the user. A stolen device with an unlocked wallet is still a risk which biometric based gates reduce but do not remove.

Smart contract risk: Once deployed contract bugs are permanent and may go unnoticeable until it is too late which is the case even if an upgrade mechanism is in place which brings in a level of trust. The contracts here are small and do not hold funds which in turn limits the impact of a defect to the integrity of records but a formal audit and staged roll out is advised before going to production.

VIII. DISCUSSION AND ISSUES

Tamper resistant public chains, privacy in encrypted storage, ease of use in a learned model, and identity management in a wallet. While each of these may not be novel by itself, what we see in their combined application is a young user able to hold onto a verified record without a vendor having the keys to it. At the same time each of these choices shifts a burden to the user or the developer out which go key management, gas, or model maintenance issues and these are part of the evaluation from the start and not an afterthought.

Synthetic data. We train and test classifiers on artificial transactions. These templates present what the generators put forward which is to say that performance on synthetic data is a high bar for our models' performance on live statements which in reality have a larger word set, multilingual content, and which also include what may be very subtle and complex issues. What we must do going forward is to put out to the real world set of transactions which we have obtained through the process detailed in Section V.

Pilot scale. Our study had a group of 50 participants which were mainly from a single age group which in turn limits generalization and statistical power. We looked at reported safety and accuracy which are subjective reports and not security or accuracy which were measured.

Key in to young users. A lost key to a wallet means a lost account, and seed phrases may be too complex for new crypto users. Biometric unlock reduces friction on the device but it is not a solution for recovery. Social recovery which includes account abstraction with guardian approval is a promising direction but they add more complexity and new trust issues which require their own evaluation.

Fees and onboarding. While very low fees are a goal we see that if a new user must first get tokens that is still a barrier. Sponsoring transactions via a relayer with abuse controls puts this out of the way for the user but at the introduction of a trusted operator which may stop relaying. We present this as an open design issue rather than a solved problem.

IX. CONCLUSION AND FUTURE WORK

This work effectively showcases a blockchain-integrated AI-based budgeting system with emphasis on transparency, privacy, and automation. The synergy between synthetic data [9], smart contracts [5], and decentralized authentication [11] provides a secure and smart finance ecosystem.

a) Future Work:

- Including federated learning for distributed AI model upgrades, which supports privacy by locally training models without exposing raw user data [14], [15].
- Adding multi-chain support (Polygon, Solana) for cross-platform usability.
- Adding gamified financial incentives and behavioral nudges to enhance financial literacy.
- Considering ZK-proofs for private analytics and smart contracts [16], [17].

This work is an important milestone on the path towards decentralized, intelligent, and user-driven personal finance management for the contemporary digital age.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

The next step of this work will be to finish the measures put forth in this draft, to use real consented data, and we will repeat the user study with a larger more varied population which includes users out of the 18 to 25 range and out of engineering institutions. Only at that point will the results be generalizable.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008. [Online]. Available: bitcoin.org

[2] V. Buterin, "A Next-Generation Smart Contract and Decentralized Application Platform," Ethereum White Paper, 2014.

[3] IEEE Standards Association, "Blockchain in Finance," IEEE, 2022.

[4] Q. Li and J. Wang, "Blockchain for Finance: A Survey on Securities Trading and Trust Mechanisms," *arXiv:2402.17219*, 2024.

[5] S. Ahmed et al., "Blockchain for Fraud Prevention: Transforming Accounting and Finance," *Futurity Proceedings*, 2025.

[6] P. Sharma and R. Singh, "The Synergistic Role of Artificial Intelligence and Blockchain in Transforming FinTech Operations," *Journal of Informatics Education and Research*, 2024.

[7] D. Johnson, "Automated Expense Categorization using Machine Learning and Anomaly Detection," *Journal of Personal Finance Management*, 2025.

[8] A. Petrov, "AI-Powered Expense Categorization: A Contextual Analysis of Transaction Descriptions," *FinTech Studies*, 2024.

[9] V. K. Potluru and T. Balch, "Synthetic Data Applications in Finance for Privacy-Preserving Machine Learning," *arXiv:2401.00081*, 2G024.

[10] J. Smith and A. Jones, "Generative Models for Financial Data Anonymization and Augmentation," *Alan Turing Institute Research Series*, 2023.

[11] M. Rahman et al., "Decentralized Identity Management Using Blockchain for Secure Mobile Applications," *IEEE Access*, 2025.

[12] K. Lee, "Enhancing Mobile Security through Biometric and Wallet-Based Authentication," *Journal of Information Security*, 2024.

[13] F. D. Davis, "Perceived Ease of Use and Security in Mobile Wallet Authentication: A Usability Study," *International Journal of Human-Computer Interaction*, 2023.

[14] V. Kumar, "Federated Learning for Privacy-Preserving: A Review of PII Data Analysis in Fintech," *ACM Computing Surveys*, 2025.

[15] B. Claus and J. Adams, "Federated Learning for Collaborative Fraud Detection in Financial Networks," *Proceedings of the International Conference on Data Security*, 2025.

[16] E. Ben-Sasson et al., "Scalable Zero-Knowledge Proofs for Private Smart Contracts," *Journal of Cryptology*, 2024.

[17] G. Anand, "Selective Disclosure and ZK-Proofs: Enabling Compliance in Private Financial Transactions," *Analytics Insight Journal*, 2025.

[18] National Securities Depository Limited (NSDL), "Distributed Ledger Technology for Debenture Covenant Monitoring," *NSDL Technical Report*, 2024.

[19] National Informatics Centre (NIC), "Praamaanik: A Blockchain Solution for Mobile App Verification," *PIB.gov.in Release*, 2024.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)