Mohit Kumar

Address: Bangalore, India, 560095   **Mobile:** +91-9717120144

LinkedIn: linkedin.com/in/mohitku   **Email:** mohit.iit16@gmail.com

Summary

Research Engineer with 7+ years of experience designing, implementing, and optimizing AI solutions at scale. Expertise in **machine learning, deep learning, and NLP**, with a strong focus on **Large Language Models (LLMs)** such as **GPT, Llama, and BERT**. Proficient in **ML frameworks (PyTorch, TensorFlow, scikit-learn)** and deployment using **Docker, and cloud platforms (AWS, GCP)**. Strong problem-solving and collaboration skills, with experience delivering high-impact AI solutions.

Work Experience

AI Labs, American Express   Bangalore, India
Senior Research Engineer   Apr 2022 - Apr 2025

- **Vocal - Voice of Customers**: Spearheaded the implementation of a new ETL pipeline, streamlining the data collection process for customer complaints and enhancing overall data integrity for advanced analysis. Developed a Retrieval Augmented Generation **(RAG)** pipeline utilizing advanced **Large Language Models (LLMs)**, significantly improving the accuracy of analysis and **reducing human efforts by more than 80%**. Collaborated with cross-functional teams to extract insights from specific datasets, delivering actionable recommendations with rapid turnaround times to support decision-making.
- **Model Interpretability for Large Language Models (LLMs)**: Developed and implemented model interpretability guidelines to ensure **transparency and trustworthiness of GPT-based Large Language Models (LLMs)** for enterprise-wide applications. Designed and introduced a novel evaluation mechanism to assess the performance, reliability, and scalability of LLMs, enabling robust deployment of LLM-based solutions across the organization.
- **TTYD - Talk to Your Data**: Responsible for developing a question answering system that could process natural language queries and deliver fast, **accurate (upto 87%)** responses to users. Deployed it into production, monitored how it was performing, identified ways to improve response times and latency, and implemented optimizations to achieve a better overall user experience.

Analytics, EXL Service   Bangalore, India
Senior Data Scientist   Jun 2019 - Aug 2021

- **Key-Concept Extraction**: Developed and published Key-Concept Extraction Engine, an unsupervised model to automatically extract topics from call transcripts, served with a user-friendly web-app with backend based on Django and PostgreSQL. Employed Dynamic topic modeling (DTM) techniques to analyze the evolution of topics of conversation over time. Achieved **85% satisfaction score** on manual validation of transcripts.

Predictive Analytics, Tvarit GmbH (https://www.tvarit.com)   Mumbai, India
Machine Learning Engineer   Aug 2018 - Dec 2018

- **Automated Predictive Analytics Tool (APA)**: Involved in development of APA tool, a Time Series Predictive Analytics tool which automatically does anomaly detection, feature selection, model tuning, and deployment as well as performance monitoring and logging.

Skills

- **Data Science/Machine learning**: Pandas, sklearn, Keras, PyTorch, TensorFlow
- **Deployment & Cloud**: Docker, AWS, GCP
- **Concepts**: Data Structures and Algorithms, Natural Language Processing, Deep learning

Publications

- **GENWISE**: https://aclanthology.org/2024.finnlp-2.8.pdf
- **Comparative Analysis of Transformers for Modeling Tabular Data**: https://arxiv.org/pdf/2311.14335

Education

Indian Institute of Technology (Banaras Hindu University), Varanasi   Varanasi, India
Bachelor of Technology   2012 - 2016

Accomplishments

- **AI Labs, American Express**: Recipient of CFR Excellence Award (2023).
- **AI Labs, American Express**: Recipient of Star performer of the year (2023).
- **EXL Service**: Recipient of Innovation award 2 years in a row (2019 & 2020).