

Varun Dixit

Email: varundixit4@gmail.com | Phone: +91 9673375005

[Linkedin](#) | [Portfolio](#) | [Github](#)

SKILLS:

Generative AI: Large Language Models (LLMs), Fine-Tuning, LangChain, LlamaIndex, Retrieval-Augmented Generation (RAG), AI Agents, Document AI, Prompt Engineering, HuggingFace.

Machine Learning/Deep Learning: Scikit-Learn, PyTorch, Predictive Modeling, Hyperparameter Tuning, Feature Engineering, Model Evaluation, Natural Language Processing, Computer Vision.

Data Analytics: Numpy, Pandas, Matplotlib, Seaborn, PowerBI, Tableau.

Languages and Databases: Python, SQL, C++, PostgreSQL, Milvus, MongoDB.

Cloud and MLOps: MLFlow, Azure DevOps, Azure ML Studio, Docker, FastAPI, Git.

EXPERIENCE:

JIO Platforms Limited, Mumbai

December 2023 - Present

Role: Associate Data Scientist; Team: Analytics COE

- Designed and deployed **AI-driven solutions** using **Gen-AI** and **ML** to optimize processes and support decision making for the **Refining and Marketing** division of **Reliance O2C**, collaborating with senior leaders in **Strategy and IT**.
- Led data digitization initiatives, including **Web Scraping, PDF Data Extraction, and Excel Parsing**, streamlining data extraction and insights generation, saving up to **₹3.8 lakhs per month** by reducing reliance on paid subscriptions.
- Developed an **Oil Market Report Summarizer**, automating the summarization of **50-80 page** reports to concise **6-8 page summaries**, cutting analyst time from **8 hours to 6 minutes**.
- Implemented a **Retrieval-Augmented Generation (RAG) chatbot** to query and analyze trends in monthly OMR reports, reducing manual data skimming and providing insights in **3-6 seconds** per query.
- Designed a **Custom PDF Parser** using **YOLO v7** for component classification and **Tesseract OCR** for data extraction, improving accuracy and reducing hallucinations by **80%** compared to generic parsers.
- Created a **News Summarizer and Sentiment Analysis Tool** to extract insights from major oil market sources, and an **AI-based Agreement Generator** for drafting legal terms and deal summaries for crude oil purchases.
- Contributed to organizational AI initiatives, developing POCs for **Multi-Agent Systems, Modular RAG components**, and projects like HR Chatbot, Charts-2-CSV, and SQL Query Generator to enhance in-house AI capabilities.

PROJECTS:

Customer Segmentation and Personalised Marketing Campaign Generator

- Developed a **K-Means** model to segment customers into **13 clusters** with a silhouette score of **0.42**, grouping customers based on purchasing patterns, and used AI to determine the characteristics of these clusters.
- Integrated the segmentation model with an **LLM** to generate personalised marketing campaigns and emails to maximize ROI.

Fine-Tuning an LLM to Summarise Conversational Text

- Fine-tuned a **4-bit quantized** LLM (Llama 3.2) using LoRA for parameter-efficient tuning on the **DialogSum dataset**, optimizing the model for conversational text summarization.
- Used **WandB** for real-time monitoring of training metrics and hyperparameter optimization, resulting in a **11.45% improvement in ROUGE-L** score after tuning, enhancing summarization performance.

EDUCATION:

K. J. Somaiya Institute of Technology, Mumbai

August 2019 - June 2023

Bachelors of Technology (B.Tech) in Electronics and Telecommunications Engineering

- GPA: 8.69/10
- Coursework: Machine Learning, Fundamentals of Data Science, Artificial Intelligence, Neural Networks and Deep Learning, Image Processing and Machine Vision, Data Structures and Algorithms, Database Management System.

EXTRA-CURRICULAR ACTIVITIES:

IEEE Students' Chapter KJSIT

July 2022 - May 2023

- As the Chairperson, led a team of 80+, managed the tech fest 'Renaissance' attracting 1000+ students, and organised over 10 workshops and seminars.

Students' Council KJSIT

January 2022 - December 2022

- As the CEO, led a team of 150 students to organise the cultural festival 'Surge' and sports festival 'Score', managed digital media, and documented all events.

CERTIFICATIONS:

Fellowship Program in Software Engineering with Specialisation in Data Science

- Scaler Academy - October 2021 to October 2022

Google Data Analytics Specialization

- Coursera - October 2023