Arpit Shrotriya

⊙ Harrison, New Jersey   ✉ as4893@njit.edu   ☎ +1-9734184125   in arpit-shrotriya   ⊙ cyclostone

Education

New Jersey Institute Of Technology *09/2024 – 05/2026*
M.S. in Computer Science

- GPA: 3.83/4.0
- **Coursework:** Machine Learning, Data Management System Design, Data Mining

Experience

Associate Data Scientist *New Delhi, India*
Veeyo Tech *07/2023 – 08/2024*

- Led full-stack development of three LLM-powered chatbot platforms using Gemini APIs and fine-tuned BERT models, integrating backend logic, API communication, and deployment workflows **to automate 65% of customer support cases**.
- Designed and maintained production-ready ETL systems for both structured and unstructured datasets; **reduced data preparation time by 40% and enabled smoother downstream model integration**.
- Engineered a high-accuracy (85%) NLP pipeline from scratch for classification and sentiment tagging, modularized using reusable Python packages and **deployed using scalable microservices architecture.**
- Automated large-scale data ingestion workflows using Python, AWS Lambda, and SQL, reliably processing over 1M records/day, **enhancing throughput by 30% with minimal downtime**

Data Engineer *New Delhi, India*
Sail Analytics *06/2022 – 06/2023*

- Designed and implemented robust Python-based pipelines for automated Excel reporting and email delivery, **saving 30% time in manual processes** and improving accuracy and consistency.
- **Developed a document processing microservice using Python OCR tools to extract and validate metadata (e.g., receipt IDs, totals) from 1000+ images with 95% accuracy.**
- Architected **ETL pipelines** for Master and Transactional data, improving data consolidation for reports in PowerBI and Tableau, leading to a **40% increase in report generation efficiency**
- Engineered scalable ETL architecture for merging master and transactional datasets, feeding into interactive dashboards (Power BI, Tableau) that **cut down reporting time by 40%.**

Skills

Python, SQL, HTML, Flask, Streamlit, TensorFlow, scikit-learn, Keras, HuggingFace Transformers, AWS (Lambda), Google Cloud (BigQuery), Docker, Pandas, Power BI, Git, OCR Libraries, LangChain, Backend Systems, Data Pipelines, LLM Integration, Microservices, API Development, ETL Automation

Projects

Search-Engine-With-LLM *Repo Link* ↗

- Developed an intelligent research assistant supporting NL search over ArXiv, Wikipedia, and live web.
- Designed modular backend with LangChain orchestration, integrated LLM reasoning.

Customer Churn Prediction *Repo Link* ↗

- Built a web-based churn prediction system featuring real-time user inputs and backend TensorFlow ANN-based predictions.
- Integrated REST-like architecture using Python functions, with a modular ML pipeline deployable across environments.

Project Salary Regressor *Repo Link* ↗

- Delivered a full-stack salary estimation tool using clean component-based code; supports real-time preds