Ish Kumar

Email : ishkumar@buffalo.edu
Linkedin: linkedin.com/in/ish-kumar
Noida, India
+91-9667763726

Education

- **University at Buffalo, SUNY, New York, USA -** Masters in Engineering Science (Data Science)
  Courses: Intro to Probability Theory, Machine Learning, Computer Vision and Image Processing, Data Model and Query Language
- **APJAK Technical University, India -** Bachelors in Electronics and Communication Engineering

Skills Summary

- **Languages and Tools:** Python (pandas, scikit-learn, NumPy, XGBoost), SQL (PostgreSQL, MySQL), R, MATLAB, Excel
- **ML & AI:** Random Forest, XGBoost, ARIMA, Prophet, Transformers, LLMs, Model Evaluation, A/B Testing
- **Data Visualization:** Tableau, PowerBI, Seaborn, Plotly, Matplotlib
- **Frameworks and Tools:** LangChain, FAISS, Django, OpenAI Embeddings, Retrieval Augmented Generation (RAG), Docker, AWS (S3, Lambda, Bedrock), Streamlit, Git

Experience

- **University at Buffalo, Computational Biology Lab - Research Assistant** Feb 2024 - Dec 2024
  - Developed Python-based data pipelines to process compound-protein interaction datasets, reducing preprocessing time by **35%** and enabling faster iteration for hypothesis testing.
  - Conducted exploratory data analysis and statistical validation of compound efficacy predictions, contributing to the refinement of scoring algorithms and model accuracy.
  - Built interacetive dashboards using **Streamlit** and **Plotly** to visualize compound-target relationships, supporting internal research reviews and stakeholder presentations.
- **Value Creation - Data Scientist (Part Time)** Aug 2020 - July 2022
  - Collaborated with business stakeholders to engineer **time series forecasting models (ARIMA, Prophet)** to predict temporary workforce demand, contributing to improved workforce allocation and a reduction in projected overstaffing.
  - Worked with senior data scientists to build a **random forest and XGBoost** model for employee churn prediction, incorporating tenure, performance ratings, and engagement metrics. This project contributed to improved prediction accuracy from **74%** to **86%**
  - **Designed Tableau dashboards with PostgreSQL integration** to monitor HR and campaign KPIs , reducing reporting lag by **24%**, leading to faster decision-making by HR teams.
  - Designed and analyzed A/B tests to evaluate fraud detection strategies, improving **F1-score** from **72%** to **81%** while mitigating false positives.
  - Developed NER and text classification models to extract product categories from reviews, reaching **78%** accuracy, enabling enhanced sentiment analysis for potential application in understanding commodity demand drivers.
  - Implemented **Gradient Boosting and Random Forest ensembles** to enhance anomaly detection in financial transactions, contributing to a boost in detection rates by **10%** compared to baseline models.

Academic Projects

- **Promptly – Chrome Extension for Prompt Optimization**
  - Designed and implemented a Chrome extension with FastAPI backend to optimize, rewrite, and enhance LLM prompts for for clarity, creativity, efficiency, safety, video generation, and code optimization
  - Integrated OpenRouter and Groq APIs for real-time, context-aware prompt enhancements.
  - Enabled seamless prompt detection and one-click optimization across platforms like ChatGPT and Notion.
  - Developed user-friendly popup, prompt history, and in-place replacement using JavaScript and Manifest V3.
- **Automated SQL Query Generator using Google Gemini**
  - Created a Streamlit application using Google Gemini to generate SQL queries from natural language, achieving **95%** accuracy across 30+ test cases covering diverse query types (joins, aggregations, subqueries).
  - Implemented dynamic SQL validation and optimization mechanisms, reducing syntax errors across the test suite and improving query execution time by an average of **15%** on sample database schemas.
  - Leveraged multi-modal AI to handle domain-specific queries, enhancing the system's ability to generate optimized SQL for various data types, resulting in a **40%** improvement in contextual accuracy.
- **Multiagent Retrieval-Augmented Generation (RAG) System**
  - Constructed a multi-agent system combining LangChain and FAISS for vector-based document retrieval and real-time query handling, improving response accuracy by **30%**,
  - Designed a query routing mechanism to dynamically switch between Wikipedia for general queries and a vector store for complex, domain-specific information, boosting query efficiency by **25%**