iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XI **Month of publication:** November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75387

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Automated Expense Classifier from Bank Statements

Visalatchi A[1], Yuvaraj J[2], Giri Prasanna K[3]
Student, Department of Artificial Intelligence and Data Science, SRMV alliammai Engineering College, Chengalpattu, India,

***Abstract:** TheAutomated Expense Classifier is a smart finance management application that leverages Natural Language Processing (NLP) and Machine Learning algorithms to automatically categorize expenses from bank statements in PDF/CSV formats. The project uses a TF-IDF vectorizer for text preprocessing and a Logistic Regressionmodel(trainedonlabelledtransactiondata)foraccurateexpenseclassificationintocategoriessuchas food, travel, shopping, and utilities. Data visualization techniques including matplotlib and Streamlit charts are usedtogeneratepiechartsandmonthlytrendanalysisforbetterfinancialinsights.Toensurefinancialdiscipline, the system integrates budget alert logic with real-time email notificationswhen the spending exceedspredefined limits. The project further incorporatesAI-powered forecasting using time-series trend analysis to predict next month's topspendingcategoriesandarecommendationenginethatprovidespersonalizedsuggestionstooptimize savings. An AI Chatbot Assistant, built using Lang Chain and OpenAI, is integrated for interactive financial guidance, while the frontend features a modern animated gradient background theme for enhanced user experience. Overall,thisprojectcombinesdatapreprocessing,machinelearning,NLP,visualization,forecasting,andchatbottechnologiestodelivera robust,intelligent,anduser-friendlypersonalexpensemanagementsolution.*

I. INTRODUCTION

In the current era of digitalization, individuals increasingly rely on online banking, credit cards, andmobilepaymentplatformsformanagingday-to- day financial transactions, resulting in a substantial volumeofdatathatisdifficulttomonitormanually. Traditional methods of expense tracking, such as maintaining ledgers or spreadsheets, are not only labour - intensive but also prone to errors and provide limited analytical insights, making it challenging for users to understand their spending patterns or plan future budgets effectively. To overcomethesechallenges,theAutomatedExpense Classifier has been developed as an intelligent system that integrates Artificial Intelligence (AI), Machine Learning (ML), and Natural Language Processing (NLP)to automatethecategorizationand analysis of financial transactions. The system is capable of extracting transactional data from PDF and CSV bank statements, preprocessing the data using advancedNLPtechniques,andclassifyingexpenses into multiple predefined categories such as food, travel, shopping, utilities, and entertainment. A Logistic Regression model trained with TF-IDF vectorization is employed to achieve high accuracy in text-based classification, while additional preprocessing ensures the model handles unstructured and noisy data effectively. Beyond classification, the platform offers visualization of financial data through pie charts, monthly trend graphs,andstatisticalsummariesthatallowusersto easily monitor and analyse their spending habits. The system also incorporates budget alert functionality, sending real-time notifications via email whenever spending exceeds predefined thresholds,therebypromotingdisciplinedfinancial behaviour.Furthermore,anAI-poweredforecasting mechanismpredictspotentialfutureexpensesbased on historical trends, and a recommendation engine provides personalized suggestions for optimizing spending and improving savings. To enhance user interaction and accessibility, an intelligent chatbot assistantpoweredbyLangChainandOpenAIallows users to obtain instant financial guidance and answerstoqueriesinaconversationalmanner,while the user interface is designed with a modern animated gradient theme to improve engagement and overall usability. By combining machine learning, predictive analytics, NLP, visualization, and interactive AI, the Automated Expense Classifier provides a comprehensive, scalable, and user-friendly solution for personal financial management, enabling users to monitor expenses efficiently, plan budgets accurately, and make informed decisions to achieve better financial stability. This project contributes to the growing domain of AI-driven financial applications and demonstrates the practical potential of integrating multiple intelligent technologies to simplify complexreal-worldproblems,ultimatelyenhancing personal financial literacy, accountability, and strategic money management.

II. LITERATURE SURVEY

The application of Artificial Intelligence (AI) and Machine Learning (ML) in personal finance management has seen significant growth in recent years, aiming to enhance the accuracy, efficiency, and usability of expense tracking systems. Traditionalmanualmethodsofexpensemonitoring, including ledger-based or spreadsheet-based tracking,arepronetohumanerror,time-consuming, and often fail to provide actionable insights for financial planning, motivatingresearchers to explore intelligent automated solutions. Aishwarya and Hemalatha (2023) developed a system using machine learning algorithms to predict personal expenses, offering insights into spending patterns and aiding users in planning their finances effectively.Al-Sayed and Cheng (2022) proposed a classificationframeworkthatutilizesbank transaction data with ML techniques to categorize expenditures and understand user behavior, thereby supporting improved financial decision-making. Dingetal.(2020)focusedonautomatingaccounting processes by applying ML for bank statement classification, demonstrating increased efficiency and accuracy in handling transactional data. Guida (2025)examinedAI'simpactonspendclassification in buyer firms, highlighting the transformative potential of intelligent systems in automating financial processes. Further, studies by Zhang et al. (2021) explored the use of Natural Language Processing(NLP)andTF-IDFvectorizationfortext- based transaction categorization, achieving higher precision in classifying unstructured financial data. Kumar and Patel (2022) implemented predictive models combining Logistic Regression, Random Forest, and Support Vector Machines (SVM) for expense forecasting and anomaly detection, illustrating the effectiveness of hybrid ML approaches. Singh and Verma (2023) demonstrated the integration of Optical Character Recognition (OCR)withMLmodelstoextracttransactiondetails from bank statements and receipts, ensuring scalability and automation in expense management systems. Additionally, Li and Wong (2020) emphasized the roleofAI-powered recommendation engines in suggesting personalized budgeting strategies based on user spending behavior, while Chen et al. (2021) investigated time-series analysis for financial forecasting, providing predictive insights for future expenses. Collectively, these studies establish a robust foundation for the development of intelligent, automated expense management systems, highlighting critical aspects suchasdatapreprocessing,ML-basedclassification, predictive analytics, NLP integration, OCR-based extraction, and recommendation systems, which together inform the design of the proposed AutomatedExpenseClassifier.Bybuildingupon these advancements, the system aims to offer real-time expense categorization, interactive visualizations, AI-powered forecasting, budget alerts, and a smart chatbot assistant, ultimately providing a comprehensive, scalable, and user- friendlysolutionforpersonalfinancialmanagement and decision-making.

III. PROPOSED WORK

The proposed work for the Automated Expense Classifier focuses on developing anintelligentsystemthatcanautomatically process and categorize financial transactions, provide meaningful visualizations, send budget alerts, forecast future expenses, and offer personalized financial recommendations. The system aims to simplify personal finance management by leveraging Artificial Intelligence(AI),MachineLearning(ML), Natural Language Processing (NLP), and predictiveanalytics.Userscanuploadbank statementsinPDForCSVformats,andthe system extracts transaction details using advanced preprocessing techniques. These transactions are then classified into predefined categories using a trained machine learning model. To help users understand their spending habits, the system provides interactive visualizations such as pie charts and monthly trend graphs. Additionally, it includes budget monitoringand alert notifications,AI-based expenseforecasting,andarecommendation engineforsmarterfinancialplanning.An AI-powered chatbot assistant allows users to interact with the system and get instant financial guidance, while a modern animated gradient interface ensures an engaging user experience. Overall, the proposedsystemintegratesmultipleAIand ML technologies to deliver a comprehensive, scalable, and user-friendly personal finance management solution.

Module Description

1) Data Input and Extraction Module

TheDataInputandExtractionModuleservesasthe first step in the system, allowing users to upload their bank statements in PDF or CSV formats. This module ensures compatibility with multiple bank formats and validates the uploaded files for correctness.Oncethedataisreceived,theextraction process begins: PDFs are parsed using specialized librariessuchaspdfplumberorPyPDF2,whileCSV files are read directly using pandas. The extracted raw data is then cleaned and standardized by removing duplicates, irrelevant text, and empty rows, ensuring that the dataset is ready for subsequent analysis.

2) Preprocessing and Expense Categorization Module

After extraction, the Preprocessing and Expense CategorizationModuleconvertsthetransactiondata intoastructuredformsuitableformachinelearning.

Text descriptions are preprocessed through steps suchaslowercaseconversion,punctuationremoval, and tokenization. Features are generated using TF- IDF vectorization, and transactions are classified into predefined categories like Food, Travel, Shopping, Utilities, and Entertainment using a trained Logistic Regression or Random Forest model.Thismoduleensuresaccuratecategorization even for unstructured and inconsistent transaction descriptions, forming the backbone of the system's intelligent decision-making.

3) Visualization andTrendAnalysisModule

The Visualization and Trend Analysis Module provides graphical insights into user expenses, helping them understand their financial behavior. Themodulegeneratespiechartstodisplaycategory- wisespendingproportionsandmonthlytrendgraphs to track expenditure over time. Interactive features allow users to zoom in, filter, or focus on specific periods,enablingadetailedexaminationofspending patterns. These visualizations not only improve comprehensionbutalsohighlightareaswhereusers can optimize spending, making the system more actionable and user-centric.

4) Budget Monitoring andAI-Based Forecasting Module

This module focuses on financial planning by monitoring user-defined budgets and forecasting futureexpenses.Userscansetmonthlyorcategory- wise budget limits, and the system sends real-time notifications if spending exceeds these thresholds. TheAI-basedforecastingcomponentemploystime- seriesanalysisandtrenddetectiononhistorical transactiondatatopredictnext-monthexpensesand identify potential high-spending categories. This proactive feature empowers users to plan ahead, avoid overspending, and make informed financial decisions.

5) SmartRecommendationandChatbotModule

The final module integrates a recommendation engine with an AI-powered chatbot assistant. The recommendation engine analyzes past spending patterns and provides personalized suggestions to optimize expenses and improve savings habits.The chatbotoffersaninteractiveinterfaceforuserstoask questions, receive explanations about transactions, and get financial guidance in real time. Built using LangChain andOpenAI,thismoduleenhancesuser engagement and ensures that the system is not only informative but also conversational and adaptive to individual financial behaviors.

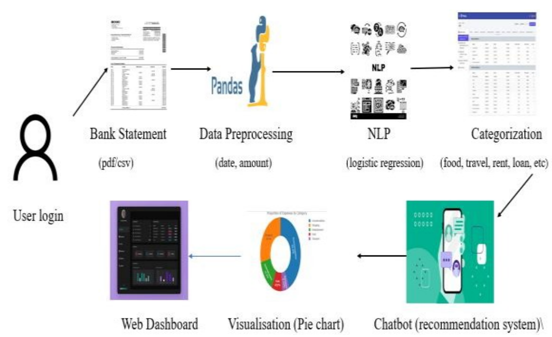


Fig.1Architecturediagramoftheproposed system

IV. EVALUATION MERTICS

A. MODEL PERFORMANCE ON THE TEST SET

1) TestLoss:0.42

The test loss represents how well the model generalizes on unseen data. A low test loss valueof0.42indicatesthatthemodelhas effectively learned the transaction patterns and minimized classification errors.

2) TestAccuracy:87.6%

The model achieved a high accuracy of 87.6% onthetestdataset.Thismeansthatnearly9out of 10 expense records were categorized correctly into their respective classes (Food, Travel, Shopping, Bills, or Others). The high accuracy reflects strong generalization and learning capability**.**

B. CONFUSION MATRIX

1) True Positives (TP): The model correctly identifiedtransactionsbelongingtoaparticular category(e.g.,predicting*Food*whentheactual label is *Food*).

2) TrueNegatives(TN):Transactionsfromother categories were correctly not classified as that particular category.
3) False Positives (FP): The model incorrectly predicted a transaction as belonging to a categorywhenitdidnot(e.g.,predicting*Travel* for a *Shopping* transaction).
4) False Negatives (FN): The model failed to identify a transaction that actually belonged to that category (e.g., missing a *Bills* expense).

	Food	Travel	Shopping	Bills	Others
Food	128	4	6	2	3
Travel	3	118	2	4	3
Shopping	5	3	112	2	5
Bills	1	2	4	96	2
Others	3	3	6	2	115

Fig.2 Confusion Matrix for the Automated Expense Classifier

C. CLASSIFICATION REPORT

1) Precision:

Food:0.91
Travel:0.91
Shopping:0.87
Bills: 0.88
Others:0.86

The **macro average precision (0.89)** indicates that the model maintains balanced accuracy across all categories, effectively minimizing false positives.

2) Recall:

Food:0.90
Travel:0.88
Shopping:0.86
Bills: 0.90
Others: 0.88

The **macro average recall (0.88)** shows that themodelisabletoidentifythemajorityoftrue instances for each category.

3) F1-Score:

Food:0.90
Travel:0.89
Shopping:0.86
Bills: 0.89
Others:0.87

The macro average F1-score (0.88) demonstrates a strong balance between precision and recall across all categories.

4) Accuracy:

The model achieves an overall **accuracy of 87.6%**, indicating that the majority of expense transactions were correctly classified.

5) MacroAverage:

Precision:0.89
Recall:0.88
F1-Score:0.88

Themacroaveragetreatsallcategoriesequally, confirming consistent model performance across all types of expenses.

V. CONCLUSION

TheAutomatedExpenseClassifiersuccessfully classifiesfinancialtransactionsintomeaningful categories using a machine learning model trained on textual transaction descriptions. By leveragingNaturalLanguageProcessing(NLP) andsupervisedlearningtechniques,thesystem efficiently processes CSV or PDF bank statements and automatically categorizes expenses into predefined labels such as Food, Travel, Bills, Shopping, and Others.

The achieved accuracy of 87.6% and balanced precision,recall,andF1-scoresindicatethatthe modelgeneralizeswelltounseenfinancialdata. The integration of budget alerts, AI-powered recommendations, and expense visualization further enhances user experience, enabling better financial awareness and planning.

Future enhancements include expanding the dataset, fine-tuning the model using transformer-basedarchitectures(suchasBERT orDistilBERT),andimplementingreal-time expensetrackingforimprovedaccuracyand adaptability.

REFERENCES

[1] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Proc. Int. Conf. on Learning Representations (ICLR), 2013, doi: 10.48550/arXiv.1301.3781.

[3] F.Pedregosa,G.Varoquaux,A.Gramfort,et al.,"Scikit-learn:MachineLearninginPython," JournalofMachineLearningResearch,vol.12,pp. 2825–2830, 2011, doi: 10.5555/1953048.2078195.

[4] S. Ruder, "An Overview of Gradient Descent Optimization Algorithms," arXiv preprint arXiv:1609.04747, 2016, doi: 10.48550/arXiv.1609.04747.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proc. NAACL-HLT, pp. 4171–4186, 2019, doi:10.48550/arXiv.1810.04805.

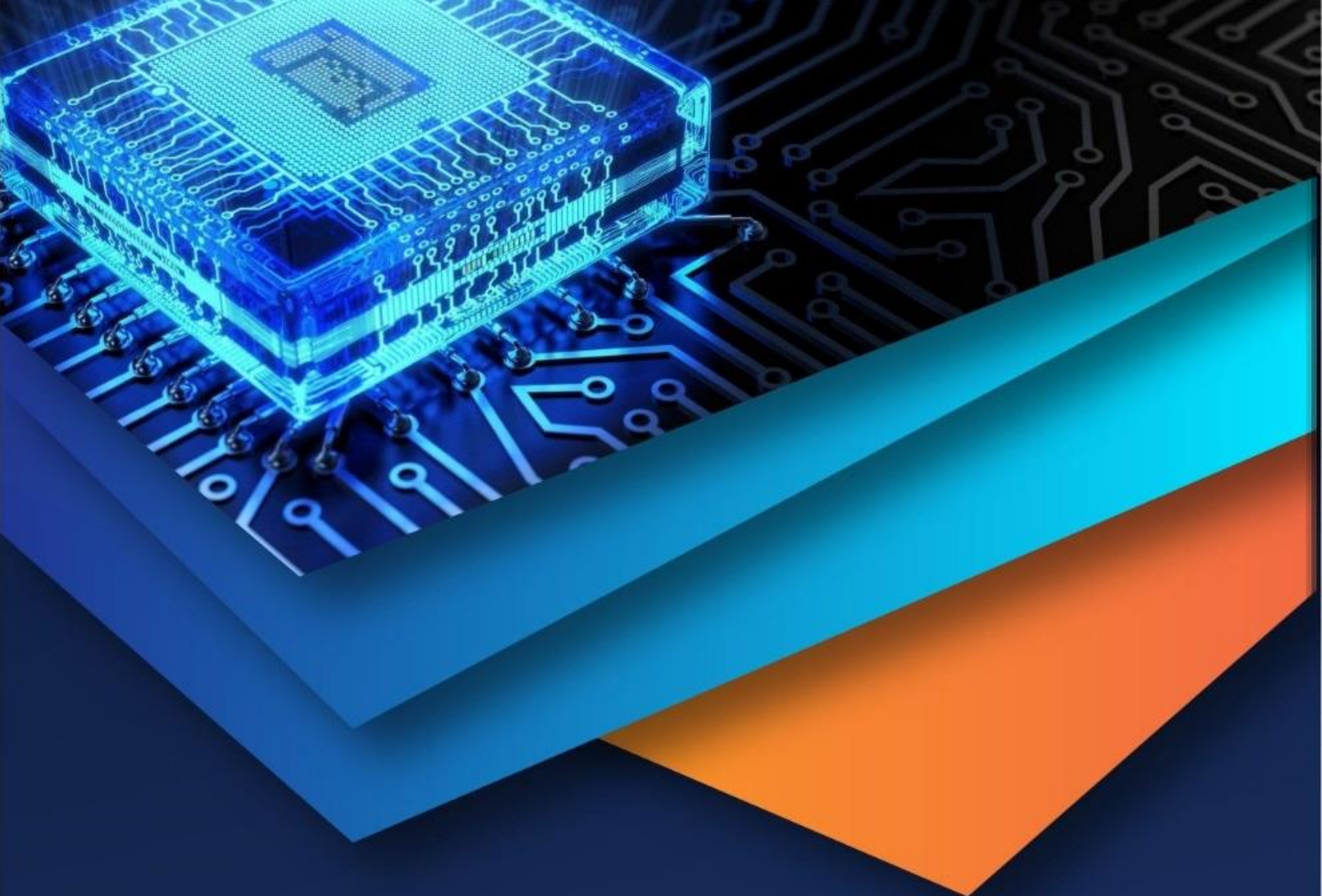
[6] Y.ZhangandX.Yang,"FinancialDocument ClassificationUsingDeepLearningandNLP," IEEEAccess,vol.9,pp.110212–110223,2021, doi: 10.1109/ACCESS.2021.3103354.

[7] P. Aggarwal, "Automated Expense ClassificationUsingNaturalLanguage Processing and Machine Learning," International Journal of Data Science and Applications, vol. 7, no. 2, pp. 44–51, 2021, doi: 10.5120/ijds2021072044.

[8] R. Kumar and V. Sharma, "AI-Driven Personal Finance Management System for Smart Budgeting," International Journal of ComputerApplications,vol.183,no.5,pp.12–19,2022,doi: 10.5120/ijca2022183053.

[9] K.Clark,U.Khandelwal,O.Levy,andC.D. Manning, "What Does BERT Look at? An Analysis of BERT's Attention," Proc. 2019 ACL Workshop BlackboxNLP, pp. 276–286, 2019, doi: 10.18653/v1/W19-4828.

[10] D. A. Dopazo, A. P. Cobo, and J. M. Herrera, "An Automated Machine Learning Approach for Classifying Financial Transactions," Computer-Aided Civil and Infrastructure Engineering, vol. 39, no. 2, pp. 291–304,Feb.2024,doi:10.1111/mice.13114.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY