FNU SUMIT KUMAR

sumit.usa23@gmail.com | +1(562) 254-7284 | Long Beach, CA | [Portfolio]

WORK EXPERIENCE

Research Assistant | CSUDH – IT Department   Carson, CA, USA | Mar 2023 – Present

- **Fine-tuned and deployed** Large Language Models (LLMs) (LLaMA, T5, BERT) on domain-specific datasets, leveraging transfer learning, **LoRA/QLoRA,** &prompt engineering, improving model relevance and task-specific performance by **30%**.
- Developed **ML models using Python** for **classification, clustering,** & **sentiment analysis**, improving prediction accuracy **30%**.
- **Designed** a multi-stage document retrieval pipeline integrating **BM25** & **Cross-Encoder re-ranking**, significantly enhancing **Retrieval-Augmented Generation** (RAG) query relevance.
- **Implemented** deep learning models using **PyTorch**, **NumPy**, & **TensorFlow**, optimizing training efficiency with hyper-parameter tuning and dropout techniques.

ETL Developer | IBM India Pvt Ltd   Bengaluru, India | Jun 2021 – Nov 2022

- Led **ETL process** development with **SAP BODS** and optimized SQL queries, reducing manual data handling by **40%.**
- **Automated** recurring reporting processes using **Python scripts** and **Power BI**, reducing manual reporting efforts by **30%**.
- **Implemented Slowly Changing Dimensions (SCD)** in **SAP Data Workflows** to accurately track historical data changes, maintaining data integrity for business intelligence and reporting purposes.
- **Built and maintained** interactive dashboards and visualizations in **Power BI**, presenting key performance indicators (KPIs) and trends to stakeholders, enabling data-driven decision-making, and improving business insights.

Analyst | Enuke Software Pvt Ltd   Gurugram, India | Jul 2020 – Jun 2021

- **Utilized SQL** to extract, manipulate, and analyze data from **relational (MySQL, PostgreSQL)** and **NoSQL databases (MongoDB)**, improving data accessibility and accuracy.
- Developed and maintained **data pipelines and ETL processes**, ensuring the smooth flow of structured and unstructured data for reporting and analysis purposes.
- Built interactive **data visualizations** and **dashboards** using **Power BI** and **Looker** to present insights to technical and non-technical stakeholders.
- Spearheaded ad-hoc **analytical projects**, providing quick insights to address pressing business challenges.

Student Intern | Nagarro Software Pvt Ltd   Bengaluru, India | Jan 2020 – Jun 2020

- Gained hands-on experience in **data cleaning, transformation,** and **manipulation** using **Python** (Pandas, NumPy) to prepare datasets for analysis.
- Implemented **SQL queries** to extract, filter, and aggregate large datasets from **MySQL** and **PostgreSQL** databases, optimizing query performance.
- Designed & developed **interactive dashboards** & **data visualizations** in **Power BI**, presenting insights to stakeholders.
- Used **Python libraries (Matplotlib, Seaborn)** to create visual reports for trend analysis and statistical insights.

PROJECTS

Sales Performance Dashboard   **Python, SQL, Power BI**

- Build a dashboard that analyzes sales performance, identifies top products, and tracks revenue over time.
- Cleaned and processed data with **Python** (Pandas) to fill missing values, detect outliers, and ensure consistency.
- Created an interactive **Power BI dashboard** to visualize key metrics like monthly sales, regional performance, & top products.

Fraud Detection System   **Python, Scikit-Learn, TensorFlow, PyTorch**

- Built a **fraud detection model** that analyzes financial transactions to detect anomalies using **Random Forest, XGBoost, and Neural Networks**. The system identifies suspicious activities by leveraging **feature engineering and outlier detection techniques**.

SKILLS

Languages: Python, Java, SQL, JavaScript, TypeScript.
Frame Works & Libraries: TensorFlow, PyTorch, Scikit-Learn, SAP BODS, Pandas, NumPy, Django, Flask.
Databases: MySQL, PostgreSQL, MongoDB, Redshift.
Operating Systems: Microsoft Windows, Linux.
Tools: Power BI, Tableau, Git, Docker, Kubernetes, AWS, Google Collab, Jupyter Notebook, CI/CD Pipelines.

EDUCATION

Masters in Computer Science | Cal State – Dominguez Hills   Jan 2023 – May 2025

- **Relevant Courses:** Design & analysis of Algorithms, Data Analysis, Software Engineering, Operating Systems, Software Project Planning, Artificial Intelligence, Advance Programming Languages (Python)