Anupam Verma

Engineer 1

Chennai, Tamil Nadu, India
9445864375
anupam215769@gmail.com
LeetCode
Github
Linkedin
Portfolio Website
Tableau Public

SKILLS

Generative AI
RAG, Fine-tuning, LangChain, LangGraph, Transformers, LLMs

Machine Learning
TensorFlow, PyTorch, Scikit-learn

NLP & Computer Vision
Hugging Face, NLTK, spaCy, Gensim, OpenCV

Data Analysis & Visualization
Tableau, Power BI

Programming
Python, C++, C

Frontend & Backend
Streamlit, FastAPI, Flask

Databases
MySQL, ChromaDB

Cloud
AWS, Azure, Docker

EDUCATION

Post Graduate Diploma in Data Science, Symbiosis Centre for Distance Learning
Jul 2023 – Jun 2025

B.Tech in CSE, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology
Jul 2019 – Jun 2023
Passed with 9.08 CGPA

EXPERIENCE

Comcast, Engineer 1
Jan 2023 – present

Generative AI

- Engineered a **Retrieval-Augmented Generation (RAG)** system with **LangChain** and **LangGraph**, querying millions of code lines and cutting manual reviews by **40%** through automated suggestions, bug detection, test generation, and more efficient documentation of new RDK contributions using **OpenAI GPT-4o** and **o1 models.**
- Developed an interactive **chatbot** with **Streamlit** as the front-end and **FastAPI** as the **RESTful API** back-end, reducing latency by **60%** while creating **custom vector embeddings** through **microservices**.
- Integrated multiple agents with a **ChromaDB** vector store, combining **Similarity and Full-Text Search** to improve retrieval accuracy by **95%** and enable more precise queries of RDK documentation and codebase.
- Implemented intelligent **query routing** to direct inquiries to RDK documentation, codebase, custom codebase, or patch analysis, cutting response time by nearly **25%** and optimizing retrieval based on query intent.
- Integrated a **Neo4j Knowledge Graph** to map and analyze RDK component-level dependencies, accelerating dependency mapping by **40%** and enabling precise identification of inter-component relationships and creation of architecture diagrams.
- Employed **Web Scraping** and **Selenium** to extract details on contributions like Test Procedures and Reasons for Change, leading to correct data extraction by **90%** of the times while downloading patches from Gerrit.
- Subsequently, the results of this analysis were updated in Jira tickets via the **Jira API**, improving cross-team visibility by **70%** and streamlining communication and tracking of code changes with their impact.

Tableau

- Developed **50+** interactive RDK dashboards in **Tableau**, including clone, code, and contribution metrics portals, boosting insights by **75%** and significantly improving data accessibility and efficiency for stakeholders.
- Leveraged advanced Tableau techniques to create visually appealing **charts, graphs, and maps**, representing complex data more effectively and enhancing **stakeholder** understanding by **50%** for greater overall comprehension.
- Implemented innovative data integration with **Tableau Prep Builder** for **ETL** process like cleansing and transformation, cutting data cleaning time by **85%** and improving loading efficiency into **MySQL** while maintaining a streamlined workflow.

PERSONAL PROJECTS

Customer Feedback Analysis [TensorFlow | Scikit-learn]

- Developed a multi-class sentiment analysis pipeline on Amazon product reviews (**negative, neutral, positive**) using **NLP techniques**. Compared a baseline Naive Bayes approach with several deep learning architectures—**Feed-forward (Dense), LSTM, GRU, Bidirectional LSTM, 1D ConvNet**, and a zero-shot **BERT** model—achieving an accuracy of up to **88.85%**, substantially exceeding the baseline performance of 79.78%

SafeHome Observer [PyTorch | OpenCV | Transformers]

- Implemented a security-focused project that integrates **YOLO12 object detection** and **BLIP image captioning** to identify and describe **persons or vehicles** in real-time. By analyzing **video feeds or recordings**, it enables automated detection of **potential intrusions**, providing robust, user-friendly home or office surveillance.

Movie Recommendation System, [Numpy | Pandas | Scikit-learn]

- Built a **content-based recommender** using the TMDB 5000 dataset by merging and preprocessing “movies” and “credits” metadata. Extracted features and applied **CountVectorizer** (with stop-word removal) to create numerical vectors. Leveraged **cosine similarity** and **K-Nearest Neighbors** to produce **top movie recommendations**

CERTIFICATES

- Fundamentals of Deep Learning by NVIDIA
- AWS Academy Machine Learning Foundations