KHUSHI AGRAWAL

(623)212-5340 • kagraw22@asu.edu • linkedin.com/in/khushi-agrawal-83aa0a202/

EDUCATION

Master of Science, Computer Science **Aug 2024 - May 2026**
Arizona State University, Tempe, AZ (**CGPA - 4.0/4.0)**
Relevant Courses: Data Processing at Scale, Data Mining, Semantic Web Mining, Statistical Machine Learning

Bachelor of Technology, Computer Science and Engineering **Jul 2019 – Jul 2023**
Vellore Institute of Technology, Vellore, India (**CGPA – 3.56/4.0)**
Relevant Courses: Machine Learning, Data Warehousing, Image Processing, Artificial Intelligence, Cybersecurity

TECHNICAL SKILLS

Programming Languages: C/C++, Python, Java, SQL
Data Engineering: PySpark, Cassandra, Hadoop, Numpy, Pandas, ETL, Scikit-learn, Matplotlib, Kafka, Numpy
Databases: MySQL, Postgres, Oracle, MongoDB
Cloud Computing and DevOps: CI/CD, Git, Github, Jupyter Notebook, Postman, AWS, Azure, GCP, Bitbucket
Certifications: AWS Certified Cloud Practitioner
Tools and Frameworks: API Integration, Jira, RESTful APIs

PROFESSIONAL EXPERIENCE

DruvStar, Tempe, USA: Data Analyst Engineer Intern **May 2025 – Present**

- Building an AI-powered data classification pipeline using Python and domain-specific models, improving detection across **20K+** documents and aligning with global privacy regulations.
- Architecting containerized microservices with Docker and Kafka for secure, offline inference—achieving **3x** throughput and supporting scalable deployment like RHOAI patterns.
- Implementing MLOps practices like A/B testing, performance benchmarking, and telemetry logging to optimize model performance—cutting latency by **40%** and error rate by **25%**.
- Collaborating cross-functionally with DevOps and security teams to embed ML inference into secure, production-grade infrastructure, improving modularity and operational resilience.

Accenture, Hyderabad, India: Advanced App Engineering Analyst **Oct 2023 – Aug 2024**

- Engineered large-scale distributed data pipelines on Hadoop and PySpark to ingest and normalize **5TB**/day of unstructured data.
- Led schema modeling and Spark job optimizations using star schema design and partitioning, improving warehouse query speed by **35%**.
- Applied domain-driven principles to construct modular data validation components, reducing system-wide data inconsistency by **70%**.
- Authored Terraform IaC templates for consistent cloud resource provisioning across staging and production environments (AWS).

Ataloud Technologies, Mumbai, India: Full Stack Developer Intern **Sep 2022 – Mar 2023**

- Spearheaded **2** live projects and developed **120+** responsive screens using ReactJS, Next.js, Material UI, and Redux, enhancing functionality and user experience for **10K+** users.
- Collaborated with backend and DevOps teams to integrate **40+** REST APIs, streamlining data flow across microservices and improving UI–backend synchronization.
- Conducted unit and regression testing to resolve **77%** of critical data bugs, ensuring accuracy of backend responses and reliability of user-facing analytics.
- Contributed to SpringBoot-based backend feature development and MongoDB integration, gaining hands-on experience with Java-based microservices and NoSQL modeling.

ACADEMIC PROJECTS

Online Judge, Remote **Spring 2025 – Present**

- Developed a multi-tenant containerized execution system using +Docker and Kafka, maintaining secure isolation for **100+** concurrent code executions.
- Built infrastructure on +Heroku with CI/CD automation and real-time event streaming to support distributed, low-latency system behavior.

LLMind, Tempe, AZ **Spring 2025**

- Orchestrated Spark-based ETL workflows to clean and structure over **1M** financial tweets for input to +FinBERT and other domain-tuned LLMs.
- Implemented pipeline monitoring using Prometheus; achieved **96.5%** accuracy in sentiment prediction on MSFT data.