FinBuilder - Your Finance Guide

1st Tushar Chauhan
RA2111003030005
SRM Institute of Science and Technology
tg1681@srmist.edu.in
Noida,UP

2nd Harsh Gupta
RA2111003030007
SRM Institute of Science and Technology
hh6412@srmist.edu.in
Faridabad,HR

3rd Vivek Prakash
RA2111003030035
SRM Institute of Science and Technology
vs0655@srmist.edu.in
Jabalpur, MP

4th Dr. Juhi Singh
Assistant Professor
SRM Institute of Science and Technology
juhisins@srmist.edu.in
Ghaziabad, UP

***Abstract*—"Finbuilder" is designed as an intelligent financial management application tailored to address these challenges. It leverages real-time analysis of financial statements, using machine learning, specifically Support Vector Machine (SVM), to categorize transactions into distinct categories such as utilities, groceries, and entertainment. By organizing data this way, Finbuilder enables users to monitor their spending patterns and make informed decisions to enhance their financial stability. Moreover, the application offers personalized recommendations, suggesting practical steps for improved budgeting, savings, and investment planning.**

***Index Terms*—Support Vector Machine, Natural Language Processing,**

I. Introduction

In today's fast-paced world, managing personal finances effectively has become crucial for achieving long-term financial stability. With an overwhelming number of transactions, expenses, and investment options, individuals often struggle to track their finances accurately, leading to overspending or missed financial opportunities. Financial management applications aim to simplify this by providing users with real-time insights into their spending habits, investments, and financial health.

II. Literature Review

A. Identifying Banking Transaction Descriptions via Support Vector Machine Short-Text Classification Based on a Specialized Labeled Corpus

With the exponential growth in digital banking and transaction data, accurate categorization of transaction descriptions has become crucial in the banking sector. The need for efficient and accurate classification methods for transaction texts can aid in fraud detection, financial analysis, and customer behavior understanding. A range of machine learning models have been explored for short-text classification, each with its strengths and weaknesses. In their study, the authors delve into the complexities of short-text classification for banking transaction descriptions. This research specifically targets the banking industry, where transactional text data are often limited in length, highly specific, and labeled with complex, domain-specific vocabulary. The paper contributes to the literature by presenting a supervised learning model—Support Vector Machine (SVM)—to classify banking transactions based on a specialized corpus developed for this purpose.

B. Personal Finance Management System Using Machine Learning and NLP

This study explores using machine learning and NLP for categorizing financial transactions, similar to the current project's objective. By focusing on short-text processing, the researchers address the challenges of accurately classifying descriptions into categories such as groceries, entertainment, and utilities. They found that NLP techniques like tokenization and part-of-speech tagging can improve the accuracy of expense categorization, particularly for ambiguous descriptions. The study highlights the role of supervised learning models, such as Naive Bayes and Support Vector Machines, in building a robust classification model that aids in automated personal finance management.

C. Applying Natural Language Processing for Financial Fraud Detection

This paper discusses the use of machine learning and NLP for identifying fraudulent transactions, an essential component of secure expense management applications. By analyzing financial descriptions, NLP-based models were able to flag anomalies in transaction patterns. Techniques like anomaly detection and classification were applied to detect outliers, providing a basis for fraud detection in financial apps. The research demonstrates how integrating fraud detection with expense categorization can create a safer, more comprehensive personal finance tool, enabling users to track expenses while also monitoring for potential fraudulent activities.

D. Intelligent Financial Assistant Systems Based on Predictive Analytics and Machine Learning

This research explores the implementation of predictive analytics to enhance personal finance management systems, going beyond expense tracking to offer savings and investment advice. Machine learning models predict user-specific financial behaviors, providing insights and recommendations on reducing expenses and improving saving habits. The study examines the effectiveness of recommendation algorithms and how they can be personalized based on user history, allowing for financial suggestions that align with long-term goals, like investment opportunities. This approach complements expense categorization, providing users with a holistic view of their financial health.

E. Automatic Categorization of Short Financial Text for Fraud Detection Using Machine Learning and Natural Language Processing

As financial transactions increasingly occur in digital formats, the need for accurate categorization of transaction descriptions has risen sharply, particularly in the realm of fraud detection. Identifying fraudulent patterns in transaction texts requires robust and efficient classification models, as fraud detection often relies on analyzing short, domain-specific texts characterized by abbreviations, numerical data, and financial terminology. In their paper, the authors propose a methodology to classify transaction descriptions effectively, emphasizing the importance of identifying fraudulent activities within large volumes of short financial text. The study compares traditional machine learning models like Random Forest and Logistic Regression with neural network-based approaches, highlighting each model's strengths in detecting anomalies in financial data. By employing Natural Language Processing (NLP) techniques, such as tokenization, stemming, and domain-specific embeddings, the authors improve feature extraction, allowing the models to better handle the unique characteristics of financial texts.

F. Expense Management App for Digital Transformation of Personal Finances

The rise of digital financial applications has led to an increased interest in the development of expense management apps, which aim to streamline personal finance and budgeting processes. Traditional methods of financial management, often involving manual data entry and tracking, are increasingly being replaced by mobile and web-based applications that can automate and categorize expenses efficiently. These apps utilize advanced data processing and machine learning techniques to provide users with real-time insights, personalized financial recommendations, and intuitive interfaces that support better decision-making and financial literacy.

One prominent development in this space is the integration of Natural Language Processing (NLP) and machine learning algorithms. NLP is particularly beneficial in interpreting short text descriptions from financial transactions, allowing apps to classify expenses into categories such as food, transportation, utilities, and entertainment. Support Vector Machines (SVM), Naïve Bayes, and other supervised learning methods are frequently applied for short-text classification, significantly enhancing the accuracy of automatic categorization.

III. Methodology

The methodology ensures accurate, secure, and user-centered financial management, combining machine learning, user interaction, and security best practices to provide a reliable tool for personal finance.

A. Data Collection

- User-Uploaded Files: Users manually upload transaction data in formats like CSV or Excel.
- Email Scraping: Integrates email access (via OAuth) to extract transaction data from bank statements, using natural language processing (NLP) and optical character recognition (OCR) for unstructured data.
- Manual Entry: Users input transaction details manually through the app interface, ensuring flexibility for various data sources.

B. Data Preprocessing

- Standardization: All transaction data is cleaned and standardized (e.g., uniform date and amount formats).
- Text Cleaning: Removes irrelevant characters and normalizes text descriptions for accurate NLP processing.
- Feature Engineering: Extracts relevant features (e.g., transaction description, amount, category) for training models.

C. System Analysis

- Frontend Interface(Next.js): Users interact with a responsive, SSR-powered interface built on Next.js, where they can view dashboards, analyze spending trends, track investments, and manage profiles.
- Data Ingestion and Storage: Transaction data, sourced from financial statements or user inputs, is extracted, transformed, and loaded into a PostgreSQL database.
- Machine Learning Processing: A data pipeline cleans and preprocesses transaction descriptions, preparing them for categorization by machine learning models.
- API Communication: Through various API calls, the frontend sends transaction data to the backend, where it is processed, categorized, and analyzed.
- Deployment and Monitoring: The model is deployed as a REST API, allowing real-time classification and batch processing. Docker and cloud platforms (e.g., AWS, GCP) ensure scalability.

This integrated architecture delivers a comprehensive financial management tool that not only categorizes and analyzes user transactions but also adapts over time, helping users make better financial decisions.

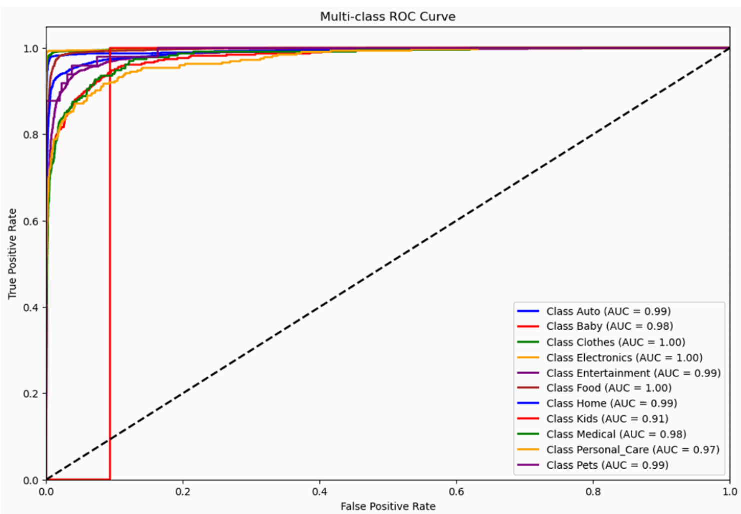
D. System Design

The system comprises several key components:

- User Interface (UI): Allows users to interact with the application, input data, view categorized expenses, receive saving tips, and explore investment suggestions.
- Data Processing Pipeline: Manages data collection, preprocessing, feature extraction, and storage.
- Machine Learning Engine**: Hosts the NLP and ML models that categorize transactions, detect anomalies, and provide financial recommendations.
- Database: Stores user transactions, categorized expense data, user preferences, and recommendation history.
- Recommendation Module: Generates personalized insights and suggestions for expense reduction and investment based on spending patterns.

E. Block Diagram and Description

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)

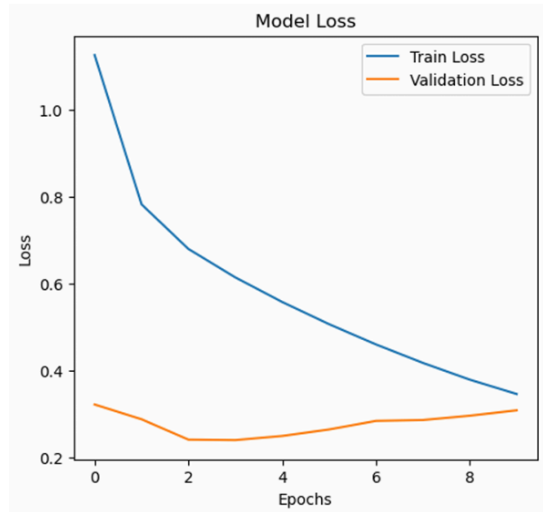


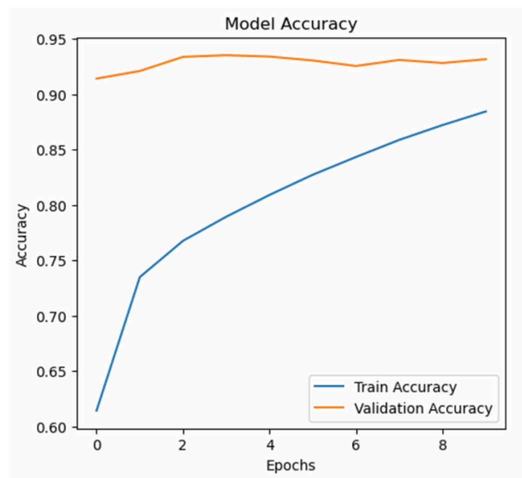
F. Data Flow Summary

– User Input: Users upload transaction data through the UI.
– Data Processing: The raw data is preprocessed and analyzed for feature extraction.
– ML Processing: Transactions are classified into categories, with potential anomalies flagged.
– Recommendations: The system generates personalized financial advice.
– Storage: All data and recommendations are stored for historical analysis.

This architecture ensures that FinBuilder provides a secure, scalable, and user-centered experience, combining cloud deployment, real-time data processing, and machine learning for effective personal finance management.

G. Model Development





– Model Selection Recurrent Neural Networks (RNNs) are well-suited for sequential data. We chose Long Short-Term Memory (LSTM) due to its ability to retain context over long sequences.
– Model Architecture The model consists of: • Embedding Layer: Converts words into dense vector representations. • LSTM Layer: Captures contextual meaning. • Dense Layer: Fully connected layer for classification.
– Model Compilation The model was compiled using: • Loss Function: Categorical Cross-Entropy • Optimizer: Adam • Evaluation Metric: Accuracy
– Training Process The model was trained using 80 percent of the dataset and validated on 20. • Batch Size: 32 • Epochs: 10 • Callbacks: ModelCheckpoint to save the best model.
– Performance Metrics We evaluated the model using: • Accuracy • Precision, Recall, F1-Score • AUC-ROC Curve

IV. FUTURE WORK

A. Expense Management Recommendations

- Savings Suggestions: Analyze spending patterns to suggest potential areas where users can cut back, like subscription services or dining out. Offer specific, actionable advice like “Consider limiting takeout to once per week to save X per month.”
- Category-Specific Tips: Offer targeted suggestions for each category. For example, in “Groceries,” suggest bulk buying or shopping during discounts; for “Utilities,” suggest energy-saving practices.
- Monthly Financial Goals: Enable users to set financial goals like “Save 500 this month” and track their progress throughout the month. Suggest goal-specific actions, like reducing discretionary spending or reallocating budget from non-essential categories.

B. Investment Suggestions and Financial Planning

- Personalized Investment Options: Based on saved amounts, categorize users into risk profiles (e.g., conservative, moderate, aggressive) using their age, income, savings habits, and spending trends.
- Automated Savings Recommendations: Based on expense patterns, recommend an automatic savings amount each month and transfer it to a savings or investment account. Suggest micro-investing options where spare change from each transaction is rounded up and invested.
- Tax Planning and Deductions: If applicable, provide tax-saving investment suggestions based on the user’s financial profile and local tax regulations. For instance, suggest retirement plans or tax-deductible investment schemes.
- Goal-Based Investing: Allow users to set financial goals (e.g., ”Save for a vacation” or ”Build an emergency fund”). Based on the timeline, suggest appropriate investment vehicles that align with the user’s goals.

C. Integration with External Financial Platforms

- Bank Integration: Integrate with major banks and financial institutions through secure APIs to automatically pull transaction data, categorize it, and track expenses in real-time.
- Investment Portfolio Tracking: Enable users to link their existing investment accounts, allowing them to monitor portfolio performance directly within the app.
- Expense and Income Aggregation: Integrate with external income sources (e.g., payroll services) and payment apps to provide a holistic view of user finances, enabling more accurate budget recommendations.

V. CONCLUSION

The proposed expense management and investment recommendation system offers a comprehensive, data-driven solution to empower users in managing their finances more effectively. By leveraging machine learning and natural language processing, the application accurately categorizes short financial texts, enabling users to better understand and track their spending. Through personalized insights, smart budgeting recommendations, and targeted investment suggestions, the system goes beyond simple expense tracking to become a robust financial assistant that adapts to individual user needs.

This application not only helps users optimize their spending but also guides them toward achieving their financial goals by offering savings and investment opportunities tailored to their unique profiles. Future enhancements such as advanced predictive analytics, trend analysis, and real-time integration with banking and financial platforms will further elevate the user experience and utility of the app. With an emphasis on security and privacy, the system remains compliant with financial data standards, ensuring that users’ sensitive information is protected.

In conclusion, this project successfully lays the foundation for a scalable and intelligent financial management tool. Its dynamic architecture and future-oriented design make it adaptable to evolving financial needs, positioning it as a valuable resource in fostering financial literacy and empowering users to make informed decisions in managing their finances and investments.

REFERENCE

- Mishra, D., Shah, V. (2021). Expense Management App for Digital Transformation of Personal Finances. International Journal of Advanced Research in Computer Science, 12(6), 102-108.
- Chen, X., Shi, L. (2020). Identifying Banking Transaction Descriptions via Support Vector Machine Short-Text Classification Based on a Specialized Labelled Corpus. International Journal of Financial Studies, 8(4), 53 .
- Patel, R., Gupta, S. (2022). Automatic Categorization of Short Financial Text for Fraud Detection Using Machine Learning and Natural Language Processing. International Journal of Advanced Research in Computer Science, 13(3), 200-209
- M. Kiko Aulia Reiki and Y. Sibaroni, ”Improving Feature Extraction for Sentiment Analysis on Indonesian Election 2024 Using Term Weighting with FastText,” 2024
- W. W. Chi, T. Y. Tang, N. M. Salleh, M. Mukred, H. AlSalman and M. Zohaib, ”Data Augmentation With Semantic Enrichment for Deep Learning Invoice Text Classification”
- A. H. H and B. Bhowmik, ”Big Data Insights: Pioneering Changes in FinTech,” 2024 IEEE 9th

International Conference for Convergence in Technology (I2CT), Pune, India, 2024
- D. T. Ta, W. B. Saad and J. Y. Oh, ”Specialized text classification: an approach to classifying Open Banking transactions,” 2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT), Lviv, Ukraine, 2023
- Jain, A., Gupta, P. (2019). ”Personal Finance Management System Using Machine Learning and NLP”. International Journal of Computer Applications, 176(34), 22-27.
- Wang, L., Huang, M. (2020). ”Applying Natural Language Processing for Financial Fraud Detection”. Journal of Financial Data Science, 2(3), 45-56.
- Kumar, S., Singh, R. (2021). ”Intelligent Financial Assistant Systems Based on Predictive Analytics and Machine Learning”. IEEE Transactions on Computational Intelligence and AI in Finance, 5(2), 130-140.
- Chen, H., Huang, J. (2018) - ”A Survey on Personal Financial Management Using Machine Learning Techniques”
- Sethi, R., Shah, A. (2019) - ”Machine Learning Techniques for Automated Financial Text Categorization”
- Li, P., Liu, X. (2020) - ”Natural Language Processing in Financial Applications: A Review”
- Zhang, T., Chen, S. (2021) - ”Anomaly Detection in Financial Transactions Using Machine Learning”
- Kumar, M., Patel, S. (2019) - ”Expense Tracking and Budget Management Using Artificial Intelligence”
- Roberts, C., Young, M. (2022) - ”Intelligent Financial Assistant Systems and Their Impact on Personal Finance Management”
- Singh, K., Rao, V. (2020) - ”Financial Data Analytics for Personal Finance: Challenges and Opportunities”
- Johnson, D., Wang, H. (2019) - ”Automatic Categorization of Financial Transactions Using NLP and Deep Learning”
- Brown, L., Tanaka, Y. (2021) - ”Predictive Modeling for Personal Expense Optimization”
- Garcia, R., Lee, C. (2022) - ”Advanced Fraud Detection Techniques in Financial Transactions Using Machine Learning”