Purva Sanjay Agarwal

857-654-6076 | agarwal.pu@northeastern.edu | LinkedIn | GitHub | Portfolio | Boston, MA

Education

Northeastern University, Khoury College of Computer Sciences — Boston, MA

Master of Science in Computer Science — Sept 2023 - Dec 2025

Teaching Assistant: Theory of Computation, Database Management Systems

Relevant Coursework: Algorithms, Web Development, Programming Design Paradigm, DBMS, Cloud Computing, Artificial Intelligence, Data Mining Techniques, MLOPs

Technical Skills

Languages and Frameworks: Java, Python, JavaScript, TypeScript, C, C++, HTML/CSS, R, Powershell, Bash Scripting, Node.js, React, React Native, Express, Bootstrap, Web3.js
Database and Cloud: SQL, PL/SQL, Oracle, SQL Server, HSQL, MongoDB, NoSQL, PostgreSQL, AWS, Firebase
Testing and Tools: JUnits, Selenium, Docker, Kubernetes, Kafka, Postman, Android Studio, IntelliJ, GIT, Linux, Terraform, ELK, MLFlow, Apache Airflow

Experience

Lendbuzz, Inc — Boston, USA

Software Engineering Co-op — Jan 2025 - Aug 2025

- Upgraded distributed **Python microservices** on **AWS EC2/ECS** by migrating to **Python 3.13** and resolving deprecations, ensuring consistent cross-platform functionality across **Linux** and **macOS** environments
- Built robust **CI/CD infrastructure** using **Docker Compose**, **Makefiles**, and **GitHub Actions** to automate testing and deployment workflows, increasing development speed and team efficiency
- Developed **real-time communication** across microservices via **Kafka (Confluent Cloud)**, **REST APIs**, and **WebSockets**, supporting high-throughput financial data streaming backed by **PostgreSQL** and **Snowflake**

Fidelity Information Services — Pune, India

Senior Analyst — Nov 2020 - Aug 2023

- Increased financial data processing efficiency by 27% by automating **ETL workflows** using **Python**, **Bash**, and **PowerShell** across multiple enterprise systems
- Reduced execution time of **large-scale SQL pipelines** by 30% by optimizing **data ingestion and transformation**, enabling near real-time financial reporting
- Collaborated with infrastructure and application teams to execute **Data Center and Windows Server migrations** with **zero downtime**, improving overall system performance
- Enhanced **database performance** by 40% through advanced **query tuning** and **configuration improvements** for back-office and exchange operations
- Resolved high-priority client issues within 2 hours by implementing complex **SQL queries**, reducing operational delays and ensuring compliance with **financial regulations**
- Partnered with engineering and operations teams to deliver **scalable backend systems**, reducing **new client onboarding time** by 30% and improving integration with financial platforms

Projects

Book Recommendation System | *Python, CI/CD Pipeline, Google Cloud Platform* — Sept 2025 - Dec 2025

- Orchestrated data pipelines using Apache Airflow and DVC, data preprocessing with BigQuery, and managed model experiments and versioning with MLflow and Vertex AI, ensuring reproducibility and performance tracking.
- Deployed Streamlit frontend and FastAPI backend services on Cloud Run, managed infrastructure with Terraform, and automated builds and deployments using GitHub Actions CI/CD.
- Implemented centralized logging and monitoring with the ELK stack on Compute Engine and Google Cloud Monitoring, ensuring pipeline reliability, system observability, and model performance tracking.

Multimodal Text Processing | *Python, Whisper, BART, BiLSTM-CRF* — Sept 2024 - Dec 2024

- Built an integrated system for speech-to-text, text summarization, and text classification using PyTorch and Whisper, BART, and BiLSTM-CRF NLP models
- Fine-tuned transformer models (BART) using libraries such as Hugging Face Transformers for optimizing transcription accuracy (WER), summarization quality (ROUGE scores), and classification performance (F1 scores)
- Developed robust workflows for real-world applications, enabling automated and scalable speech-to-text pipelines