iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79467

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Personal Expense Leakage Detection and Budget Optimization using Machine Learning

Abiya F R[1], Parvathy S[2], Devende M R[3], Ajmalsha S R[4], Viji C[5]
[1, 2, 3, 4]Department of Computer Science, (Data Science), St. Thomas Institute for Science and Technology, Trivandrum, India
[5]Assistant Professor, Department of Computer Science, St. Thomas Institute for Science and Technology, Trivandrum, India

Abstract: Effective personal money management is increasingly challenged by expense leakage, which refers to the gradual and often unnoticed loss of money through unused subscriptions, repeated small transactions, and irregular spending behavior. Most existing budgeting tools focus on summarizing expenses after they occur and provide limited support for identifying hidden or inefficient spending patterns. The primary aim of this project is to assist individuals in managing their finances more effectively by detecting and reducing unnecessary expenses. This project presents a Personal Expense Leakage Detection and Budget Optimization system developed using Python and machine learning techniques, employing a dual-layer unsupervised learning framework in which the Isolation Forest algorithm is used to identify abnormal transactions such as unexpected charges and billing inconsistencies, while K-Means clustering groups frequent low-value transactions that may be overlooked individually but have a significant cumulative impact. The system further incorporates features such as identification of unused recurring subscriptions, prediction of end-of-month balance based on current spending trends, analysis of behavioral spending patterns to reduce impulsive purchases, and visualization of the long-term financial impact of small recurring expenses. Experimental evaluation using synthetic transaction data demonstrates that the proposed system is more effective than traditional rule-based budgeting methods in detecting hidden spending patterns, indicating that the integration of machine learning and behavioral analysis into personal finance tools can significantly improve money management, reduce financial waste, and support long-term financial stability.

Keywords: Expense leakage detection, Budget optimization, Machine learning, Isolation Forest, K-Means clustering, Anomaly detection, Financial data analysis, Spending behavior analysis

I. INTRODUCTION

Personal financial management has become increasingly important in modern society due to rising living costs and the widespread use of digital payment systems. Individuals often struggle to track their expenses effectively, leading to poor budgeting and financial instability.

One of the major challenges in personal finance is expense leakage, which refers to the gradual loss of money through unnoticed small transactions,unused subscriptions, and irregular spending habits. These expenses may appear insignificant individually but can accumulate over time and significantly impact overall financial health.

Traditional budgeting tools and expense tracking applications mainly focus on recording and summarizing past transactions. However, they provide limited support in identifying hidden spending patterns or predicting future financial outcomes. As a result, users may not be fully aware of inefficient spending behaviors or opportunities to optimize their budget. With the advancement of data analytics and machine learning, it is now possible to develop intelligent systems that go beyond basic tracking and provide deeper insights into financial behavior.

This project proposes a Personal Expense Leakage Detection and Budget Optimization system that leverages machine learning techniques to analyze transaction data and identify hidden expense patterns. The system uses the Isolation Forest algorithm to detect abnormal transactions such as unexpected or irregular spending. In addition, K-Means clustering is applied to group frequent low-value transactions that contribute to expense leakage. By combining anomaly detection and clustering techniques, the system provides a comprehensive analysis of spending behavior.

Furthermore, the system includes features such as detection of unused subscriptions, prediction of end-of-month balance, and visualization of spending patterns. These features help users understand their financial habits and make informed decisions to reduce unnecessary expenses. The proposed solution aims to improve financial awareness, promote better budgeting practices, and support long-term financial stability through intelligent data-driven analysis

A. Research Problem

The increasing use of digital payment systems and online transactions has made personal financial management more complex. While individuals have access to various expense tracking tools, most of these systems focus only on recording and summarizing past transactions. They fail to identify hidden spending patterns such as recurring small expenses, unused subscriptions, and irregular spending behavior, which collectively contribute to expense leakage. As a result, users often remain unaware of inefficient spending habits, leading to poor budgeting and financial instability over time.

Furthermore, existing systems lack intelligent analytical capabilities to detect abnormal transactions and provide meaningful insights into user spending behavior. The absence of machine learning techniques limits their ability to analyze large volumes of financial data and identify patterns that are not easily visible. There is also a lack of predictive features that can help users anticipate future expenses and manage their budgets effectively. These challenges highlight the need for a data-driven system that can automatically detect expense leakage, identify unusual spending patterns, and support informed financial decision-making.

B. Purpose of the Study

The purpose of this study is to develop a system that analyzes personal transaction data to detect expense leakage and optimize budget usage. The system aims to identify abnormal transactions using Isolation Forest and group spending patterns using K-Means clustering. It also provides insights into recurring expenses, unused subscriptions, and future spending trends to help users make better financial decisions.

C. Contribution

This project presents a machine learning-based system for detecting expense leakage and optimizing personal budgets. It integrates Isolation Forest for anomaly detection and K-Means clustering for analyzing spending patterns. The system also includes features such as identification of recurring expenses, subscription tracking, and budget prediction. The proposed solution provides a simple, efficient, and scalable approach for improving financial awareness and decision-making.

D. Motivation

The increasing use of digital payments and subscriptions has made it difficult for individuals to track their expenses effectively. Small, frequent transactions and unused services often go unnoticed, leading to financial loss over time. This project is motivated by the need to improve financial awareness and provide a smart system that helps users identify hidden expenses and manage their budgets more efficiently using data-driven techniques.

E. Paper Organization

The remainder of this paper is organized as follows. Section II presents the related work in the field of expense analysis and machine learning techniques. Section III describes the methodology of the proposed system. Section IV discusses the results and analysis. Finally, Section V concludes the paper and outlines future work.

II. RELATED WORK

A. Background

Personal finance management has gained significant attention with the increasing use of digital transactions and online payment systems. Many existing tools help users track expenses and generate summaries, but they often lack the ability to detect hidden spending patterns such as recurring small expenses and unused subscriptions. As a result, users may experience expense leakage without clear awareness.

Recent studies have explored the use of machine learning techniques in financial data analysis to improve expense tracking and decision-making. Anomaly detection methods are used to identify unusual transactions, while clustering techniques help group spending patterns and understand user behavior. These approaches provide deeper insights compared to traditional rule-based systems.

However, most existing solutions focus either on anomaly detection or basic expense tracking and do not provide a complete system that integrates multiple techniques for expense leakage detection and budget optimization.

This project addresses these limitations by combining anomaly detection and clustering methods into a single framework for effective financial analysis.

B. Machine Learning in Financial Analysis

Machine learning techniques have been widely used in financial data analysis to improve expense tracking and decision-making. Algorithms such as anomaly detection and clustering help identify unusual transactions and group spending patterns. Isolation Forest is commonly used for detecting outliers in financial data, while K-Means clustering is effective in identifying behavioral patterns. These techniques provide deeper insights compared to traditional rule-based systems.

C. Limitations of Existing Systems

Existing expense tracking systems primarily focus on recording and summarizing user transactions, but they lack advanced analytical capabilities to identify hidden spending patterns. These systems typically provide basic reports such as total expenses or category-wise breakdowns without detecting recurring small transactions or unused subscriptions that contribute to expense leakage. As a result, users may not be aware of inefficient spending habits, leading to poor budget management and financial loss over time.

In addition, most existing systems do not incorporate intelligent techniques such as anomaly detection or behavioral analysis. They are unable to identify unusual transactions, predict future spending trends, or provide actionable insights for budget optimization. Furthermore, the absence of machine learning integration limits their ability to adapt to dynamic financial behavior. These limitations highlight the need for a more advanced system that combines multiple analytical approaches to effectively detect expense leakage and support better financial decision-making.

III. METHODOLOGY

A. System Overview

The proposed Personal Expense Leakage Detection and Budget Optimization system is designed to analyze financial transaction data and identify hidden spending patterns using machine learning techniques. The system follows a structured and data-driven approach that integrates multiple stages, including data collection, preprocessing, normalization, anomaly detection, clustering, and visualization. The primary objective of the system is to help users detect expense leakage, which often occurs due to small recurring expenses, unused subscriptions, and irregular spending behavior that may not be easily noticeable.

The system begins with the collection of transaction data, which is typically stored in a structured dataset format such as CSV. This dataset contains various attributes, including transaction amount, category, frequency, and subscription details. These attributes provide the necessary information required for analyzing user spending behavior. The collected data is then passed through a preprocessing stage, where it is cleaned and prepared for further analysis. This includes handling missing values, removing inconsistencies, and selecting relevant features that contribute to effective financial analysis.

After preprocessing, the dataset undergoes normalization using the StandardScaler technique. This step ensures that all features are brought to a common scale, preventing any improves the performance of machine learning algorithms and ensures more accurate and stable results. Once the data is normalized, it is ready to be processed by the core analytical components of the system.

The system employs a dual-layer machine learning approach to analyze transaction data. The first layer uses the Isolation Forest algorithm for anomaly detection. This algorithm identifies abnormal transactions by isolating data points that differ significantly from the normal spending pattern. These anomalies may include unexpected high-value expenses, duplicate transactions, or irregular subscription charges. Detecting such anomalies helps users identify potential financial risks and unusual spending activities.

The second layer of the system applies K-Means clustering to group transactions based on similarity in spending behavior. This clustering process divides the dataset into multiple groups, where each group represents a specific spending pattern. For example, clusters may represent frequent small expenses, moderate spending, or occasional high-value transactions. This helps in identifying hidden patterns in user spending, particularly recurring low-value expenses that contribute to expense leakage over time.

The outputs from both the anomaly detection and clustering modules are combined to generate meaningful insights about the user's financial behavior. The system analyzes these results to identify recurring expenses, unused subscriptions, abnormal transactions, and overall spending distribution. These insights are essential for improving financial awareness and helping users make better budgeting decisions.

Finally, the system presents the results through a visualization layer. Graphical representations such as charts, graphs, and dashboards are used to display the analysis in an easy-to-understand format. This allows users to quickly interpret their financial data and take appropriate actions to reduce unnecessary expenses. Additionally, the system may include predictive features such as estimating future expenses, based on current spending trends.

Overall, the system provides a comprehensive and intelligent solution for personal financial management by combining machine learning techniques with data analysis and visualization. It not only detects expense leakage but also empowers users to optimize their budgets and achieve better financial stability.

B. Data Collection and Preprocessing

The system uses a transaction dataset stored in CSV format, which contains attributes such as transaction amount, category, frequency, and subscription-related information.

The dataset may include both normal and irregular transactions to simulate real-world financial scenarios. During preprocessing, the data is cleaned by handling missing values, removing inconsistencies, and converting data into a suitable format for analysis. Feature selection is also performed to extract important attributes that contribute to expense analysis. This step ensures that only relevant information is used for further processing, improving the efficiency and accuracy of the machine learning models.

C. Data Normalization

Data normalization is an important preprocessing step used to scale the features of the dataset to a common range. In financial transaction data, different features such as transaction amount, frequency, and category values may have varying scales. Without normalization, features with larger values may dominate the analysis and affect the performance of machine learning models.

In this system, normalization is performed using the StandardScaler technique, which transforms the data to have zero mean and unit variance. This ensures that all features contribute equally during model training and prevents bias toward any particular feature. Normalization also helps improve the stability and efficiency of algorithms such as Isolation Forest and K-Means clustering.

By applying normalization, the system achieves better accuracy in detecting anomalies and grouping transactions into meaningful clusters. It enhances model performance and ensures consistent results across different datasets. Overall, data normalization plays a crucial role in preparing the dataset for effective machine learning analysis.

D. Anomaly Detection using Isolation Forest

The Isolation Forest algorithm is used in the proposed system to detect abnormal transactions in financial data. It is an unsupervised machine learning technique designed specifically for anomaly detection, where the goal is to identify data points that significantly differ from normal behavior. Unlike traditional methods, it works by isolating anomalies through random partitioning of the dataset.

The algorithm constructs multiple decision trees by randomly selecting features and split values. Each data point is recursively partitioned until it is isolated, and the number of splits required to isolate a point is called the path length. Anomalies generally have shorter path lengths because they are easier to separate from the rest of the data, while normal data points require more splits.

In this system, the algorithm analyses features such as transaction amount, frequency, and category to identify unusual spending patterns. Transactions with higher anomaly scores are labelled as abnormal, helping to detect unexpected expenses, duplicate payments, or irregular charges. Overall, Isolation Forest provides an efficient and effective method for identifying anomalies and improving financial decision-making.

E. Clustering using K-Means

K-Means clustering is used in the proposed system to group transactions based on similarity in spending behaviour. It is an unsupervised machine learning algorithm that partitions the dataset into a predefined number of clusters by minimizing the distance between data points within the same cluster. The goal is to form meaningful groups that represent different patterns in financial data.

The algorithm begins by selecting initial cluster centroids and assigning each data point to the nearest centroid using a distance measure such as Euclidean distance. After all data points are assigned, the centroids are recalculated as the mean of the points in each cluster. This process is repeated iteratively until the clusters stabilize. The final clusters represent distinct spending behaviours present in the dataset.

In this system, K-Means is applied to normalized transaction data using features such as transaction amount, frequency, and category. The clusters help identify patterns like frequent low-value expenses and occasional high-value transactions, which contribute to expense leakage. Although the algorithm is efficient and easy to implement, it requires selecting an appropriate number of clusters and may be affected by outliers. Overall, K-Means improves the understanding of spending patterns and supports better financial decision-making.

F. Result Generation and Visualization

The result generation and visualization module plays a crucial role in transforming the outputs of machine learn
ing models into meaningful and interpretable insights for users. After processing the transaction data through anomaly detection and clustering algorithms, the system consolidates the results to provide a comprehensive analysis of user spending behavior and expense leakage.

The first stage of result generation involves collecting outputs from both the Isolation Forest and K-Means models. The Isolation Forest algorithm produces anomaly labels and scores for each transaction, indicating whether a transaction is normal or abnormal. Transactions identified as anomalies represent unusual spending patterns such as unexpected high-value expenses, duplicate charges, or irregular subscription payments. These anomalies are highlighted as potential areas of concern for the user.

Simultaneously, the K-Means clustering algorithm groups transactions into distinct clusters based on similarity in spending behavior. Each cluster represents a specific pattern, such as frequent low-value transactions, moderate regular expenses, or occasional high-value purchases. These clusters provide insights into how different types of expenses contribute to overall financial behavior. By analyzing cluster characteristics, the system identifies recurring spending patterns that may lead to expense leakage over time.

After obtaining results from both models, the system integrates these outputs to generate meaningful insights. For example, transactions that are both part of a frequent cluster and exhibit abnormal characteristics can be flagged as critical expense leakage points. The system also identifies recurring transactions and unused subscriptions, which are common sources of unnecessary financial loss.

The visualization component is designed to present these insights in an intuitive and user-friendly manner. Graphical representations such as pie charts, bar graphs, and line charts are used to display spending distribution across different categories, trends over time, and the proportion of normal versus abnormal transactions. These visualizations help users quickly understand their financial patterns without requiring technical knowledge.

In addition, dashboards are used to provide a consolidated view of financial insights. The dashboard may include key indicators such as total expenses, identified anomalies, cluster distribution, and predicted future spending trends. These interactive visual elements allow users to explore their data and gain deeper insights into their financial habits.

The system may also include predictive analysis, such as estimating the end-of-month balance based on current spending patterns. This helps users plan their expenses and avoid overspending. Overall, the result generation and visualization module enhances the usability of the system by converting complex analytical outputs into clear, actionable insights.

By combining anomaly detection, clustering, and visualization, the system provides a powerful tool for identifying hidden expense leakage and supporting effective budget optimization. This ensures that users can make informed financial decisions and improve their overall financial stability.

Table I
Comparative Analysis of Existing Systems

Title	Method Used	Advantage	Disadvantage
Personal Finance Management using Machine Learning	Classification & basic ML algorithms	Automates expense tracking and categorization	Cannot detect hidden expense patterns.
Anomaly Detection in Financial Transactions	Isolation Forest	Effectively identifies abnormal transactions.	Does not analyze long-term behavior.
Consumer Spending Analysis using Clustering	K-Means Clustering	Groups similar spending patterns.	Not focused on budget optimization.
Smart Budgeting and Expense Prediction Systems	Time-series / Trend analysis.	Predicts future expenses.	Cannot detect expense leakage.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

IV. IMPLEMENTATION

A. System Workflow

The system workflow describes the complete process through which the proposed Personal Expense Leakage Detection and Budget Optimization system analyzes financial transaction data and generates meaningful insights. The workflow begins with the input of transaction data, which is typically provided in the form of a dataset containing details such as transaction amount, category, frequency, and subscription-related information. This dataset serves as the foundation for all further analysis performed by the system.

Once the data is provided, it undergoes a preprocessing stage where it is cleaned and prepared for analysis. During this step, missing values are handled, inconsistent entries are corrected, and irrelevant features are removed. Feature selection is also performed to extract important attributes that contribute to expense analysis, such as normalized amount, transaction frequency, and category information. This ensures that the dataset is structured and suitable for machine learning processing.

After preprocessing, the dataset is normalized using the StandardScaler technique. Normalization transforms the data into a standard scale with zero mean and unit variance, ensuring that all features contribute equally to the analysis. This step is essential because machine learning algorithms are sensitive to variations in data scale, and normalization improves the accuracy and stability of the models. The normalized data is then processed using two parallel machine learning approaches. The first approach involves anomaly detection using the Isolation Forest algorithm. This algorithm works by isolating data points that differ significantly from the majority of the data. Transactions that require fewer splits to isolate are considered anomalies. These anomalies may represent unusual or suspicious transactions, such as unexpected high expenses, duplicate payments, or irregular subscription charges.

Simultaneously, the second approach applies K-Means clustering to group transactions based on similarity in spending behavior. The algorithm divides the dataset into clusters, where each cluster represents a specific spending pattern. For example, one cluster may represent frequent small transactions, while another may represent occasional high-value expenses. This clustering helps identify hidden patterns in user spending, especially recurring small expenses that contribute to expense leakage over time.

After both anomaly detection and clustering processes are completed, the results are combined to provide a comprehensive analysis of the user's financial behavior. The system identifies key insights such as abnormal transactions, recurring expenses, and potential areas of unnecessary spending. These insights are crucial for helping users understand their financial habits and take corrective actions. Finally, the processed results are presented through a visualization layer. The system generates graphical outputs such as pie charts, line graphs, and dashboards that display spending distribution, trends over time, and detected anomalies. These visualizations make it easier for users to interpret the data and gain a clear understanding of their financial status. Additionally, the system may provide predictive insights, such as estimating the end-of-month balance based on current spending patterns. Overall, the workflow ensures a systematic and efficient approach to analyzing financial data, combining preprocessing, machine learning, and visualization techniques to deliver accurate and meaningful results for expense leakage detection and budget optimization.

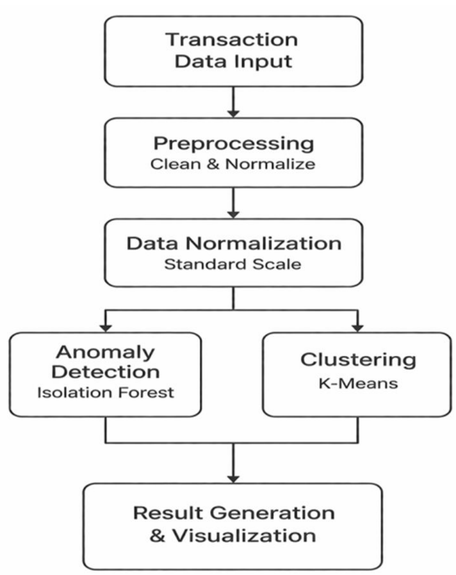

Fig 1. Proposed System Workflow

B. Challenges and solutions

The development of the proposed system involves several challenges related to data quality and preprocessing. Financial datasets may contain missing values, inconsistent entries, and irrelevant features, which can affect the performance of machine learning models. To address this issue, data cleaning techniques such as handling missing values, removing inconsistencies, and selecting relevant features are applied. This ensures that the dataset is accurate and suitable for analysis.

Another major challenge is the use of unsupervised learning techniques, where labelled data is not available for evaluation. This makes it difficult to measure performance using traditional metrics such as accuracy or precision. To overcome this, the system focuses on identifying meaningful anomalies and useful spending patterns. Visualization and practical relevance of the results are used as key indicators to evaluate system effectiveness.

The system also faces challenges related to parameter selection and scalability. Algorithms like K-Means require the number of clusters to be defined, and improper selection may affect results. Similarly, Isolation Forest requires tuning of parameters for optimal performance. These challenges are addressed through experimentation and testing with different configurations. Additionally, efficient libraries and optimized processing techniques are used to ensure scalability and better performance with larger datasets.

C. Testing environment

The proposed system is developed and tested using a Python-based environment with commonly used data science libraries such as Pandas, NumPy, Scikit-learn, and Matplotlib. The implementation is carried out using tools like Jupyter Notebook or Visual Studio Code, which provide an efficient platform for data analysis, model development, and visualization. These tools enable smooth integration of preprocessing, machine learning, and result generation modules.

The system is tested using a transaction dataset in CSV format, containing features such as transaction amount, category, frequency, and subscription details. Synthetic and sample datasets are used to simulate real-world financial scenarios. The dataset is preprocessed and normalized before being passed to the machine learning models to ensure accurate and consistent results.

The testing process includes evaluating the performance of both the Isolation Forest and K-Means algorithms. The system is checked for its ability to detect anomalies and identify meaningful spending patterns. Visualization outputs such as charts and graphs are analysed to verify the correctness of results. Overall, the testing environment ensures that the system functions effectively and produces reliable insights for financial analysis.

V. RESULTS

A. Evaluation Metrics

The proposed system is based on unsupervised machine learning techniques, and therefore traditional supervised evaluation metrics such as accuracy, precision, recall, and F1-score are not directly applicable. Since the system does not rely on labeled data, its performance is evaluated based on the effectiveness of anomaly detection, clustering quality, and the usefulness of generated insights.

For anomaly detection, the evaluation focuses on how effectively the Isolation Forest algorithm identifies transactions that deviate significantly from normal spending behavior. Transactions flagged as anomalies are examined to determine whether they represent unusual or suspicious financial activities, such as unexpected high-value expenses, duplicate transactions, or irregular subscription charges. The relevance and correctness of these detected anomalies serve as an important indicator of system performance.

For clustering, the evaluation is based on how well the K-Means algorithm groups transactions into meaningful clusters. The quality of clustering is assessed by analyzing whether transactions within the same cluster share similar characteristics, such as spending frequency and amount. Effective clustering results in clear separation between different spending patterns, enabling better understanding of user financial behavior.

Visualization also plays a significant role in evaluation. The system generates graphical outputs such as pie charts, bar graphs, and trend lines that represent spending distribution and patterns over time. These visualizations help in verifying the correctness of the results and ensure that the detected patterns are meaningful and interpretable. Additionally, the practical usefulness of the insights generated by the system, such as identification of recurring expenses and potential savings opportunities, is considered a key factor in evaluating overall performance.

B. Performance Analysis

The performance of the proposed system demonstrates its effectiveness in detecting expense leakage and analyzing user financial behavior. The Isolation Forest algorithm successfully identifies abnormal transactions that deviate from normal spending patterns, such as unexpected high-value expenses and irregular payments.

These anomalies help in highlighting potential financial risks and unusual activities that may otherwise go unnoticed. The model efficiently processes transaction data and provides meaningful outputs without the need for labeled datasets.

The K-Means clustering algorithm further enhances system performance by grouping transactions into distinct clusters based on similarity. These clusters reveal important spending patterns, such as frequent low-value expenses, moderate recurring costs, and occasional high-value transactions. This grouping enables the identification of hidden expense leakage caused by repeated small transactions, which have a significant cumulative impact over time. The clustering results are consistent and provide a clear understanding of user spending behaviour.

In addition, the system performs efficiently in terms of computation and scalability. The use of optimized libraries such as Scikit-learn ensures faster processing even for larger datasets. The integration of visualization techniques improves the interpretation of results, allowing users to easily understand financial insights. Overall, the system provides accurate, reliable, and practical results, making it more effective than traditional rule-based expense tracking methods.

VI. DISCUSSION

A. Impacts

The proposed Personal Expense Leakage Detection and Budget Optimization system has a significant impact on improving personal financial management by introducing intelligent, data-driven analysis. Unlike traditional expense tracking systems that only record transactions, the proposed system actively identifies hidden spending patterns and abnormal transactions. This enables users to gain deeper insights into their financial behaviour and take corrective actions to reduce unnecessary expenses. As a result, users can manage their finances more effectively and avoid gradual financial loss caused by unnoticed spending.

Another important impact of the system is the enhancement of financial awareness among users. By analysing transaction data and presenting it through clear visualizations, the system helps users understand where their money is being spent. It highlights recurring expenses, unused subscriptions, and frequent small transactions that often go unnoticed. This increased awareness encourages responsible financial behavior and promotes better budgeting practices, ultimately leading to improved financial discipline and savings.

Furthermore, the integration of machine learning techniques such as Isolation Forest and K-Means clustering provides a more advanced and scalable solution for financial analysis. The system not only detects anomalies but also identifies long-term spending patterns, offering a comprehensive understanding of financial habits. This approach can be extended to real-world applications such as banking systems and financial management platforms. Overall, the system contributes to improved financial stability, smarter decision-making, and the development of intelligent digital financial solutions.

B. Advantages

The proposed system offers several advantages over traditional expense tracking methods by incorporating machine learning techniques for deeper financial analysis. It effectively detects abnormal transactions using the Isolation Forest algorithm, enabling users to identify unusual expenses such as unexpected charges or irregular spending patterns. Additionally, the use of K-Means clustering helps in grouping transactions based on spending behavior, making it easier to identify recurring small expenses and hidden patterns that contribute to expense leakage. This provides users with a more comprehensive understanding of their financial habits.

Another major advantage of the system is its ability to present results through clear and interactive visualizations. Graphs, charts, and dashboards simplify complex financial data and make it easily interpretable for users. The system also supports better decision-making by providing insights into spending trends and potential savings opportunities. Furthermore, since the system is based on unsupervised learning, it does not require labeled data and can be applied to various types of financial datasets. Overall, the system enhances financial awareness, improves budgeting efficiency, and helps users achieve better control over their expenses.

C. Limitations

Despite the advantages and positive impact of the proposed system, several limitations exist that affect its overall performance and practical applicability. One of the primary limitations is the dependency on the quality of the input dataset. Since the system relies on transaction data for analysis, any inaccuracies, missing values, or inconsistencies in the dataset can significantly affect the results. Poor data quality may lead to incorrect anomaly detection or misleading clustering results. Another major limitation is the use of unsupervised learning techniques, which do not rely on labeled data. While this approach is suitable for detecting patterns and anomalies in unknown datasets, it also introduces challenges in performance evaluation.

Unlike supervised models, where accuracy and other metrics can be calculated, unsupervised models require qualitative evaluation based on interpretation and usefulness of results. This makes it difficult to measure performance quantitatively.

The system also faces challenges related to parameter selection and model tuning. Algorithms such as K-Means require the number of clusters to be defined in advance, and selecting an inappropriate number of clusters can lead to poor grouping of data. Similarly, the effectiveness of the Isolation Forest algorithm depends on parameters such as contamination rate and number of estimators. Improper tuning of these parameters can reduce the accuracy of anomaly detection.

Another limitation is the lack of real-time processing capability. The current system operates on static datasets and does not support continuous monitoring of financial transactions. In real-world scenarios, real-time analysis is important for detecting unusual transactions immediately and preventing potential financial loss. The absence of real-time functionality limits the system's applicability in dynamic environments.

Additionally, the system is not integrated with real-world financial systems or banking APIs. This means that users must manually provide transaction data, which may reduce convenience and scalability. Integration with secure financial platforms would significantly improve the usability and effectiveness of the system.

D. Future Directions

Future work on the proposed system can focus on enhancing its capabilities, improving performance, and enabling real-world deployment. One of the most important directions is the implementation of real-time data processing. By integrating streaming data pipelines, the system can analyze transactions as they occur and provide instant alerts for abnormal spending behavior. This would greatly improve the system's usefulness in real-world financial management.

Another key direction is the incorporation of advanced machine learning techniques. While the current system uses unsupervised algorithms, future versions can explore hybrid approaches that combine supervised and unsupervised learning. This could improve the accuracy and reliability of anomaly detection and allow for better classification of transactions.

The system can also be enhanced by integrating with banking systems and financial applications through secure APIs. This would enable automatic data collection and eliminate the need for manual data input. Such integration would provide users with a seamless experience and allow continuous monitoring of financial activities.

In addition, the development of a mobile application interface can improve accessibility and usability. A mobile-based system would allow users to track and analyze their expenses on the go, making financial management more convenient. Scalability can also be improved by optimizing the system to handle large datasets and multiple users simultaneously.

Finally, future research can focus on improving model interpretability and transparency. Providing clear explanations for detected anomalies and clusters will help users better understand the system's outputs and build trust in its recommendations. Overall, these future enhancements aim to transform the system into a comprehensive, real-time, and intelligent financial management solution.

VII. CONCLUSION

This project presented a Personal Expense Leakage Detection and Budget Optimization system that utilizes machine learning techniques to analyze financial transaction data. The system effectively detects abnormal transactions using the Isolation Forest algorithm and identifies spending patterns through K-Means clustering. It helps uncover hidden expense leakage caused by recurring small transactions and unused subscriptions, which are often overlooked in traditional expense tracking systems.

The experimental results indicate that the proposed system provides more meaningful insights compared to conventional methods by combining anomaly detection, clustering, and visualization techniques. The system enables users to better understand their spending behavior and take corrective actions to reduce unnecessary expenses.

In future work, the system can be enhanced by incorporating real-time data processing, advanced predictive models, and integration with mobile applications to provide more personalized financial recommendations. Overall, the proposed solution contributes to improved financial awareness, efficient budget management, and long-term financial stability.

REFERENCES

[1] M. Sakthivel, P. Roshini, K. Roja, P. Maha Lakshmi, and V. Keerthi, "Personal Expense Tracker Application, " International Journal of Research Trends and Innovation (IJRTI), Chennai, India, 2023.

[2] K. Soundharya, J. Abdul Baasith, A. Mohammed Ayaan, and R. Elangovan,"Personal Expense Tracker," Journal of Emerging Technologies and Innovative Research (JETIR)