Enhancing Business Intelligence Through NLP and Contextual AI Synergy

Kalyanaraman P
School of Computer Sc. and Engg.
VIT University, Vellore, India
pkalyanaraman@vit.ac.in

Arockia Sachin A
School of Computer Sc. and Engg.
VIT University, Vellore, India
arockiasachin.a2023@vitstudent.ac.in

***Abstract*—In the ever-evolving data-driven business landscape, retailers are increasingly dependent on Business Intelligence (BI) tools to gain insights and make informed decisions that drive growth and operational efficiency. However, the complexity of traditional BI systems often restricts access to users with technical expertise in data analysis and SQL, limiting broader organizational impact. This project proposes a novel BI application that integrates Natural Language Processing (NLP) and advanced machine learning techniques to bridge this gap. By employing a user-friendly chatbot interface, the application empowers users to interact with their financial data through intuitive natural language queries, facilitating seamless access to insights, industry trends, and predictive analytics. The system architecture is designed to convert natural language queries into structured SQL commands using transformer models, allowing non-expert users to engage with complex datasets. Real-time data processing capabilities, supported by streaming platforms like Apache Kafka, ensure that insights are continuously updated, providing dynamic and actionable recommendations. Additionally, automated dashboard generation and visualization tools such as Apache Superset present data in easily interpretable formats, enhancing decision-making efficiency. This comprehensive approach not only democratizes access to data-driven insights but also provides personalized strategies tailored to individual retailer needs. Despite challenges such as ensuring data privacy and scalability, the proposed BI application aligns with the growing demand for accessible and actionable business intelligence, offering a robust solution that enhances competitiveness in the modern retail sector.**

***Index Terms*—**

I. Introduction

In today's data-driven retail landscape, the ability to rapidly analyze and interpret vast amounts of data is critical for making strategic decisions that drive business success. Business Intelligence (BI) tools have become essential for retailers, providing valuable insights to optimize operations, enhance customer experiences, and increase profitability. However, traditional BI systems often present significant challenges, requiring specialized expertise in data analysis and proficiency in SQL. This limitation creates barriers for non-technical users, slowing down the decision-making process and reducing the broader organizational impact. The concept of BI dates back to 1865, when Richard Millar Devens first introduced the term to describe how competitive advantages could be gained by acting on timely information. Since then, BI has evolved considerably, moving through key phases that include data warehousing, online analytical processing (OLAP), big data, and, most recently, real-time analytics. Despite these advancements, the complexity of modern BI systems continues to restrict their use to a limited group of technically skilled users, thereby hindering widespread organizational adoption. In recent years, the integration of Artificial Intelligence (AI) into BI has significantly transformed traditional systems. AI-driven BI tools have introduced advanced capabilities such as predictive analytics, real-time data processing, and personalized insights, yet challenges remain in making these technologies accessible to non-technical users across the organization. To address this gap, this paper proposes the development of a next-generation BI application that leverages cutting-edge Natural Language Processing (NLP) and machine learning techniques to democratize data access. Central to the system is a user-friendly chatbot interface, which enables users to interact with their data through natural language queries, removing the need for technical expertise in SQL or data manipulation. By converting natural language queries into structured SQL commands using transformer models, the system allows even complex data analysis to be performed intuitively, empowering a broader range of users to engage with data and extract meaningful insights. The proposed BI application also incorporates real-time data processing capabilities, powered by streaming platforms like Apache Kafka, ensuring that the insights provided are continuously updated and reflective of the most current data. Automated dashboard generation and visualization tools, such as Apache Superset, further enhance decision-making by presenting data in clear, actionable formats. Predictive analytics capabilities allow the system to forecast future trends based on historical data, enabling retailers to proactively adapt their strategies in response to changing market conditions. Personalized recommendations are also tailored to each retailer's specific context, making the insights not only relevant but also actionable. While the system offers numerous benefits, its development and deployment come with challenges. Ensuring robust data privacy and security is paramount, particularly in handling sensitive financial and operational information. Scalability is another critical factor, as the application must be able to handle varying data volumes and user loads across different retail environments. Despite these challenges, the proposed solution aligns with the growing demand for accessible, real-time business intelligence tools.

By democratizing access to data and providing actionable insights through an intuitive interface, this application has the potential to revolutionize decision-making processes in the retail sector, making businesses more agile, data-driven, and competitive in today's rapidly evolving market.

II. New Literature Survey

The field of Natural Language to SQL (NL2SQL) systems has seen significant advancements, particularly in their application to enterprise data marts and large-scale databases. Recent research by [Dong et al., 2021] presents a sophisticated model tailored for the banking sector. This model leverages a combination of advanced techniques, including semantic parsing, feedback loops, and data simulation, to enhance the accuracy of SQL query generation from natural language inputs. A key innovation in this system is the integration of a BERT-based encoder, which processes natural language inputs and supports complex multi-table queries through a novel table expansion technique. The findings of this study demonstrate the model's superior performance over traditional rule-based and earlier neural network approaches, particularly when trained with template-simulated data specific to the banking domain. The significance of this work lies in its potential to democratize data access within organizations, enabling non-technical users to extract insights through natural language queries, thereby improving operational efficiency and decision-making processes. Complementing this work [Li et al., 2024] introduce BIRD, a benchmark designed for largescale databases grounded in Text-to-SQL tasks. The benchmark includes a comprehensive dataset of 12,751 Text-toSQL pairs across 95 databases, highlighting the challenges of handling noisy database values, external knowledge grounding, and SQL execution efficiency in large databases. Despite employing advanced models like GPT-4, the study reveals that current text-to-SQL models still fall short in real-world applications, achieving only a 54.89% execution accuracy compared to the human benchmark of 92.96%. This research underscores the need for further advancements in database value comprehension and efficiency in SQL generation, particularly in the context of large-scale databases. In the domain of automated database querying, [Majhadi and Machkour, 2024] propose CHAT-SQL, a deep learning-based NLIDB system that utilizes a sequence-to-sequence model with Long Short-Term Memory (LSTM) networks. This system is designed to interpret user queries in natural language and automatically generate corresponding SQL queries, overcoming traditional NLIDB limitations. The authors demonstrate that CHAT-SQL effectively handles complex query translations, outperforming existing systems in terms of accuracy and flexibility. This research is significant in its potential to make database interactions more user-friendly, empowering users with varying levels of technical expertise to efficiently retrieve and manipulate data. [Pourreza and Rafiei, 2024] proposed DIN-SQL, a novel method designed to enhance the performance of NL2SQL tasks by decomposing the complex process of translating natural language into SQL queries into smaller, more manageable sub-tasks. This approach incorporates modules such as schema linking, query classification and decomposition, SQL generation, and self-correction to improve the accuracy and reliability of SQL query generation. The DIN-SQL model demonstrated significant improvements over traditional finetuned models and existing prompting methods, achieving stateof-the-art results on challenging benchmarks like Spider and BIRD. This research underscores the effectiveness of breaking down complex queries into simpler components and leveraging adaptive prompting strategies tailored to the complexity of the task, offering a more refined and effective method for enhancing NL2SQL model performance. [Gu et al., 2023] introduced SCPrompt, a novel framework designed to enhance few-shot Textto-SQL translation tasks by dividing the translation process into two distinct stages: structure generation and content population. The first stage focuses on guiding pre-trained language models (PLMs) to generate SQL structures with placeholders, while the second stage involves filling in these placeholders with specific content such as table names and column names. This divide-and-conquer approach allows for more efficient handling of limited training data, addressing the generalization challenges that typically arise in few-shot learning scenarios. SC-Prompt leverages a hybrid prompt strategy that combines learnable vectors with fixed textual prompts, ensuring the PLMs are better guided during both stages of the translation process. Additionally, the framework incorporates fine-grained constrained decoding methods to further enhance the accuracy of the generated SQL queries. Extensive experiments demonstrated that SC-Prompt significantly outperforms state-of-theart Text-to-SQL methods, particularly in low-resource settings, underscoring the effectiveness of this structured and contentdriven approach to few-shot Text-to-SQL tasks. [Guo et al., 2023] introduced an innovative framework that enhances Text-to-SQL tasks by leveraging large language models (LLMs) like GPT-3.5 through a process of de-semanticization and skeleton retrieval. The framework addresses the challenge of accurately generating SQL queries from natural language questions by first stripping the input question of its schema-specific details, creating a generalized question skeleton. This skeleton allows for more effective retrieval of similar examples, which are then used to construct prompts for the LLM. By focusing on the structural similarities between questions, rather than their specific content, the model can generate SQL queries that are more accurate and aligned with the intended query logic. Additionally, the framework includes a fallback mechanism that revises SQL queries when initial attempts fail, further refining the output. This method has shown to significantly outperform state-of-the-art models on benchmarks like Spider, demonstrating superior generalization abilities across various domains. This research highlights the importance of structural alignment in prompting LLMs for complex tasks, offering a more robust and adaptable solution for Text-to-SQL generation. [Gao et al., 2024] presented DAIL-SQL, a sophisticated Text-to-SQL framework that leverages large language models (LLMs) to achieve state-of-theart performance on challenging benchmarks like Spider. The authors conducted a systematic evaluation of existing prompt engineering techniques, focusing on question representation, example selection, and example organization. Their findings highlighted the importance of structure-aware representations and optimized example selection strategies to enhance the effectiveness of LLMs in generating accurate SQL queries. DAIL-SQL introduces a novel approach that combines schema knowledge with example selection based on SQL skeleton similarities, significantly improving the token efficiency and overall accuracy of the model. The framework sets a new benchmark with an execution accuracy of 86.6%, surpassing previous leading models and demonstrating the potential of fine-tuning open-source LLMs for specific tasks. This research underscores the value of integrating detailed structural information and careful example curation in advancing the performance of LLM-based Text-to-SQL systems. [Shi et al., 2021] introduces the GAP framework, enhancing text-to-SQL models through generation-augmented pre-training. Implementing tasks such as Masked Language Modeling, Column Prediction, Column Recovery, and SQL Generation, GAP focuses on improving the model's ability to handle both natural language and structured database schema representations. Through synthetic data generation, the model is pre-trained on tasks that simulate real-world SQL queries and table structures. The findings demonstrate significant improvements, with GAP achieving state-of-the-art performance on the SPIDER and CRITERIA-TO-SQL datasets, with accuracy improvements of up to 71.8%. The framework effectively tackles challenges in complex query generation and column inference, showcasing strong generalization across domains. This work is significant for its ability to make database querying more accessible and accurate for non-expert users, impacting business intelligence and healthcare applications. [Guo et al., 2019] introduces the IRNet model, which addresses the challenge of translating natural language (NL) into SQL for cross-domain databases. The novel SemQL intermediate representation simplifies complex queries, allowing the model to handle tasks like GROUPBY and HAVING more efficiently. IRNet's architecture comprises three phases: schema linking, SemQL synthesis, and SQL inference. Schema linking connects NL with database schemas, SemQL synthesis generates an abstract query representation, and SQL inference converts SemQL into SQL using domain knowledge. Findings show that IRNet achieves 46.7% accuracy on the Spider benchmark, with a 19.5% improvement over prior models. Incorporating BERT further enhances performance to 54.7%. The SemQL intermediate representation is crucial, improving accuracy by 14.4% over models generating SQL directly. This work significantly improves handling complex queries in cross-domain databases, highlighting the potential of intermediate representations in bridging NL and SQL tasks. [Kim et al., 2020] provides a comprehensive analysis of the current state of NL2SQL research, addressing methods, challenges, and evaluation frameworks. The authors categorize NL2SQL methods into rule-based and deep learning-based approaches, highlighting deep learning models' superiority in handling complex queries. A key implementation feature is a taxonomy that classifies NL2SQL methods by input processing, translation techniques, and schema linking. The findings show that deep learning models outperform rule-based ones but still struggle with complex SQL queries involving joins and nested structures. The paper identifies flaws in current evaluation metrics and proposes a unified framework based on semantic equivalence to more accurately assess query results. The significance of this paper lies in its contribution to bridging NLP and database interactions, providing valuable insights for future research and practical applications, especially in real-world cross-domain databases. [Lee et al., 2022] explores the application of decision trees in predictive analytics (PA) within business analytics (BA), highlighting their significance in sectors such as customer relationship management (CRM), healthcare, manufacturing, and supply chain management. Decision trees are presented as a supervised machine learning (ML) algorithm that aids in classification and regression tasks by splitting data into branches based on feature values. This method's interpretability and simplicity make it particularly advantageous for decision-making processes, as compared to more complex models such as neural networks. The paper outlines a comprehensive implementation framework, including steps like project identification, data collection, data cleaning, and decision tree modeling. Findings demonstrate that decision trees have been successfully applied in industries for predicting customer churn, managing healthcare resources, detecting fraud, optimizing manufacturing processes, and improving supply chain logistics. The paper concludes by underscoring the scalability, efficiency, and versatility of decision trees, along with suggestions for future research on integrating them with other machine learning algorithms and enhancing their real-time decision-making capabilities. [Wang and Aviles, 2023] discusses integration of machine learning (ML) with business intelligence (BI) has gained substantial attention, as outlined in this paper, which focuses on leveraging predictive algorithms to enhance operational efficiency in business processes. The paper explores how ML models, particularly regression and neural networks, can optimize decision-making by analyzing historical data. By utilizing sales data from a supermarket from January 2021 to December 2022, the study applies ML models to forecast trends in sales, supply chain, and marketing strategies. Two models are emphasized: linear regression, which predicts outcomes based on linear relationships between variables like pricing and seasonality, and backpropagation neural networks, which better capture non-linear dependencies in factors such as customer satisfaction and weather. The findings highlight the superiority of neural networks, achieving a 93% accuracy rate compared to 86% for linear regression. This underscores their ability to model complex relationships in data, particularly for businesses dealing with dynamic, non-linear variables. Furthermore, the paper discusses how integrating these models with enterprise systems via APIs allows for real-time data transmission and predictive insights, helping businesses adapt swiftly to market demands. The iterative nature of ML in BI systems ensures continuous improvement as new data is collected, making operations more efficient. [Apeko et al., 2023] presents a novel approach to predicting used car prices by leveraging Bayesian Networks (BNs). This research highlights the volatility of the used car market, exacerbated by events like the COVID-19 pandemic and supply chain disruptions. The paper introduces three BN models—generalized, luxury-specific, and non-luxury-specific—to improve prediction accuracy, especially in scenarios with incomplete data. The implementation involves data preprocessing from a Craigslist-sourced dataset and constructing a Directed Acyclic Graph (DAG) to represent relationships between attributes such as year, condition, and mileage. By utilizing conditional probability tables (CPTs), the BN models perform probabilistic inference, enabling accurate predictions even with missing information. Through experiments and performance evaluation, the research demonstrates that the specialized luxury BN model outperforms the others, achieving an accuracy of 83.5%. Additionally, the models were benchmarked against existing car pricing tools like Kelley Blue Book, offering advantages in handling uncertainties and incomplete data. The significance of this work lies in its ability to reduce sales risks and improve transparency in the used car market. By providing lower and upper price bounds, the study enables better decision-making for both buyers and sellers. [Phillips-Wren et al., 2021] offer a comprehensive framework that aims to bridge the conceptual gaps between Business Intelligence and Analytics (BI&A) and Decision Support Systems (DSS). The authors emphasize the importance of reconnecting BI&A with its DSS roots, particularly to enhance decision-making processes. BI&A systems, with their focus on real-time data processing and advanced analytics, often neglect the decision-making frameworks that were central to DSS development. The paper identifies key differences and overlaps between the two fields, such as the focus on structured decision-making in DSS and the more unstructured, data-driven approaches in BI&A. Through a systematic literature review and empirical interviews with BI&A practitioners, the authors propose a process-level architecture that integrates traditional DSS components with modern BI&A capabilities. This integration offers organizations the ability to leverage both structured and unstructured data for enhanced decision-making, while also addressing gaps in current BI&A systems, particularly in the areas of cognition and user roles in decision-making. [Chandgude and Kawade, 2023] examine the pivotal role of Artificial Intelligence (AI) and Machine Learning (ML) in driving business growth through improved decision-making. They emphasize how AI and ML analyze large datasets, recognize patterns, and predict trends, helping businesses make data-driven decisions. AI's ability to automate tasks and simulate human intelligence enhances efficiency, particularly in areas like sales, marketing, and operations. ML, in turn, allows systems to learn from historical data, improving decision-making over time. The paper highlights applications such as predictive maintenance, fraud detection, and AI-driven chatbots that improve customer interactions. Additionally, AI is shown to strengthen cybersecurity by detecting potential risks quickly. ML also optimizes supply chain management by forecasting customer behavior and adjusting inventory accordingly. Overall, the authors conclude that AI and ML are essential tools for enhancing decision-making, increasing operational efficiency, and driving business profitability, making them crucial for sustained growth in a competitive market. [Adeniran et al., 2024] delves into the intersection of BI and PA in the financial sector, offering critical insights into how their integration optimizes decision-making. The authors argue that the use of BI to interpret historical data alongside PA to predict future outcomes transitions banks from reactive to proactive strategies. This is vital in a highly competitive, data-driven market where timely decisions can significantly impact profitability and customer satisfaction. The proposed framework for integration involves a unified data architecture that consolidates information from various sources, including customer transactions and regulatory reports. Through data processing and integration, banks can generate insights using BI tools, while PA models forecast behaviors such as customer needs, market trends, and risks. The study underscores the importance of high-quality data management for ensuring accurate analysis and decision-making, further highlighting operational benefits like enhanced customer engagement and risk management. Although the complexity of implementation poses challenges, such as a shortage of skilled professionals and regulatory constraints, the research reveals that overcoming these obstacles is key to maintaining competitiveness in modern banking. Ultimately, this paper emphasizes the transformative potential of BI and PA in revolutionizing financial decision-making, forecasting, and operational efficiency. [Chuma and De Oliveira, 2023] explore the potential of ChatGPT as a tool for enhancing business decision-making processes. By examining the AI's capabilities in generating responses to specific business-related questions, the paper reveals both its strengths and limitations. The authors test ChatGPT by posing three distinct business scenarios: a merger between two Swedish supermarket chains (ICA and COOP), the risks of investing in Petrobras, and factors influencing online shopping behavior. The findings show that ChatGPT excels at summarizing information and providing high-level insights. For instance, it effectively identifies the merger's general impacts and offers a basic risk assessment for Petrobras, highlighting operational, financial, and political risks. However, its responses lacked the nuanced details and expert-level insights required for more complex decision-making, such as advanced financial analysis or comprehensive legal implications. This study is significant as it showcases the evolving role of generative AI in streamlining business processes. While ChatGPT can assist in organizing and summarizing information, it cannot replace the deep analysis provided by human experts. It is positioned as a productivity tool that can support decision-making for non-experts but still requires human oversight for complex or high-stakes scenarios. The paper highlights the potential for future improvements in AI models and their growing relevance in business decision-making as AI technology continues to advance. [Goel et al., 2023] offers a significant contribution to the growing body of literature on AI applications in finance. By focusing on predictive analytics, the authors illustrate the immense potential of AI in analyzing complex datasets, detecting patterns, and improving decision-making processes in financial management. The paper covers several AI techniques, such as deep learning (DL) and reinforcement learning (RL), specifically in areas like stock price prediction, portfolio management, fraud detection, and credit risk analysis. Through the integration of CNNs and RNNs, the study highlights the role of AI in capturing sequential patterns and temporal dependencies in financial data. Additionally, reinforcement learning algorithms such as Q-learning and SARSA are employed to refine decision-making in dynamic market environments. The results demonstrated that AI-based models achieved a high degree of accuracy in financial forecasting, outperforming traditional models like decision trees. Despite the promising results, the study also identifies challenges such as data quality and model interpretability, indicating areas for future research. This work underscores AI's transformative potential in financial management, from enhanced accuracy in stock price predictions to more efficient credit risk assessment and fraud detection, and emphasizes the need for continuous advancements to address AI's current limitations. [Singh et al., 2023] highlights the transformative role of machine learning (ML) in improving business operations across various sectors. The authors emphasize the integration of ML in enhancing decision-making, noting its increasing relevance in the era of Industry 4.0, where massive datasets are generated by technologies such as IoT, healthcare systems, and cybersecurity frameworks. Dubey presents an in-depth review of key ML algorithms such as classification, regression, clustering, dimensionality reduction, and reinforcement learning, discussing their practical applications in business. Classification methods like Decision Trees and SVM are showcased in use cases such as spam detection and customer segmentation, while regression and cluster analysis are applied to time-series forecasting and HR analytics, respectively. The paper also identifies the challenges businesses face in algorithm selection and data quality, underscoring the significance of understanding ML principles. Furthermore, it explores the potential for hybrid ML models and highlights areas for future research, particularly in cybersecurity and smart cities. [Deng et al., 2022] contribute to the field of data visualization with their development of DashBot, a system that uses deep reinforcement learning to automate the generation of analytical dashboards. DashBot's reinforcement learning agent interacts with a dynamically constructed environment to learn optimal strategies for dashboard creation, guided by a reward structure based on visualization knowledge. The system has demonstrated superior performance in generating insightful and aesthetically pleasing dashboards, particularly in handling multi-view visualizations. This research is significant for its potential to streamline the dashboard creation process, making it more accessible to users without deep technical expertise in data analysis or visualization. [Franciscatto et al., 2022] developed a chatbot system designed to facilitate interaction with multidimensional datasets through natural language queries. The chatbot, built using the Xatkit framework, integrates with an open database containing billions of records, allowing users to query the database conversationally. The system was evaluated through an empirical user study, which revealed that the chatbot effectively assisted users in querying the database without prior knowledge of the schema or query languages. This research addresses a significant gap in the accessibility of complex databases, providing a conversational interface that makes data interaction more intuitive for nontechnical users. The potential of natural language interfaces in data visualization is explored by [Kavaz et al., 2023] through a scoping review on chatbotbased Natural Language Interfaces for Data Visualization (VNLIs). The review highlights the current implementations of V-NLIs, which primarily focus on navigating tabular data and generating basic visualizations. However, these systems often lack advanced guidance for users and support only basic query types. The authors suggest that future research should explore the integration of V-NLIs with advanced technologies like augmented reality (AR) and virtual reality (VR) to enhance user interaction and the complexity of supported visualizations. This study underscores the critical role of V-NLIs in making data visualization more accessible and interactive, particularly as they evolve to support more complex queries and dynamic user interactions. [Narechania et al., 2020] developed NL4DV, a toolkit designed to streamline the development of natural language interfaces for data visualization. Implemented as a Python package, NL4DV processes natural language queries and generates analytic specifications in JSON format, encompassing data attributes, analytic tasks, and visualization specifications compatible with VegaLite. The toolkit abstracts the complexities involved in natural language processing and visualization design, making it accessible to developers without specialized expertise in these fields. NL4DV has demonstrated its ability to accurately interpret a wide range of natural language queries and convert them into structured commands for data visualization, supporting various visualization types and integrating seamlessly with data science tools like Jupyter notebooks. This tool is significant in democratizing data visualization, allowing non-expert users to engage with complex data through intuitive natural language queries. [Luo et al., 2021] introduced ncNet, a groundbreaking Transformer-based sequence-to-sequence model designed to bridge the gap between Natural Language and Visualization (NL2VIS). Unlike traditional approaches that rely on semantic parsers and heuristic algorithms, ncNet utilizes deep neural networks to translate natural language queries directly into visualization specifications. The model is trained on the nvBench dataset, which comprises 25,750 natural language and visualization pairs across 750 tables from 105 domains. ncNet employs several novel optimizations, including attention forcing and visualization-aware translation, to enhance the accuracy and relevance of the generated visualizations. The introduction of ncNet, supported by nvBench, represents a significant advancement in the NL2VIS domain, enabling the creation of more accurate and contextually appropriate visualizations from natural language queries. This development is crucial for the broader application of data visualization in realworld scenarios, making it more accessible to users without a deep background in data science or visualization techniques. [Maddigan and Susnjak, 2023] introduce an innovative system that bridges the gap between natural language processing and data visualization. Chat2VIS leverages large language models like ChatGPT and GPT-3 to translate natural language prompts into visualizations, aiming to simplify the process for non-technical users. The study highlights the system's ability to overcome traditional challenges such as ambiguity and underspecification in natural language queries. By incorporating prompt engineering techniques and utilizing a combination of metadata and Python code primers, the system accurately generates visualizations from diverse datasets. The research demonstrates the robustness of Chat2VIS through case studies, showing that the system can handle typographical errors and underspecified queries while maintaining a high level of accuracy and flexibility. This work is significant for democratizing data visualization, making it accessible to users without programming expertise, and points to future possibilities for LLMs in NL2VIS applications. [Wu et al., 2021] delve into how AI technologies are transforming the landscape of data visualization by treating visualizations as new data forms. The paper categorizes AI4VIS research across key tasks such as visualization generation, enhancement, and analysis. It offers a detailed taxonomy that formalizes visualization data into graphics, programs, and hybrids, addressing the unique representation challenges of each. The survey further explores the potential of AI in generating visualizations, enhancing them with interaction or retargeting, and analyzing large collections of visual data through tasks like mining, querying, and comparison. Key findings highlight the rapid adoption of AI in automating the creation and extraction of visualizations and addressing data representation issues, such as the loss of semantic information in raster graphics. The survey identifies the role of machine learning in recommending visualizations based on templates, while also emphasizing gaps, such as the difficulty of reverse-engineering novel chart types and the lack of robust quality assessment techniques. The paper contributes a unified framework for AI in visualization, guiding future research by identifying opportunities to enhance AI's ability to understand and process complex visual encodings and large datasets. [Shen et al., 2022] provide an extensive review of Visualization-oriented Natural Language Interfaces (V-NLIs), which have become increasingly relevant due to advancements in Natural Language Processing (NLP) technologies. authors categorizes V-NLI systems based on an extended information visualization pipeline, consisting of seven stages: query interpretation, data transformation, visual mapping, view transformation, human interaction, dialogue management, and presentation. The authors highlights the challenges faced by V-NLIs, such as the inherent ambiguity of natural language queries and the need for robust systems capable of handling conversational contexts. One of the key challenges in V-NLIs is query interpretation, where systems like FlowSense and NL4DV use semantic parsing and keyword mappings to break down queries and infer user intent. Another significant stage is data transformation, where systems convert raw data into insights by supporting complex operations like relational joins and aggregations. Visual mapping focuses on selecting appropriate graphical elements for representing data, often using machine learning algorithms to recommend the best visualization based on the dataset. Other stages, such as human interaction and dialogue management, ensure that V-NLIs are interactive, user-friendly, and capable of handling multi-turn conversations. authors is notable for its comprehensive analysis of V-NLI systems and its focus on addressing the practical challenges involved in natural language-driven data visualization systems. [Yu et al., 2018] is a pivotal contribution to the domain of natural language interfaces for database querying. It stands out as a benchmark dataset designed to evaluate the generalization ability of models beyond single-domain SQL query generation. Prior datasets like WikiSQL often allowed models to memorize patterns due to their confinement to a single domain, leading to limitations in real-world applications where databases and queries are inherently diverse and complex. Spider addresses these challenges by providing 10,181 natural language questions mapped to 5,693 unique SQL queries across 200 databases spanning 138 domains. The diverse nature of these databases, paired with the complexity of SQL queries involving multiple tables, joins, and aggregations, pushes models to generalize their capabilities across varied database schemas. Evaluations using state-of-the-art models like SQLNet and TypeSQL revealed the models' limitations, achieving low exact match accuracy, especially in cross-domain settings. Spider's design, forcing models to generalize rather than memorize, significantly raises the bar for future research in the NL2SQL domain, particularly in cross-domain generalization, complex query understanding, and database schema comprehension.

A. *Gaps in Existing System*

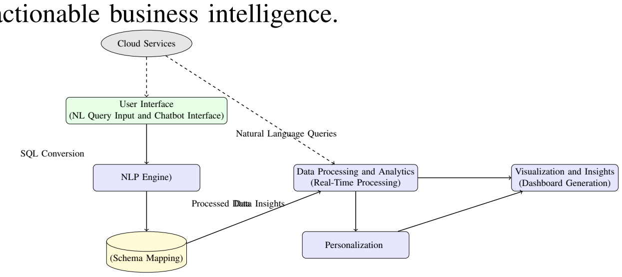
The growing reliance on Business Intelligence (BI) tools for decision-making reveals significant limitations, particularly in systems using natural language processing (NLP) for SQL queries and visualizations. Text-to-SQL models handle simple queries effectively but struggle with complex SQL structures, ambiguous inputs, and domain-specific databases, especially in real-time environments where performance and latency are crucial. Existing BI tools provide limited support for incomplete queries, real-time complex visualizations, and advanced chart customization, making them less suitable for sophisticated business needs. AI-driven BI models also lack scalability for small and medium-sized enterprises (SMEs), which often lack the computational resources needed for large-scale AI adoption. Integrating these models into legacy systems without significant infrastructure changes remains a challenge. Additionally, current models struggle to generalize across specialized industry schemas, such as healthcare and finance, and face difficulties in maintaining cost-efficiency and scalability in real-time applications. Another gap lies in the lack of integration with external knowledge and multi-turn dialogues, which limits the system's ability to refine queries iteratively and provide context-specific insights. Moreover, BI systems lack interpretability, offering minimal explanations for why certain queries or visualizations are generated, which diminishes user trust in decision-critical environments. The absence of personalization features, such as adapting to user preferences or expertise levels, further limits accessibility and usability. Support for multi-modal inputs—text, voice, and visual data—remains rare, constraining the flexibility of user interactions. Additionally, data security and privacy concerns, particularly in highly regulated industries like healthcare and finance, are inadequately addressed. This lack of privacy-preserving algorithms and compliance with legal frameworks, such as GDPR and HIPAA, leaves considerable gaps in the practical deployment of AI-driven BI systems.

III. Proposed System

The proposed architecture for the Business Intelligence (BI) system integrates advanced Natural Language Processing (NLP) technologies, such as GPT-4, to facilitate a user-friendly interface that allows users to interact with complex data systems conversationally. This architecture harmoniously integrates with external data sources via cloud services, enabling enriched analytics that amalgamate internal and external datasets. Central to the system is an NLP engine that interprets queries articulated in natural language and seamlessly translates them into structured SQL queries. These queries are subsequently executed against a robust relational database management system (RDBMS) tailored for high-performance and complex query handling. In addition to data retrieval and querying, the system is equipped with a dynamic data processing unit that employs real-time processing and predictive analytics, utilizing technologies such as Apache Kafka to ensure timely and relevant data insights. Personalized recommendations and strategic suggestions are generated by analyzing user interaction patterns and historical query data. Insights are then visualized through automated dashboards created using tools like Apache Superset, which simplify the presentation of complex data and enhance interpretability. The workflow is meticulously designed to maximize user engagement by obviating the need for SQL expertise, thereby simplifying the data query process and providing tailored intelligence to facilitate informed decision-making. This scalable architecture is particularly well-suited to meet the evolving demands of modern retail businesses, emphasizing operational efficiency, user personalization, and comprehensive analytical insights.

A. Challenges and Feasability

The proposed Business Intelligence (BI) application is highly feasible due to advancements in Natural Language Processing (NLP) and machine learning technologies. Models like BERT, RoBERTa, and T5 offer robust capabilities for understanding and converting natural language into SQL queries, enabling seamless interaction with complex datasets. These models, combined with tools like Apache Superset for visualization and cloud services for scalable data processing, provide a solid foundation for building a user-friendly BI system. The application is cost-effective and accessible, leveraging existing NLP frameworks and cloud platforms to democratize data access for non-expert users, thereby reducing the barriers to data-driven decision-making. However, several challenges need to be addressed. Ensuring high accuracy in natural language query interpretation and SQL conversion requires sophisticated NLP models and extensive training data to manage language ambiguities. Additionally, handling sensitive financial data demands stringent security measures to comply with data privacy regulations like GDPR and CCPA. The application must also be scalable to accommodate growing data volumes and user demand. Despite these challenges, the application remains highly relevant, offering retailers a competitive edge through personalized insights and strategic recommendations. By enabling data-driven decision-making and empowering non-technical users, the BI application aligns with the increasing adoption of AI technologies in business processes, catering to the rising demand for personalized and actionable business intelligence.



IV. Performance Evaluation

V. Experimental Design And Results

VI. Conclusions

References

[Adeniran et al., 2024] Adeniran, N. I. A., Efunniyi, N. C. P., Osundare, N. O. S., and Abhulimen, N. A. O. (2024). Integrating business intelligence and predictive analytics in banking: A framework for optimizing financial decision-making. *Finance Accounting Research Journal*, 6(8):1517–1530.

[Apeko et al., 2023] Apeko, J. D., Osunmakinde, I. O., Abdulgader, M. M., and Nwosu, K. C. (2023). Predictive analytics on used car prices using business intelligence of bayesian networks for sales risk reduction. In *2023 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, pages 1–6. IEEE.

[Chandgude and Kawade, 2023] Chandgude, V. and Kawade, B. (2023). Role of artificial intelligence and machine learning in decision making for business growth. *International Journal of Advanced Research in Science, Communication and Technology*, pages 54–58.

[Chuma and De Oliveira, 2023] Chuma, E. L. and De Oliveira, G. G. (2023). Generative ai for business decision-making: A case of chatgpt. *Management Science and Business Decisions*, 3(1):5–11.

[Deng et al., 2022] Deng, D., Wu, A., Qu, H., and Wu, Y. (2022). Dashbot: Insight-driven dashboard generation based on deep reinforcement learning. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):690–700.

[Dong et al., 2021] Dong, K., Lu, K., Xia, X., Cieslak, D., and Chawla, N. V. (2021). An optimized nl2sql system for enterprise data mart. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 335–350. Springer.

[Franciscatto et al., 2022] Franciscatto, M. H., Del Fabro, M. D., Trois, C., De Bona, L. C., Cabot, J., and Gonçalves, L. A. (2022). Talk to your data: a chatbot system for multidimensional datasets. In *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 486–495. IEEE.

[Gao et al., 2024] Gao, D., Wang, H., Li, Y., Sun, X., Qian, Y., Ding, B., and Zhou, J. (2024). Text-to-sql empowered by large language models: A benchmark evaluation. *Proceedings of the VLDB Endowment*, 17(5):1132–1145.

[Goel et al., 2023] Goel, M., Tomar, P. K., Vinjamuri, L. P., Reddy, G. S., Al-Taee, M., and Alazzam, M. B. (2023). Using ai for predictive analytics in financial management. In *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pages 963–967. IEEE.

[Gu et al., 2023] Gu, Z., Fan, J., Tang, N., Cao, L., Jia, B., Madden, S., and Du, X. (2023). Few-shot text-to-sql translation using structure and content prompt learning. *Proceedings of the ACM on Management of Data*, 1(2):1–28.

[Guo et al., 2023] Guo, C., Tian, Z., Tang, J., Wang, P., Wen, Z., Yang, K., and Wang, T. (2023). Prompting gpt-3.5 for text-to-sql with de-semanticization and skeleton retrieval. In *Pacific Rim International Conference on Artificial Intelligence*, pages 262–274. Springer.

[Guo et al., 2019] Guo, J., Zhan, Z., Gao, Y., Xiao, Y., Lou, J.-G., Liu, T., and Zhang, D. (2019). Towards complex text-to-sql in cross-domain database with intermediate representation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.

[Kavaz et al., 2023] Kavaz, E., Puig, A., and Rodríguez, I. (2023). Chatbot-based natural language interfaces for data visualisation: A scoping review. *Applied Sciences*, 13(12):7025.

[Kim et al., 2020] Kim, H., So, B.-H., Han, W.-S., and Lee, H. (2020). Natural language to sql: Where are we today? *Proceedings of the VLDB Endowment*, 13(10):1737–1750.

[Lee et al., 2022] Lee, C. S., Cheang, P. Y. S., and Moslehpour, M. (2022). Predictive analytics in business analytics: decision tree. *Advances in Decision Sciences*, 26(1):1–29.

[Li et al., 2024] Li, J., Hui, B., Qu, G., Yang, J., Li, B., Li, B., Wang, B., Qin, B., Geng, R., Huo, N., et al. (2024). Can llm already serve as a database interface? a big bench for large-scale database grounded text-to-sqls. *Advances in Neural Information Processing Systems*, 36.

[Luo et al., 2021] Luo, Y., Tang, N., Li, G., Tang, J., Chai, C., and Qin, X. (2021). Natural language to visualization by neural machine translation. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):217–226.

[Maddigan and Susnjak, 2023] Maddigan, P. and Susnjak, T. (2023). Chat2vis: generating data visualizations via natural language using chatgpt, codex and gpt-3 large language models. *Ieee Access*, 11:45181–45193.

[Majhadi and Machkour, 2024] Majhadi, K. and Machkour, M. (2024). Chat-sql: Natural language text to sql queries based on deep learning techniques. *Journal of Theoretical and Applied Information Technology*, 102(12).

[Narechania et al., 2020] Narechania, A., Srinivasan, A., and Stasko, J. (2020). Nl4dv: A toolkit for generating analytic specifications for data visualization from natural language queries. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):369–379.

[Phillips-Wren et al., 2021] Phillips-Wren, G., Daly, M., and Burstein, F. (2021). Reconciling business intelligence, analytics and decision support systems: More data, deeper insight. *Decision Support Systems*, 146:113560.

[Pourreza and Rafiei, 2024] Pourreza, M. and Rafiei, D. (2024). Din-sql: Decomposed in-context learning of text-to-sql with self-correction. *Advances in Neural Information Processing Systems*, 36.

[Shen et al., 2022] Shen, L., Shen, E., Luo, Y., Yang, X., Hu, X., Zhang, X., Tai, Z., and Wang, J. (2022). Towards natural language interfaces for data visualization: A survey. *IEEE transactions on visualization and computer graphics*, 29(6):3121–3144.

[Shi et al., 2021] Shi, P., Ng, P., Wang, Z., Zhu, H., Li, A. H., Wang, J., dos Santos, C. N., and Xiang, B. (2021). Learning contextual representations for semantic parsing with generation-augmented pre-training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13806–13814.

[Singh et al., 2023] Singh, A., Dwivedi, A., Dubey, S., and Lakhmani, V. (2023). Integrating machine learning in business decision making: Application and future directions. In *2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, pages 397–401. IEEE.

[Wang and Aviles, 2023] Wang, F. and Aviles, J. (2023). Enhancing operational efficiency: Integrating machine learning predictive capabilities in business intellgence for informed decision-making. *Frontiers in business, economics and management*, 9(1):282–286.

[Wu et al., 2021] Wu, A., Wang, Y., Shu, X., Moritz, D., Cui, W., Zhang, H., Zhang, D., and Qu, H. (2021). Ai4vis: Survey on artificial intelligence approaches for data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):5049–5070.

[Yu et al., 2018] Yu, T., Zhang, R., Yang, K., Yasunaga, M., Wang, D., Li, Z., Ma, J., Li, I., Yao, Q., Roman, S., et al. (2018). Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-sql task. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3911–3921.