Next-Gen Expense Tracking with Artificial Intelligence Integration

Sumathi S[1*], Dhashwathi M B[2], Pagalavan M[3] and Varsini P[4]

Head of the Department /AI&ML, Sri Eshwar College of Engineering, Coimbatore, India[1]

Department of Artificial Intelligence and Data Science, Sri Eshwar College of Engineering, Coimbatore, India[2]

Department of Artificial Intelligence and Data Science, Sri Eshwar College of Engineering, Coimbatore, India[3]

Department of Artificial Intelligence and Data Science, Sri Eshwar College of Engineering, Coimbatore, India[4]

Abstract

In this dynamic financial world, individuals are usually challenged to track their expenses with traditional tools and methods. Most traditional expenditure control tools fail to provide time-sensitive reports or deep insights customized to the specific financial behavior of an individual. This leaves the customers stuck with static reports requiring hours of tedious human input, hence underperforming their ability to respond quickly and decisively. The integrated dynamic data analytics with natural language processing will enable streamlined financial management using real-time personal financial information on customizable dashboards and advice through dashboards. Contrasting the conventional expense trackers that rely upon static reports and manual input, this solution allows users to retrieve information, compare expenditures, and set budgets using natural language queries. The virtual financial advisor facilitates improved user accessibility as well as participation. Real-time data visualization helps the users make better decisions. When all of these are put together, they escape the ills of traditional tools and provide users with a friendly and entertaining ground for the management of modern finances.

Keywords

Expense tracker, Natural Language Processing, Power BI, Large Language Model (LLM), Real-time analytics.

1. Introduction

In the Money management is an elementary step for anyone who wishes to be in control of his expenses and savings. Traditional expense tracking tools are no longer efficient to keep pace with all the growth in financial data; users are overwhelmed by manual procedures and static reports. Customers need more responsiveness and flexibility, and at this point, the limitations of traditional tools begin to show. The management of finances is very complicated, which often puts pressure on the consumers who do not acquire knowledge without which they can't act fast and sense-fully. Sensing this need, an engineering expense tracker was designed that uses latest technology to provide a perfect solution for giving customer financial autonomy. Using Power BI, it transforms unprocessed financial data into dashboards that are becoming more and more attractive and interactive to enable more in-depth study. A chatbot with a Large Language Model, designed to enhance user interaction with questions in natural languages, goes well with this visualization. The chatbot learns the behavior of the users, making it more personalized experience with advice about the individual expenditure patterns. This makes tracking effortless and is coupled with enhancing user engagement by offering insight into what is personal to the individual's spending habits. This flexibility also allows the system to meet the varying needs of the users including budget watchers and those with certain investment opportunities they seek to maximize. Simplifying the spending management process, while making the experience increasingly personal and engaging will give a user's ability to navigate their financial landscapes with clarity and confidence. That is through its capability of filling the gap between the simplicity and usability of access by users and the complexity of data, enabling the management of personal finances to be brought to another level.

1.1 Problem Statement

As the financial landscapes continue to grow in complexity, the personal financial management of people becomes highly challenging because the technological advancement is happening at such a fast rate. Most rudimentary expense tracking methods including spreadsheets or the most basic budgeting apps are bound to fall short of users' expectations. The

Conventional tools often depend on labor-intensive manual data entry, making them susceptible to human errors and inefficient for users who need real-time insights. The static nature of reports that these kinds of tools generate further hampers the capacity of users to access dynamic and actionable data. In addition, the real-time tracking fails to give users proper clarity about their existing spending and financial status for timely adjustments concerning the shifting economic conditions. Interactive elements are also missing in most of the out-of-date systems that dominate the existing market. Those systems fail to provide proactive financial advisory support or analytical statements. Users cannot predict what will be their future expenses and which spending pattern they have followed by the advanced analytics. Thus, they often cannot make well-informed decisions that would fortify their financial security. In an age where the norm is personalized digital experiences and automation, these more traditional tools leave users underserved. This opens up a great gap in the market for a smarter and better-intuitive solution that can present real-time data, integrated with advanced financial analysis, with further personalized insights. It is very vital to provide an innovative way in expense management which includes interactivity, automation, and personalization to keep the individuals up-to-top of their financial health.

1.2 Objective and Scope

The development of an advanced expenditure tracking system incorporates an interactive chatbot powered by LLMs with real-time financial data visualization through Power BI, thus permitting direct experience enhancements through dynamic, user-friendly interfaces that facilitate easy simple interactions and customized financial information. Overcoming the drawbacks of the conventional system of expense tracking such as manual data entry and static reporting, the system provides updates on spending patterns and offers real-time projections of expenses in the future. The project provides access and the ability to analyze financial data, so consumers can use it as a way of making clear and confident decisions on their finances in the long run. The main objective is to provide a fresh and interactive expense tracker that would make personal finance management easier.

1. Track daily expenses efficiently through an intuitive interface.
2. Query financial data using natural language with the SQL-to-LLM chatbot.
3. Visualize spending patterns via a linked Power BI dashboard.
4. Categorize and organize expenses for better financial insights.
5. Output customized reports to track and control individually tailored personal budgets.

2. Literature Survey

In their investigation of natural language processing for database querying, the authors present a two-phase methodology, the first phase involves pre-processing input with deep learning techniques using an LSTM model and Conditional Random Fields (CRF) for intent detection and slot filling. The second phase converts processed input into SQL queries, allowing efficient data retrieval. The system generates results in both text and speech, with future work aimed at improving accuracy and expanding its applications. [1]

The study compares several LLM-based text-to-SQL methods such as SQLQueryChain, DANKE, and C3 with GPT-4 on the same task to build a Natural Language Interface for databases. C3 with GPT-4 performed well when it came to converting queries but costed highly and had higher runtimes. The usage of tokens as well as SQLQueryChain execution time were significantly better. Many techniques struggled to handle highly complex database schemas effectively and exhibited reduced accuracy with intricate queries. The results indicate a need for further research to tackle these challenges in managing large and complex datasets. [2]

This overview describes the technology behind creating an agile and fast application that utilizes JavaScript, Angular, and PrimeNG to make a responsive front end while emphasizing Angular's efficient routing and reusable components. ExpressJS is employed on the backend to manage route distribution while ensuring perfectly smooth interactions. MongoDB is a NoSQL database that provides dynamic organization and retrieval of data. The technologies ensured the scalability of a high- performance application. [3]

The authors present ChatBI, a solution for the NL2BI task in production systems, which improves performance by matching MRD scenarios using two efficient Bert-like models. It leverages view technology for schema linking, transforming it into a Single View Selection problem. A phased process enhances SQL generation accuracy by decoupling complex semantics from LLMs, reducing errors. It demonstrates its superiority in terms of accuracy and efficiency compared to mainstream NL2SQL methods across various product lines. [4]

This paper provides insight into how analytical thinking and visualization of big data will be useful for enhancing HR policy and reducing costs through accounting analytics models in workforce and project management. It flags predictive and prescriptive analytics into decision-making enhancement, showing better internal control through dashboards that track trends. Further, it suggests the cost-tracking improvement by the use of expense management software, such as Helio, which is easily integrated with accounting systems. [5]

The author introduces Chat2Data, a three-layer framework that combines vector databases, large language models (LLMs), and retrieval-augmented generation (RAG). The application's first layer embeds domain-specific knowledge to reduce hallucinations, the second limits interactions involving the LLMs to optimize performance, and the third improves accuracy by breaking down complex tasks into subtasks for multi-round reasoning. This approach enhances system performance while solving usual challenges in data-driven applications. [6]

3. Proposed Solution

The proposed solution for an expense tracker system that caters to assisting users deal with handling personal financial matters by using advanced technologies. The mechanism includes processing data and applying models based on machine learning applied across algorithms based on detailed categorizations, with interactive visualizations to offer a comprehensive overview of user financial matters. This system captures user expense data, categorizes, analyses using algorithms with machine learning basis, and offers insights through an intuitive dashboard. The system will be interacted by the users in their monitoring and tracking of spending habits, setting financial goals, and coming up with good decisions. This solution is well positioned to simplify personal finance management for both parties, coupling the robust integration of a tech stack, thus providing a smooth experience.

3.1 Methodology

The methodology of the expense tracker involves gathering user expense data, pre-processing it to maintain accuracy and consistency, categorizing and analyzing the data with machine learning algorithms, splitting the data into training and testing sets to optimize the model, and evaluating the model's overall performance. Then, based on the results of the testing, the model is deployed into the live environment for live tracking and visualization of expenses based on the dashboard. Fine-tuning of the various parameters is required, and different techniques of machine learning will further enhance the accuracy, reliability of the model, and prevent overfitting. In this manner, the approach given above would be efficient and friendly to users for financial management.

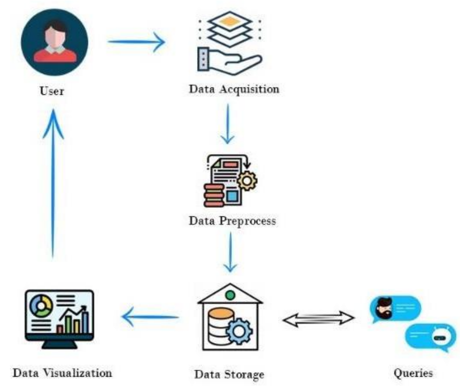

Fig: 3.1 BLOCK DIAGRAM

The project is divided into four key sections. They are Data Acquisition, Data Preprocessing, Data Storage and Data Visualization, illustrated in Figure 3.1.

3.1.1 Data Collection and Storage:

Manual and Automated Input: Users can manually input their financial transactions from their banks or credit card statements. Transactions include information such as date, category (e.g., groceries, utilities, entertainment), description, and amount

Database: All the data of transactions are stored in a MySQL database for this purpose. Keeping all the data in a single resource with integrity and easy retrieval is possible. The backend has been implemented with Python using Fast API, which ensures that a robust and responsive API is used for interaction between the frontend and backend. The schema of the database consists of tables for users, transactions, budgets, and categories. Possible dependencies have been created with relational constraints to optimize queries for faster data access.

3.1.2 Data Preprocessing and Normalization:

Data Cleaning: The raw financial data undergoes cleaning to remove null values, standardize date formats, and ensure all fields contain valid entries. Using Pandas, these preprocessing steps enable efficient data manipulation, ensuring that the dataset is ready for analysis without inconsistencies or errors.

Categorization and Transformation: Each transaction is categorized either manually by users or automatically using NLP techniques that infer categories from transaction descriptions. Numerical features, such as transaction amounts, are normalized to maintain consistent scaling, which enhances the performance of machine learning models. Categorical variables are transformed into machine-readable formats to support effective analysis and predictions.

3.1.3 Model Training and Model Building:

Feature Extraction and Data Split: To help train the machine learning model, salient features including date, category, and amount are extracted. In order to properly assess the model and make sure it generalizes adequately to new data, it separates the dataset into training and testing sets.

Model Training: In order to identify and assess expenditure trends, a Support Vector Machine (SVM) model is trained. The model learns relationships in the data to forecast future spending behavior by utilizing information such as transaction amounts and spending categories. Hyperparameters such as kernel type and regularization are fine-tuned to achieve optimal accuracy and performance.

3.2 Algorithm

A Support Vector Machine (SVM)-based machine learning model is used by the expense tracker system for classification and prediction, along with supplementary data processing methods to guarantee seamless user engagement. The algorithm focuses on analyzing user spending habits, predicting future spending behaviors, and providing intelligent financial insights.

3.2.1 Hyperplane Classification:

Spending patterns are predicted and user expenses are categorized using Support Vector Machines (SVM). The goal of SVM, a supervised learning method, is to identify the optimal hyperplane for dividing the data into distinct classes. These classes can be used to reflect distinct spending behaviors in the cost tracker, such as "high," "medium," or "low" spending, across a variety of categories including grocery, utilities, entertainment, etc. Finding patterns in user transactions and categorizing them appropriately is the main goal, which enables the model to provide insights based on historical behavior.

3.2.2 Feature Vector Formation:

The model’s input consists of feature vectors, which are extracted from transaction data. These feature vectors include important fields like transaction amount, spending category, date, and day of the month. Each transaction is converted into a numerical representation, and these features help the model understand spending patterns across different periods and categories. By analyzing the relationships between features, the SVM model learns to distinguish different spending habits as shown in fig 3.2, making it capable of accurately predicting future financial behaviors.

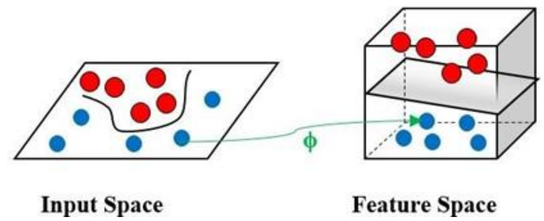

Fig: 3.2 SUPPORT VECTOR MACHINE

3.2.3 Maximizing Margin:

One of the strengths of SVM lies in its ability to maximize the margin between different classes. The margin refers to the distance between data points from different classes and the decision boundary (hyperplane). The larger the margin, the better the model generalizes. The SVM model efficiently distinguishes between various spending categories by optimizing the margin, guaranteeing robust performance on both training data and fresh, unseen transactions. This characteristic of SVM enhances the model's accuracy and reduces the risk of overfitting, making it more dependable for real-world applications.

3.2.4 Cross Validation:

During training, k-fold cross-validation is used to increase the model's performance and robustness. Using this method, the model is trained k times once the dataset is split up into k pieces. The remaining k-1 parts are used for training, and a different portion is used as the validation set each time. By testing the model on different data segments, this technique helps to reduce overfitting. The approach guarantees that the SVM model can generalize to unseen data by employing cross-validation, which improves the model's accuracy in forecasting spending patterns for a variety of customers.

3.3 Working Principle

3.3.1 React with TypeScript:

The frontend interface of the expense tracker is designed with React and TypeScript. These Two technologies make up a powerful modern JavaScript library and a good typing system. TypeScript offers a strong typing system that minimizes development errors by ensuring that variables, functions, and components conform to the expected types. This leads to more reliable and maintainable code, reducing the likelihood of runtime errors that could impact the user experience. This makes it possible to build modular components that can be reused effectively to make the building blocks of the user interface. Advanced search and filtering options are provided to allow users to quickly explore their transaction history or monitor their budget adherence. It is easier to update the user interface dynamically based on user input, maximizing the level of responsiveness at one's fingertips.

3.3.2 FASTAPI:

FastAPI is essentially the central API layer where the frontend application communicates with the database and machine learning components. It handles all kinds of incoming requests related to the user interface such as adding new transactions, fetching financial records, or even interacting with machine learning models for predictions as shown in fig 3.3. It's fast, efficient, making use of asynchronous processing that will handle multiple requests concurrently and ensure users experience prompt response times. The API performs all CRUD operations- Create, Read, Update, and Delete, so that users can input the relevant financial data like transactions, update previous entries, and even fetch detailed reports on their spending habits. Also, the security and integrity of data are maintained to the fullest, ensuring that every user's personal information remains safe.

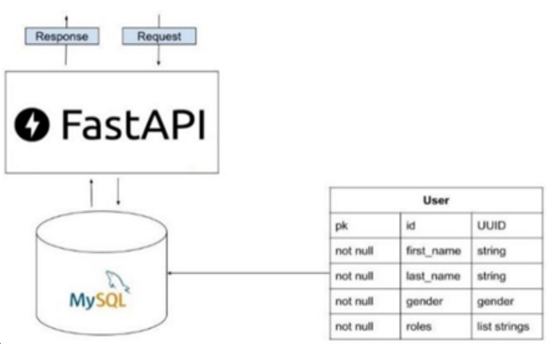

Fig: 3.3 WORKING OF FASTAPI

3.3.3 MYSQL

All the user profiles, transactions, budget categories, and financial data are stored in a relational database MySQL. The role of the database is to ensure that datais consistent, normalized, and well-integrated. Users, expense categories, budgets, and their transactions are all part of a schema which represents these relationships as tables. There is the need of data processing which should synchronize the data processing and computation time. Querying the schema is efficient which is why it is possible for data to be retrieved very quickly whenever users want reports or analyses. To ensure fast and reliable data access, indexing techniques are used in MySQL, enabling complex queries (like filtering transactions by date or category) to run efficiently. The backend system also ensures that the data is normalized, reducing redundancy and ensuring data integrity for seamless financial tracking and management.

3.3.4 SVM Model:

The expense tracker system's machine learning component utilizes a Support Vector Machine (SVM) model to analyze the spending behavior and forecasting future financial actions. Once trained, the SVM model can classify user spending into different categories (such as "high", "medium", or "low" spending) and detect anomalies. For example, if a user typically spends around \$50 on dining out each week but suddenly logs a \$300 restaurant bill, the system can flag this as an unusual expense, prompting the user to verify if the transaction was correct or possibly fraudulent. Moreover, it allows the model to point toward future spending and help the user configure a budget and alter his or her financial behavior through its recommendations.

3.3.5 SQL-TO-LLM ChatBot:

A chatbot driven by Large Language Models (LLM) for natural language processing (NLP) is part of the cost tracker. Instead of scrolling through intricate menus, users can now engage with the system using conversational language. Users may enquire, for instance, "How much did I spend on travel in the last month?" As seen in fig. 3.4, the chatbot then converts the user's natural language inquiry into a matching SQL statement that is run on the MySQL database. The consumer is given the retrieved data in a comprehensible style, like "You spent \$500 on travel in September". By making the system more approachable and intuitive, this conversational interface greatly enhances the user experience, particularly for individuals who might not feel comfortable switching between screens to get financial data.

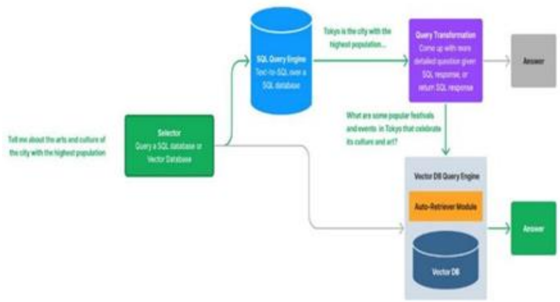

Fig: 3.4 SQL-TO-LLM CHATBOT WORKING

3.3.6 POWERBI:

To provide comprehensive insights into user spending habits, the system integrates Power BI for data visualization. Power BI transforms raw financial data into visually appealing and informative reports, using various types of charts and graphs. These reports help users understand their spending behaviors over time, across different categories, or in comparison to their budgets. One aspect that Power BI provides is customizable dashboards. Additionally, Power BI allows for real-time updates, so the visualizations update to reflect the most recent data as soon as new transactions are entered or budgets are modified. This enables users to make informed, proactive financial decisions without delay.

4. Result and Discussion

The proposed expense tracker system has demonstrated impressive performance, achieving an accuracy rate of 95%. This high accuracy indicates the system's ability to effectively categorize user expenses and detect irregularities in spending patterns. The SVM model, which was trained on a diverse set of historical transaction data, played a critical role in recognizing patterns and providing accurate predictions for future spending habits. Additionally, advanced preprocessing techniques, such as data normalization and feature extraction, ensured that the model could handle varied datasets efficiently, contributing to the system's robust performance. With such reliable accuracy, the system offers users detailed financial insights, personalized budget recommendations, and alerts for unusual transactions, making it a valuable tool for managing both personal and business finances.

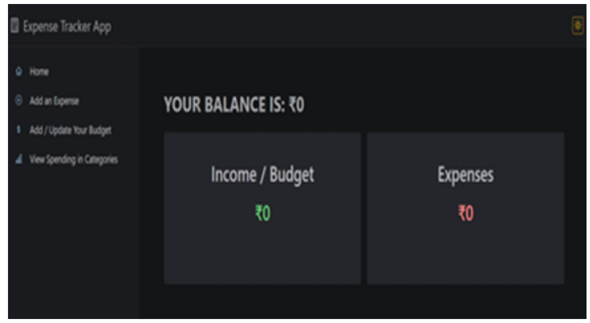

Fig: 4.1 WEBSITE OVERVIEW

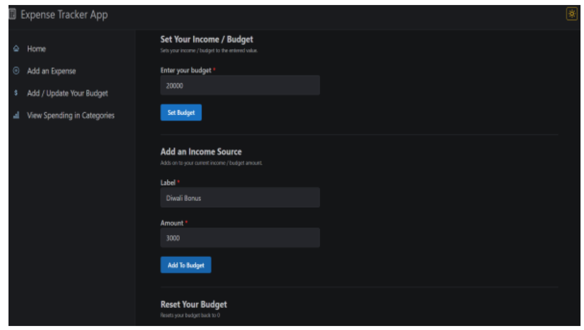

Fig: 4.2 SETTING BUDGET

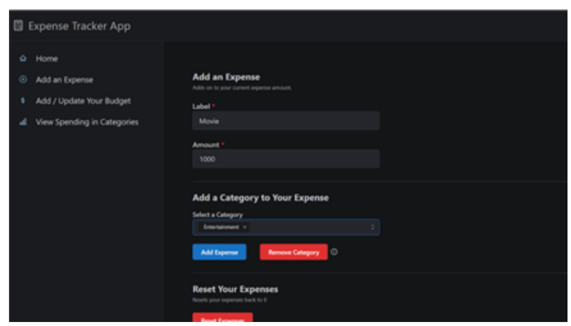

Fig: 4.3 ADDING EXPENSES

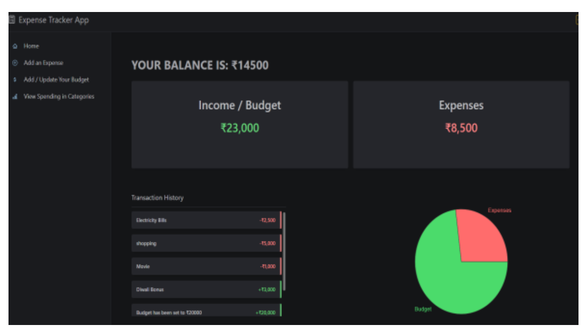

Fig: 4.4 VISUALIZING EXPENSES

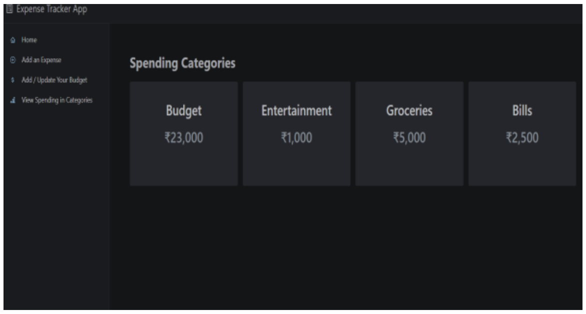

Fig: 4.5 SPENDING IN CATEGORIES

5. Conclusion and Future Scope

The expense tracker project demonstrated advanced use of recent technologies like Power BI, NLP, and machine learning to simplify personal finance management. It has shown that a chatbot combined with an easy-to-use dashboard actually enriches the user experience with true real-time insights into financial data. The ability to track and then visualize easy expenses forms a strong base for further development improvements. The project integrates natural language processing for a user-friendly, query-based interaction in expense management.

- The scope of the research explores the improvement of the chatbot to correctly answer complex financial queries.

- More complex machine learning models would help better categorize and expense, thereby giving more accurate insights.

- The goal is to create a scalable solution for the personal as well as for the business.

- Future advancements are expected to boost overall user satisfaction in personal and corporate finance management.

Acknowledgment
None.

Conflicts of interest
The authors have no conflicts of interest to declare.

References

[1] "Chat-SQL: natural language text to SQL queries based on deep learning techniques" by Machkour, M. and Majhadi, K.2024 Applied and Theoretical Information Technology Journal, 102 (12).

[2] Zhang, Qinggang, Junnan Dong, Hao Chen, Wentao Li, Feiran Huang, and Xiao Huang. "Structure guided large language model for sql generation." arXiv, 19 Feb(2024).

[3] "ChatBI: Towards Natural Language to Complex Business Intelligence SQL". Shao Y, Wei Z, Lian J, Liu X, Dong Y, Yan H, Wang M, Dong M, and Wan T. May 1, 2024, arXiv.

[4] Coelho, G.M., Nascimento, E.R., Garcıa, G.M., Feijó, L., Victorio, W.Z., Izquierdo, Y.T., de Oliveira, A.R., Coelho, G.M., Lemos, M., Garcia, R.L., Casanova, M.A., and Leme L.A.P. 2024. “Text-to-SQL Meets the Real-World”. In 26th Int. Conf. on Enterprise Info. Sys.

[5] Li G, Zhou X, Zhao X. “ChatToData: An Interactive Data Analysis System with RAG, Vector Databases and LLMs”. Proc. VLDB Endow. 2024.

[6] Toh, Y.X., Chong, L.Y. and Chong, S.C., 2024, June. “A secured self-financial management application with RSA encryption technology”. In AIP Conference Proceedings (Vol. 3153, No. 1). AIP Publishing.

[7] Salleh, C.K. and Wong, M.N.M., 2023. “Personal Finance and Budgeting Mobile Application, ‘CashSave’”. Applied Information Technology and Computer Science, 4(1), pp.1372-1387.

[8] Hong Chen, Jun Zhu, Renjie Wei, Jing Zhang, Hanbing Liu, Ju Fan, Hongyan Pan, Xiaokang Zhang, Cuiping Li, and Haoyang Li. 2024. “CodeS: Towards Building Open-source Language Models for Text-to-SQL”.

[9] Immanuel Trummer. 2022. “Codexdb: Synthesizing code for query processing from natural language instructions using gpt-3 codex”. Proceedings of the VLDB Endowment, 15(11):2921–2928.

[10] Rafiei, D. and Pourreza, M.. (2023). “Din-sql: Decomposed in-context learning of text-to-sql with self-correction”. arXiv preprint arXiv:2304.11015.

[11] Yanbin Zhao, Kai Yu,Su Zhu, Zhi Chen, Ruisheng Cao , and Lu Chen. 2021. “LGESQL: line graph enhanced text-to-SQL model with mixed local and non-local relations”. arXiv (2021).

[12] “C3: Zero-shot text-to-sql with chatgpt”. Yuhang Ge, Chao Zhang, Yunjun Gao, Xuemei Dong, Yuren Mao, Dongfang Lou, Jinshu Lin et al. 2023.arXiv (2023).

[13] Yichen Qian, Dawei Gao, Haibin Wang, Bolin Ding, Yaliang Li, Xiuyu Sun, and Jingren Zhou. 2023. “Text-to-sql empowered by large language models: A benchmark evaluation”. arXiv (2023).

[14] Scott, Clark. 2010. “New Mobile Apps re Using HTML 5, CSS and JavaScript”. Accessed May 20, 2023.

[15] Davood Rafiei and Mohammadreza Pourreza. 2024. “Din-sql: Decomposed incontext learning of text-to sql with self-correction”. Advances in Neural Information Processing Systems 36 (2024).

[16] C. S. Lee, M. Moslehpour and P. Y. S. Cheang, “Predictive Analytics in Business Analytics: Decision Tree,” Advances in Decision Sciences, vol. 26, no. 1, 2022, doi: 10.47654/V26Y2022I1P1 30.

[17] vans, Philip. 2022. “Review of Power BI in Technical Services.” TCB: Technical Services in Religion & Theology 30 (1): 10–10

[18] Guoliang Li, Xuanhe Zhou, and Zhaoyan Sun. 2024. “DB-GPT: Large Language Model Meets Database”. Data Science and Engineering (2024), 1 10.

[19] Zhao Yan, Bing Wang, Jiaqi Bai, Changyu Ren, Jian Yang, Xinnian Liang, Zhoujun Li, and Qian Wen Zhang. 2023. “Mac-sql: Multi-agent collaboration for text-tosql”. arXiv (2023).

[20] Hootan Nakhost, Ruoxi Sun, Pengcheng Yin, Hanjun Dai, Sercan O Arik, Tomas Pfister and Rajarishi Sinha. 2023. “Sql-palm: Improved large language model adaptation for text-to-sql”. arXiv (2023).

[21] Sumathi, S.; Madhumitha, S.; Sushma, R.; Suriya, M.; Subitsha, M.; Swetha, S., "Certain Investigations on Cognitive based Movie Recommendation system using Pairwise Cosine Similarity" 2023 9th International Conference on Advanced Computing and Communication Systems, ICACCS 2023 DOI: 10.1109/ICACCS57279.2023.10112790

[22] Improved optimizer with deep learning model for emotion detection and classification
W.C., Joseph, Willson C.,J.W., Kathrine, Jaspher Willsie, S., Vimal, S., ... X.P., Blanco-Valencia, Xiomara P., E., Verdú Pérez, Elena,Mathematical Biosciences and Engineering, 2024