

# 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: 3.66/4.0
- **Coursework:** DSA, Reinforcement Learning, Machine Learning, Data Management System Design, Data Mining, Intro to BigData

## Experience

### Associate Data Scientist

*New Delhi, India*

*Veeyo Tech*

07/2023 – 08/2024

- Engineered and launched three distinct LLM-powered chatbots utilizing Gemini APIs and fine-tuned BERT models, **resolving roughly 65% of** basic customer support inquiries without human intervention
- Designed and optimized ETL workflows for structured and unstructured data, **reducing data processing time by 40%**, enabling faster model training and deployment
- Built and fine-tuned a large-scale NLP pipeline for text classification and sentiment analysis, **achieving an 85% accuracy rate on production data**
- Engineered scalable data pipelines using Python, SQL, and AWS Lambda, automating the ingestion of 1M+ records per day from data sources, **enhancing workflow efficiency by 30%**

### Data Engineer

*New Delhi, India*

*Sail Analytics*

06/2022 – 06/2023

- **Automated client reporting** through creation of data pipelines **employing Python and HTML**, generating Excel files and delivering emails, **resulting in a 30% reduction** in manual report generation
- **Leveraged Python OCR packages**, extracting key information like total amount, date, and receipt ID from **over 1,000** receipt photos with **95% accuracy**, delivering valuable data to clients quickly
- 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**
- **Streamlined** secure data movement between SAP HANA and Google BigQuery, enabling secure data movement for the deployment of ETL pipelines, resulting in a **20% reduction in data transfer time**

## Skills

Python, SQL, TensorFlow, Keras, HuggingFace Transformers, LLMs, Pydantic AI, Pandas, BigQuery, ETL Pipelines, AWS, Google Cloud, Power BI, Streamlit, Flask

## Projects

### Search-Engine-With-LLM

[Repo Link](#) 

- Architected an AI research assistant utilizing Streamlit and LangChain, enabling natural language searches across ArXiv, Wikipedia, and web content, ultimately reducing research time for data scientists by 30%
- Configured Streamlit for UI, LangChain for orchestration, Groq LLM for NLP, SerpAPI for search & Python

### Customer Churn Prediction

[Repo Link](#) 

- Constructed and deployed an interactive churn prediction tool using TensorFlow and Streamlit, empowering business users to forecast churn risk with 95% accuracy based on user input
- Integrated ANN, pandas, numpy, scikit-learn, TensorFlow, and Streamlit for modeling & deployment

### Project Salary Regressor

[Repo Link](#) 

- Developed an interactive salary estimation dashboard using Streamlit, enabling users to input skills and experience for instant salary projections, achieving high satisfaction rating of 4.8/5 stars

- Harnessed ANN, scikit-learn, TensorFlow, Streamlit, Python, and SHAP for model development