

Data And AI Synergy Hub: Driving Migration and Transformation

Ms. Nethra S

Department of Artificial Intelligence and Data Science
Kumaraguru College of Technology, Coimbatore, India

Abstract—In the current scenario, businesses grapple with the persistent challenges of predictive modeling complexity and ineffective data communication in customer data analysis, impeding efficient decision-making processes. Our AI platform offers a timely solution by providing interactive visualizations, efficient query handling via a chatbot, and systematic data and AI transformation. The ongoing struggle of manually sifting through data for insights underscores the critical need for our solution in optimizing operations and enabling informed decision-making.

Keywords—AI Synergy; Data interpretation; Large Language Model; Predictive modeling complexity; Query handling; SAP Analytics Cloud; Zephyr

I. INTRODUCTION

Initially, the exploration of data visualization and prediction was conducted within the SAP Analytics Cloud (SAC) platform, a cloud-based solution offered by SAP. SAC boasts powerful features for data analysis, visualization, and predictive modeling. However, due to the lack of transparency in prediction algorithms within the cloud environment, concerns arose regarding the opacity of the underlying processes driving predictive insights, leading to uncertainties in decision-making. Hence the mission embarked on developing a comprehensive solution that not only addresses these concerns but also enhances the entire process of data transformation and migration. The objective was to bridge the gap between data visualization, prediction, and migration while ensuring enhanced transparency and precision in predictive modeling.

The research paper introduces a platform that involves several key steps. Firstly, data extraction and transformation are performed. This ensures that the data is cleansed, structured, and prepared for analysis. Subsequently, predictive modeling is carried out using advanced techniques, with an emphasis on transparency and interpretability. Furthermore, the platform focuses on providing visualizations and integrating a sophisticated chatbot solution for seamless handling of customer queries. This chatbot enables users to ask questions and receive instant responses based on the analyzed data. By offering intuitive visualizations and efficient query handling capabilities, the platform aims to revolutionize the way businesses approach data-driven decision-making. This streamlined approach empowers users to derive actionable insights and make informed decisions with confidence.

II. RELATED WORK

In today's rapidly evolving digital landscape, businesses are increasingly relying on AI-driven platforms to extract insights from vast amounts of data. However, several persistent challenges hinder the effectiveness and adoption of these platforms. Hence, this review will explore emerging trends and future directions in AI technology, including advancements in machine learning algorithms, natural language processing capabilities, and automation techniques. By synthesizing insights from diverse sources, this section seeks to provide valuable insights and recommendations for building a more effective and adaptive AI platform that meets the evolving needs of businesses in the digital age. The paper by Sabai Phyu et.al[2] explained the Prophet library's features as a platform for predicting the local meteorological conditions. The estimation

of absolute and relative forecasting errors, at least for a weekly time frame, showed that the Prophet package is a good tool for forecasting changes in temperature. However, challenges in this paper primarily revolve around accurately considering various factors such as the model's sensitivity to different input parameters, the inherent uncertainty in meteorological data, and the potential for abrupt changes in weather patterns.

The paper by Yash Saxena et.al[3] stated that as the signals from the stock market are not linear in nature the Prophet model produced results with little data and is dependable. Time-series data with strong seasonal impacts are best suited for it. In this paper Prophet was able to handle anomalies well and is robust against missing data. Besides its advantages, this paper still poses challenges that the stock market can be influenced by a myriad of factors beyond seasonal trends, including geopolitical events, economic indicators, and investor sentiment. Ensuring that the model effectively captures and incorporates these diverse influences is crucial for accurate forecasting.

An article by Bineet Kumar Jha et.al[2] proposed a Facebook Prophet tool to anticipate supermarket sales data. A few forecasting models, including the additive model, the Autoregressive integrated moving average (ARIMA) model, and the FB Prophet model, have been studied in the suggested research project. According to the proposed research work, FB Prophet performs better as a prediction model in terms of low error, prediction, and fitting. While the article highlights the benefits of using Facebook Prophet and compares it with other forecasting models such as additive models and ARIMA, selecting the most appropriate model for a specific problem domain remains a challenge. The performance of forecasting models can vary depending on factors like data quality, feature selection, model parameters, and the presence of outliers or anomalies in the data.

A paper by Vincent Velasco et.al[4] explains a systematic literature review that evaluates research on the application of AI chatbots to disease prediction. Chatbots are capable of improving the speed and accuracy of disease diagnosis through the use of machine learning algorithms and methodologies. To enhance the capabilities of AI chatbots and transform healthcare delivery for better patient outcomes and

disease management, further research and development work is required. Integrating AI chatbots seamlessly into existing healthcare systems and workflows can be challenging due to interoperability issues, legacy infrastructure, and resistance to change among healthcare providers.

A paper by Mahdi Banisharif et.al[3] explains that the development of a BI chatbot is a challenging undertaking that calls for the cooperation of technical and BI specialists. In this research paper, a model-driven method is presented for automatically creating customized BI chatbots for businesses in order to solve this issue. The solution has two components, one for allowing the user to model the business-specific chatbot and another for automatically generating the chatbot code. The challenges in this paper that need to be addressed include Heterogeneity of Organizational Requirements. Organizations have diverse needs and structures, leading to variability in BI requirements. Developing a BI chatbot that caters to these diverse requirements poses a challenge in terms of customization and personalization.

An article by Hrushikesh Koundinya K et.al[3] aims to implement an online chatbot system to assist users who access college websites with the help of tools that expose Artificial Intelligence techniques like Natural Language Processing. Users will be able to communicate with the college chatbot using natural language input, and the chatbot will be trained using the proper Machine Learning techniques so it will be able to generate a response. The challenges that need to be solved include generating appropriate and accurate responses to user queries that require a robust knowledge base and advanced algorithms. Ensuring that the chatbot selects the closest matching response and provides relevant information is essential for enhancing user satisfaction.

In conclusion, the diverse range of research papers reviewed showcases the versatility and effectiveness of AI-driven solutions across various domains. From weather forecasting to stock index probability prediction, traffic matrix prediction, and even disease diagnosis, AI technologies such as the Prophet model and chatbots have demonstrated remarkable capabilities in enhancing prediction accuracy, streamlining processes, and transforming traditional practices. The studies underscore the significance of

leveraging machine learning algorithms, natural language processing techniques, and model-driven approaches to address complex challenges and deliver tailored solutions. Furthermore, the consistent findings across different applications highlight the robustness and reliability of AI methodologies, indicating their potential to revolutionize diverse sectors and improve decision-making processes.

III. PROPOSED SYSTEM

The proposed concept as in Fig. 1, presents a holistic framework to elevate data utilization and insight extraction for businesses, with a specific emphasis on leveraging data. This system integrates a series of interconnected steps aimed at maximizing the value derived from data and facilitating informed actions.

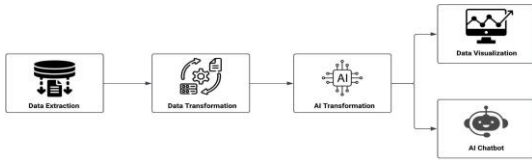


Fig. 1. Block Diagram of System Flow

Firstly, the process commences with the critical phase of Data Connection and Extraction, constituting the foundational step in harnessing valuable sales data. This pivotal stage entails the deployment of a robust data extraction mechanism meticulously designed to navigate diverse data sources with precision. It operates as a bridge between disparate data environments, orchestrating the extraction of pertinent sales data with efficiency. By consolidating data from various sources, the platform lays a robust foundation for subsequent analysis and interpretation, empowering organizations to derive actionable insights and strategic recommendations. Following data extraction, the methodology progresses to Data Transformation. Here, comprehensive data transformation processes are implemented to convert raw sales data into structured, analyzable formats. Data preprocessing techniques are applied to handle missing values, outliers, and data inconsistencies, ensuring data quality and reliability. Next, AI Transformation is employed to leverage predictive

modeling techniques for forecasting future sales quantities accurately. Predictive insights derived from these models are then utilized to optimize revenue streams, improve inventory management strategies, allocate resources efficiently, and enhance overall customer satisfaction levels.

In the subsequent step of Data Visualization, dynamic and interactive visualizations of sales data trends and forecasts are created. Intuitive dashboards and customizable reports are designed to meet specific user requirements, enabling stakeholders to explore insights, identify trends, and make informed decisions effectively. Finally, Chatbot Integration is employed to provide conversational access to sales data insights and analytics. A user-friendly chatbot interface is developed, seamlessly integrating with data ingestion and processing mechanisms to enable real-time interaction. Natural language processing capabilities interpret user queries accurately, generating relevant responses based on underlying sales data.

IV. METHODOLOGY

A. Data Extraction

To implement the proposed system as shown in Fig. 2 effectively, the first crucial step involves creating custom connectors tailored to specific data sources and destinations. These connectors serve as bridges between disparate systems, ensuring seamless data transfer and compatibility across platforms. Custom connectors are designed to understand the data structure and protocols of each source and destination system, facilitating efficient and reliable data extraction and transfer processes. Once the custom connectors are configured, they facilitate the extraction of data from customer databases and other relevant sources. This extracted data is then transferred to designated destinations for further processing, such as data warehouses or analytical platforms. During this transfer process, it's essential to ensure data integrity and security to safeguard sensitive information.

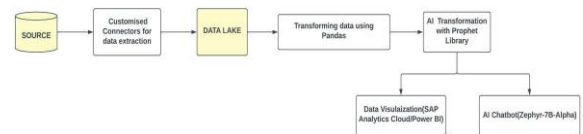


Fig. 2. System Overview

B. Data Transformation

Following the successful extraction of data, the next step involves data transformations. Leveraging the capabilities of Pandas, a powerful Python library for data manipulation and analysis, efficient data transformations can be executed. Pandas provides a wide range of functions and methods for cleaning, filtering, aggregating, and reshaping data. Specifically, Pandas allows for the conversion of raw data into structured, analyzable formats. This involves tasks such as parsing dates, handling missing values, standardizing data formats, and aggregating data at different levels of granularity. By applying these transformations, the raw data is transformed into a more structured and uniform format, making it ready for further analysis and insights generation. Furthermore, Pandas enables the implementation of complex data processing pipelines, allowing for the automation of repetitive tasks and the integration of various data manipulation techniques. This enhances the efficiency and scalability of the data transformation process, enabling organizations to handle large volumes of data effectively.

C. AI Transformation

In the phase of advanced predictive modeling, sophisticated algorithms such as the Prophet library are deployed and meticulously trained using historical sales data. To begin, historical sales data spanning the designated time period is utilized to train the Prophet model. This process involves feeding the model with a rich dataset comprising information about past sales quantities, corresponding timestamps, and any relevant contextual features. The model then analyzes this historical data to identify underlying patterns and establish relationships between sales performance and various influencing factors. One of the key strengths of the Prophet library lies in its ability to automatically detect and incorporate seasonality patterns inherent in the data. Errors are calculated in terms of mean absolute error, mean squared errors, and Root Mean Square errors to check for accurate predictions. By analyzing the periodic fluctuations in sales over time, the model can capture recurring patterns associated with daily, weekly, monthly, or yearly cycles.

D. Data Visualization

After the forecasting process, the next step involves leveraging powerful visualization tools such as the SAP Analytics Cloud Platform (SAC) or Power BI to craft visually compelling and highly informative representations of sales data trends and forecasts. These visualization dashboards serve as critical tools for stakeholders to gain insights and make data-driven decisions. Firstly, the data extracted and transformed from various sources is imported into the visualization tool. In Fig. 3, the visualization provides a user-friendly interface for importing and organizing this data, allowing users to easily access and manipulate datasets. Advanced features such as interactive filters, drill-down capabilities, and linked visualizations allow users to explore sales data from multiple perspectives and uncover hidden insights. Hence, these visualization dashboards empower stakeholders to make informed decisions, drive business growth, and stay ahead of market trends.



Fig. 3 . Data Visualization in SAC

E. Chatbot Integration

In the final phase of the methodology, Chatbot Integration is strategically employed to provide conversational access to data. The Zephyr-7B Alpha LLM serves as the core engine behind the chatbot's natural language processing capabilities. Fig. 4, shows that Zephyr outperformed the other models with a high MT Bench score. Being the fine-tuned version of Mistral-7B, Zephyr-7B Alpha LLM processes these queries with remarkable accuracy, leveraging its deep linguistic understanding to extract relevant information from the underlying sales data. PySpark, a powerful framework for distributed data processing, is utilized to efficiently handle the extraction and manipulation of both

original and predicted data from CSV files. It provides robust capabilities for data ingestion, transformation, and analysis, enabling seamless access to the structured data stored in CSV format. On the other hand, Langchain serves as a critical component responsible for orchestrating the flow of information between PySpark and the Zephyr-7B Alpha LLM-powered chatbot interface. It facilitates seamless integration and communication between these disparate systems, ensuring smooth and efficient data retrieval and processing. Thus, the chatbot can effectively deliver the requested insights to the user in a natural language response.

Model	Author	Method	Avg. MT Bench
Llama-2-70b-chat-hf	Meta	RLHF	6.856250
zephyr-7b-alpha	Hugging Face	DPO	7.091750
Mistral-7B-Instruct-v0.1	Mistral	SFT	6.694969

Fig. 4 . Comparison of Zephyr with other Models

V.CONCLUSION

In conclusion, the proposed project represents a comprehensive and innovative approach to harnessing the power of artificial intelligence and advanced data analytics for optimizing sales data utilization and decision-making processes within businesses. By integrating cutting-edge technologies such as the Prophet forecasting library and Zephyr-7B Alpha LLM-powered chatbot interface, the project offers a holistic solution to address the diverse needs of organizations in today's data-driven landscape. By leveraging Langchain and PySpark for seamless data retrieval and processing enhanced the utility and scalability of the solution. In essence, it is a transformative approach to data analysis and interpretation, offering businesses the tools and technologies needed to drive growth, optimize operations, and achieve sustainable success in the ever-evolving marketplace. Through continuous innovation and refinement, this would lay the foundation for data-driven decision-making and strategic excellence in the digital age.

VI. RESULT

Fig. 5, shows the answers given by the chatbot for the queries asked by the user from the data present in the CSV file. Moreover, by incorporating natural language processing (NLP) capabilities, the chatbot has enhanced its ability to understand user queries and respond with better-framed sentences for improved user understanding. This combination of data retrieval and NLP has resulted in a more effective and user-friendly interaction experience, enabling users to obtain the desired information more efficiently and accurately.

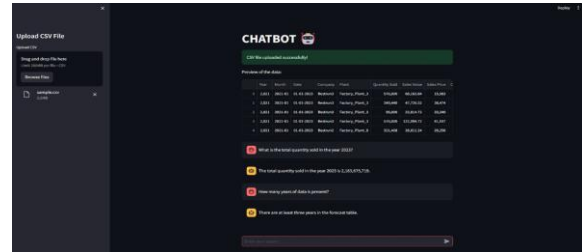


Fig. 5 . Sample Conversation

VII. REFERENCES

- [1]. Gnanasekaran, H., Dhinakaran, P., & Köse, U., 2023, November. Time-series Forecasting of Web Traffic Using Prophet Machine Learning Model. SRM Institute of Science and Technology, T.C. Süleyman Demirel Üniversitesi.
- [2]. Jha, B. & Pande, S., 2021, April. Time Series Forecasting Model for Supermarket Sales using FB-Prophet. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). DOI: 10.1109/ICCMC51019.2021.9418033. CMR Institute of Technology.
- [3]. Oo, Z.Z. and Phyu, S., 2020, December. Time Series Prediction Based on Facebook Prophet: A Case Study, Temperature Forecasting. International Journal of Applied Mathematics Electronics and Computers, 8(4), pp.263-267. DOI: 10.18100/ijamec.816894. License: CC BY-SA 4.0.
- [4]. Sharma, K., Bhalla, R., & Ganesan, G., 2022, August 29 – 30. Time Series Forecasting Using FB-Prophet. In ACM-2022: Algorithms Computing and Mathematics Conference, Chennai, India.
- [5]. Sankar, R., 2018, February. EMPOWERING CHATBOTS WITH BUSINESS INTELLIGENCE BY BIG DATA INTEGRATION. International

Journal of Advanced Research in Computer Science, Volume 9, No. 1, pp. (DOI: 10.26483/ijarcs.v9i1.5398). Cochin University of Science and Technology.

[6]. Kumar, A.K., 2020, December. Smart College Chatbot using ML and Python. In ICSCAN 2020. DOI: 10.1109/ICSCAN49426.2020.9262426. Matrusr Engineering College.

[7]Chien, S.Y.; Hwang, G.J. A research proposal for an AI chatbot as virtual patient agent to improve nursing students' clinical inquiry skills. ICAIE 2023, 2023, 13. [Google Scholar]

[8]. Ozkaya, I., May-June 2023. Application of Large Language Models to Software Engineering Tasks: Opportunities, Risks, and Implications. IEEE Software, Volume 40, Issue 3, pp. 4-8. DOI: 10.1109/MS.2023.3248401. Publisher: IEEE.

[9] Sudharson, D., Bhuvaneshwaran, A., Kalaiarasan, T. R., Sushmita, V., March 2023. A Multimodal AI Framework for Hyper Automation in Industry 5.0. 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), pp. 282-286. Publisher: IEEE.

[10] Bhuvaneshwaran, A., Amitha, K., Bhavya, A., Kamali, S., Karishma, R., Deisy, C., Sridevi, S.,

December 2023. Sentimental Analysis Using Deep Learning Models. 2023 International Conference on Energy, Materials and Communication Engineering (ICEMCE), pp. 1-6. Publisher: IEEE.

[11] Vani, K., Kasthuri, M., Bhuvaneshwaran, A., March 2023. Enabling Cloud Storage Auditing with Verifiable. 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), pp. 287-291. Publisher: IEEE.

[12] Bhuvaneshwaran, A., Mallikarjunan, K. N., Shalinie, S. M., 2019. BAIT: Behavior Aided Intruder Testimony Technique for Attacker Intention Prediction in Business Data Handling. International Journal of Business Intelligence and Data Mining, Volume 14, Issue 1-2, pp. 177-198. Publisher: Inderscience Publishers (IEL).