

Karunya Srinivasan

London, CA | karunyasrini23@gmail.com | (647) 882-0881 | linkedin.com/in/karunya-s | github.com/Karunya003

Summary

Data Engineer with a Master's degree and 2+ years of experience designing and automating data pipelines, ETL workflows, and scalable data systems. Skilled in Python, SQL, and Azure for building data ingestion, transformation, and integration processes supporting analytics and machine learning.

Skills

Languages & Libraries: Python (Pandas, NumPy, Scikit-learn, XGBoost, Matplotlib, Seaborn, Plotly), SQL

Data Engineering & ML: ETL/ELT, Data Modeling, Data Pipelines, CI/CD, Data Integration, NLP, Model Optimization

Frameworks: TensorFlow, Keras, PyTorch, HuggingFace, LangChain

Databases: MySQL, PostgreSQL, MongoDB

Tools: Docker, Microsoft Azure, Git, Jupyter, Streamlit

Education

University of Western Ontario, MSc in Computer Science

Sept 2024 – Sept 2025

- **GPA:** 4.0/4.0

- **Courses:** Intro to Machine Learning, Advanced AI, Database II, Brain Inspired AI, Parallel Computing

- **Research:** Designed and benchmarked explainable AI pipelines using PyTorch, evaluating model performance across diverse data conditions.

Anna University, BE in Computer Science

2020 – 2024

Experience

Data Research Assistant, University of Western Ontario – London, Canada

July 2025 – present

- Designed and maintained data ingestion and transformation pipelines for multilingual text datasets using Python.
- Developed automated ETL workflows to preprocess and structure unstructured text for model training.
- Implemented a RAG architecture with FAISS vector database and CI/CD integration for scalable data retrieval and updates.
- Optimized embeddings and evaluated pipeline output using BLEU to ensure data quality and consistency.

Freelance Data Analyst, E-commerce Platform – Remote

Mar 2023 – July 2024

- Built and automated data pipelines combining SQL and Python for real-time sales and customer data processing.
- Performed data wrangling, transformation, and modeling to support pricing, forecasting, and performance dashboards.
- Optimized data workflows that improved reporting efficiency and supported 12% higher conversion outcomes.

Developer Intern, National Association of Software & Service Companies – Chennai, India

Mar 2023 – May 2023

- Engineered data preprocessing and ETL scripts for text-based datasets to train NLP models in Python (Pandas, Scikit-learn).
- Developed and deployed chatbot APIs integrating structured query handling, reducing manual intervention by 40%.
- Implemented data validation and monitoring mechanisms for query logs processed through Telegram API.

Projects

Natural Language Query Agent

- Built a Streamlit application integrating LangChain SQL Agents with local LLM (Ollama) for dynamic query execution.
- Automated data ingestion, table creation, and real-time SQL generation for seamless dataset querying.

ML-Powered Taxi Fare Prediction Using Docker

- Developed a LightGBM-based regression pipeline ($R^2=0.84$) processing 50K+ records for fare prediction.
- Automated data extraction, transformation, and feature engineering using Dockerized workflows.

Depression Detection Using Twitter Posts

- Designed and trained an LSTM-CNN pipeline for text classification using 32K tweets with Word2Vec embeddings.
- Achieved 99% accuracy by optimizing data preprocessing, stratified sampling, and regularization.

News Summarizer using Web Scraping & NLP

- Automated large-scale article extraction and transformation from news sites using BeautifulSoup and Requests.
- Implemented batch text processing and optimized runtime by $6\times$ for efficient data summarization.