IIMB

तेजस्वि नावधीतमस्तु

15th Symposium on Business Analytics and Intelligence 2024-25

BOOK OF ABSTRACTS

List of Projects

Project Title: Automating Data Intelligence: AI-Driven Cataloguing & Chatbot for multi-national manufacturing company

Abstract: The Company's AI-Driven Data Transformation Project addresses key challenges in modern data management, including manual metadata generation and complex data mapping processes that impede efficiency and rapid insight generation. Current practices such as manual cataloging and heavy reliance on technical experts for data mapping result in delayed data reports, analysis and increased operational costs.

This project introduces an AI-powered web application, leveraging Azure OpenAI's LLM models to automate metadata creation, streamline data mapping, and dynamically generate transformation code. The solution provides a user-interface for validation of AI generated metadata and mapping, ensuring accuracy and compliance, and a conversational AI chatbot that enables data engineers to interact with and query metadata and data models using natural language. As a result, the time required for data catalog optimization and report creation traditionally dependent on extensive manual mapping is dramatically reduced, thus accelerating report delivery and overall decision-making.

Key outcomes include making data 'AI-ready' by optimizing catalogs and metadata, achieving 75% automation efficiency, reducing manual data handling by 40%, and accelerating decision-making by 30%. By minimizing dependence on specialized roles and integrating robust security measures for data governance, the project not only lowers operational costs but also drives higher adoption of AI/ML technologies, fosters collaboration, and enhances data usability.

By uniting automated, AI-enabled data processes with an intuitive interface, this project empowers company's data engineering team to rapidly generate insights from optimized, compliant, and future-ready data assets unlocking faster, more reliable, and scalable data-driven decision-making.

Project Title: Dynamic Capacity planning for Multi-Tiered Service SLAs

Abstract: There are lot of challenges in predicting support ticket volumes and required engineer capacity based on multiple factors such as ticket priority, service tier, engineer productivity, and case complexity. This is to optimize resource allocation based on different service tiers (e.g., Basic, Activate, Enterprise) and their associated Service Level Agreements (SLAs) for a technology product company.

Currently the change in the ticket flow is not known for specific priority, account type and for specific region. Hence optimal staffing could not be determined due to above factors.

This capacity planning tool aims to not only estimate required engineer headcount by mapping support volume to productivity levels but also will enable dynamic adjustments if SLAs or Service Tier (Basic, Enterprise etc.) is changed.

This tool will also help in reducing costs through proactive planning and self-service mechanisms, enhancing engineer productivity, improving forecasting accuracy to mitigate risks from seasonal fluctuations.

The prediction/forecasting models will be developed using statistical/ML techniques for time series data. The output of this forecasting will be fed to the Queuing model which will predict the required engineer headcount for different scenarios. This will help in introducing real-time adjustments based on incoming ticket volume.

Project Title: From Notes to Action: AI-Powered Prescription Structuring with Generative AI to Reduce Hospital Queue Time.

Abstract: Medical prescriptions in hospitals are often recorded in free-text format, especially when physicians find digital systems sluggish or unintuitive. This results in underutilization of structured modules in hospital management systems (HMS), leading to billing delays, pharmacy inefficiencies, and missed opportunities for operational optimization. At a leading healthcare facility, prescriptions are typically handwritten and later transcribed manually by pharmacists or lab staff for billing, contributing to average queue times of 7 minutes, with over 3 minutes spent interpreting each prescription.

This project introduces a smart prescription extraction system that converts unstructured medical notes into structured, actionable data. It extracts key elements such as patient details, diagnosis, medicines, lab/radiology tests, precautions, and follow-up instructions using a Generative AI model based on the LLaMA transformer architecture. To ensure accuracy, it incorporates a retrieval-augmented mechanism that cross-validates medicine names against the hospital's internal billing SKU list. Medicine identification leverages a combination of Normalized Exact Match, Strength-Based Match, Prescription-Based Match, and Prefix-Based Match, along with a prediction module to resolve ambiguous or partial entries.

Beyond structuring, the system automates downstream actions: extracted medicines are converted into pharmacy bills, lab tests are mapped to procedure codes, and follow-up notes are prepared—eliminating redundant manual effort. This automation reduces delays across pharmacy, diagnostics, billing, and discharge workflows.

Overall, the solution shortens the hospital journey, improves prescription accuracy, reduces manual errors, and enhances patient experience. It also lays the foundation for future clinical decision support and predictive analytics using structured prescription data.

Project Title: IntelliMove: A RAG-Powered platform for actionable Insights.

Abstract: Data is the new oil, but it's crude without refinement. In the era of data-driven decision-making, enterprises face a critical challenge: transforming fragmented, unstructured urban mobility data into actionable insights. Traditional dashboards often fall short—limited by rigid interfaces and the need for technical expertise—leaving valuable knowledge untapped.

This project addresses that gap by introducing a scalable, domain-adapted Retrieval-Augmented Generation (RAG) model tailored to provide insightful needs. The goal is to empower real-time, natural language interaction with enterprise mobility documents, thereby accelerating and democratizing decision-making across teams.

The platform is built on a modular architecture featuring a robust document ingestion pipeline that supports diverse formats such as PDF, CSV, DOCX, and PPT. Using Sentence Transformers for embedding and semantic indexing via FAISS and ChromaDB, the system ensures fast and accurate retrieval. On the generative side, models like GPT-4.0 Mini and Mistral-7B are evaluated for producing context-rich, document-grounded responses.

To enhance relevance and coherence, the system integrates advanced features including query expansion, chat history tracking, and intelligent document chunking. Early testing indicates strong performance in both response latency and relevance across a wide range of mobility-related queries, with a target accuracy of 90% based on benchmarked queries and expert evaluation.

Designed for accessibility, the platform supports both structured Q&A and exploratory analysis, making it valuable for marketing teams, analysts, and non-technical users alike. A lightweight, interactive front end is being developed using Streamlit. Future enhancements include multimodal input support (e.g., geospatial data, visual charts) and domain-specific fine-tuning of language models for even greater precision.

This work underscores the transformative potential of GenAI-powered retrieval systems in enterprise environments—where interpretability, flexibility, and speed are not just advantages, but necessities.

Project Title: Next Best Action Recommendation for Physician Engagement using Reinforcement Learning

Abstract: In the competitive pharmaceutical industry, Healthcare provider (HCP) engagement involves complex, multi-channel promotional strategies-including emails, calls, digital interactions, face-to-face meetings, conferences etc. with variable waiting periods between touchpoints. However, the lack of a systematic, data-driven approach to sequencing these engagements often leads to inefficient outreach and missed opportunities. This project aims to design a reinforcement learning driven Next Best Action model to predict the optimal sequence of promotional interactions that maximize the probability of conversion while minimizing competitor wins. Leveraging historical data on HCP engagement (e.g., email & text interactions, conference attendance, meeting success, etc.) and prescribing outcomes, the model integrates clustering, Markov Decision Processes (MDPs) for sequential decision-making under uncertainty, and machine learning algorithms to capture temporal dependencies in interaction sequences.

The solution will allow the marketing and sales teams to tailor outreach strategies dynamically, optimize waiting periods between interactions, and improve resource allocation across channels. By analysing the sequence of actions and their outcomes, the model provides actionable insights for both new and ongoing practitioner engagements. For new practitioners, it suggests the optimal initial action; for existing sequences, it recommends the next action to maximize sales probability. This data-driven approach aims to enhance promotional efficiency, reduce wasted efforts, reduce promotional fatigue among HCPs and improve sales conversion rates, ultimately providing a scalable solution to strengthen practitioner relationships and drive business growth.

Project Title: Optimizing Job and Candidate Matching with AI and Graph Theory

Abstract: Job matching is a critical process of connecting qualified candidates with appropriate job opportunities, creating value for both employers and job seekers. Some of the challenges in effective job matching include the difficulty in objectively evaluating the alignment between candidate skills and job requirements on a scale. Currently, many recruitment processes rely heavily on manual resume screening and keyword matching, which is time-consuming, subject to human bias, and often fails to identify optimal candidate-job pairings. Therefore, this project aims to develop a data-driven job matching system that leverages LLM, Generative AI, and Graph Theory to enhance recruitment decision-making.

The project aims to develop an LLM and Graph Theory-based solution for recruitment processes which can enable the following:

1. Bidirectional ranking of resumes and job descriptions
2. Skill-based community detection for candidates
3. Skill-based community detection for companies

Our methodology leverages Llama3-8b-8192 as our primary model after exploring over 30 other LLMs including Mixtral-8x7b-32768, DeepSeek-r1-distill-llama-70b, and Cohere AI. These models extract key KPIs from resumes (key skills, location, education, experience, total years, project descriptions) and JDs (skills, responsibilities, minimum and maximum years of experience, education, location). We then perform similarity matching using cosine similarity, Euclidean distance, and other vector embedding techniques to rank matches bidirectionally.

Implementing this AI-powered matching system would directly result in reduction of time spent on manual resume screening. This would enable more accurate candidate evaluation for HR professionals and enhance their decision-making capabilities. This will also reduce the threat of poor hiring decisions and employee turnover, thereby ensuring minimal recruitment costs and better talent acquisition for businesses.

Project Title: Personalized Care Plan Generation Using Time Series Data from Fitness Tracker Devices and LLMs

Abstract: The increasing availability of continuous physiological data from wearable devices presents a unique opportunity to transform personalized healthcare. However, these granular time series signals—such as heart rate, step count, SpO_2, temperature variation, and sleep patterns—are often underutilized due to challenges in interpretation and lack of integration with clinical systems. This project addresses that gap by developing a data pipeline that fuses time series analytics with Large Language Models (LLMs) to automatically generate individualized care plans in natural language.

Multivariate time series data is collected from consumer-grade wearables like Apple Watch, Fitbit, and Oura Ring. After applying normalization, feature engineering, and fixed-window aggregation techniques, we compile per-patient data along with metadata such as diagnostic codes and patient groups. These structured summaries are then passed into open-source LLMs such as LLaMA 3 and LLaMA 4 using prompt engineering methods. The system outputs coherent, context-aware recommendations tailored to the user's physiological trends and clinical background.

The project utilizes Weights & Biases (W&B) for experiment tracking and prompt evaluation. A scoring framework is implemented to assess the LLM outputs on relevance, personalization, and clarity, enabling comparison across prompt variants and model architectures. The approach has shown promising results in compiling and contextualizing patient observations into actionable care suggestions.

In collaboration with Azodha—a platform enabling integration of EHRs with external tools—this project contributes to building intelligent, consumer-facing healthcare applications. The integration of wearable health signals with generative AI provides a scalable, low-friction pathway for enhancing digital health interventions and delivering personalized, data-driven care at scale.