Ayush Sharma

ayushsharma.1074@gmail.com | [LinkedIn](#) | [Github](#) |

Education

Simon Fraser University

Master of Science in Computing Science; GPA: 3.85 *Sep. 2022 – Apr. 2024*

Vellore Institute Of Technology

B.Tech in Computing Science; GPA: 8.22 *May. 2019 – Apr. 2019*

Technical Skills

Languages: Python, Java, SQL, JavaScript, HTML/CSS, JSP
Frameworks: Flask, Apache Spark, PyTorch, Keras, Scikit-learn, Spring Boot
Technologies: Jira, Git, Databricks, Azure, Hadoop, MySQL, DB2, Cassandra, Machine Learning
Concepts: Data warehousing, Data lakes, CI/CD pipelines, ETL

Experience

YVR Airport Authority Vancouver, BC

Data Scientist *May 2023 – Dec 2023*

- Developed machine learning models for predicting different types of total flight baggage using **XGBoost** with **93%** accuracy, leveraging **Databricks**, **PySpark**, **MLflow**, **PyTorch**, **scikit-learn**, **Pandas**, and **SQL**, automating a 2-month manual process into daily operations.
- Managed full project cycle from design to deployment, implemented data pipelines in Databricks, and performed advanced **data wrangling** for effective model implementation on a dataset of around 6 million records.
- Enhanced airport logistics by automating pier/carousel assignments using a **data pipeline** with Databricks, **PySpark**, and **NumPy**.

Deloitte (Offices of the US) India

Software Developer *Aug. 2019 – Aug. 2022*

- Developed SSO-enabled web portal for Medicaid policies with **Java Spring Boot**, **SQL**, and **JavaScript**, reducing development time by 90%.
- Built automated file transfer framework with Java Spring Boot and IBM tools, managing 500-700 files daily.
- Created an automated framework using Spring Boot, MoveIT automation, **IBM DataPower**, and IBM Integration Bus to transfer 200-300 MQ messages per day on a fixed schedule between external and internal queues.
- Led a team of three in enhancing portal features, ensuring high client satisfaction and securing long-term partnership.

Projects

Neural Style Transfer *May. 2024 – Jun. 2024*

- Implemented neural style transfer to apply artistic styles to images using **PyTorch** and **OpenCV**.
- Converted it into an Python application using **streamlit** framework, to deploy and run on local or cloud servers.
- Used **L-BFGS** a second order optimization method which reduced the convergence rate to 100 epoch, as compared to Adam optimizer which took around 3000 epoch for convergence, and added **total variation** loss a a regularizer, to make the output image more smooth and increase the quality of the output image.

Large Language Models for Medical Question Answering *Feb. 2024 – Apr. 2024*

- Explored and fine-tuned **large language models** (T5, GPT-2, BLOOM) for medical question-answering, assessing accuracy using the PyTorch framework and Transformers library.
- Implemented **LoRA,** and **quantization** PEFT techniques to train models on limited hardware, utilizing metrics like **SacreBLEU** and **ROUGE** for evaluation.

Graph based Playlist Continuation using LightGCN *Jan. 2023 – Apr. 2023*

- Developed a collaborative filtering **recommendation system** using Spotify Million Playlist Dataset to recommend songs in user's playlist.
- Implemented a graph based **LightGCN** recommendation model using **PyTorch Geometric**, **DeepSNAP** and **NetworkX**, achieving a ROC AUC score of **0.84**.

Using Blockchain for Implementing Distributed Storage System *Jan. 2023 – Apr. 2023*

- Developed a blockchain-based distributed storage system using **Google Cloud Platform**, **Python**, **Solidity**, and **JavaScript**.
- Enhanced data integrity and availability by implementing file redundancy across multiple **Google Storage Buckets** and load-balanced backup assignments; integrated version control for consistency.

Vancouver Transit Delay Prediction using LSTM *Oct. 2022 – Dec. 2022*

- Developed an **LSTM** model to predict transit delays using **PySpark**, **Python**, and **Keras**, achieving a low mse loss of 0.02.
- Reduced development time by 60% for dataset creation by developing a framework for fusing real-time and static data using **Apache Spark**.