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

- **Outlier AI - Prompt Engineer (Freelance)** Aug 2024 - May 2025
  - Worked on a freelance, project-driven basis to evaluate and benchmark AI-generated code for data analysis tasks using Python libraries like NumPy, pandas, and scikit-learn.
  - Designed and iterated on prompts for LLMs, applying few-shot and chain-of-thought approaches to enhance code correctness and debugging.
  - Completed **100+** AI-generated code reviews assessing outputs for correctness, efficiency, and adherence to Python best practices, leading to a **30%** improvement in feedback precision and supporting more effective RLHF cycles.
  - Evaluated and fine-tuned model prompts and outputs, resulting in a measurable increase in output reliability and feedback precision.
- **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, Project-Based)** Aug 2020 - July 2022
  - Worked as a part-time, project-based data scientist, supporting business stakeholders with analytics and machine learning solutions across multiple initiatives.
  - Developed and maintained **time series forecasting models (ARIMA, Prophet)** to predict temporary workforce demand, contributing to a reduction in projected overstaffing by approximately **15%**
  - Contributed to employee churn prediction models (Random Forest, XGBoost), helping improve model accuracy from **74%** to textbf86% over several project iterations.
  - Designed Tableau dashboards integrated with PostgreSQL, which reduced reporting lag, leading to faster decision-making by HR teams.
  - Participated in the design and analysis of A/B tests for fraud detection, supporting improvements in F1-score from **72%** to **81%** while reducing false positives
  - Supported NLP projectsNER, Text classification for product review categorization, achieving **78%** classification accuracy and enabling more effective sentiment analysis.
  - Assisted in enhancing anomaly detection for financial transactions, contributing to a **10%** increase in detection rates 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.
- **Multi-Agent Retrieval-Augmented Generation (RAG) System**
  - Designed and implemented a modular, multi-agent RAG system combining Retriever, Summarizer, Answer Generator, and Explainability agents for efficient information retrieval and LLM-powered question answering.
  - Integrated Wikipedia API and augmented web search for robust, configurable retrieval, supporting both general and technical queries.
  - Built an interactive Streamlit UI enabling real-time query handling, explainability toggles, and dynamic configuration of retrieval parameters.