

YASH MITTAL

(408)-207-2705 | yash.mittalz.co.in | ym.mittal.college@gmail.com | linkedin.com/in/yashmittal | github.com/yashmittal

EDUCATION

Arizona State University, USA

Concurrent Bachelor of Computer Science and Data Science

May 2027

Dean's List Award '23, '24 and '26, President for ACM at ASU (ASU's largest student organization)

GPA 3.8

SKILLS

Language: Python, R, Java, C++, HTML, CSS, MySQL, MATLAB, Go

Machine Learning/AI: NLP, Sentiment Analysis, Regression, Scikit-learn, TensorFlow (Basic), Computer Vision, Matplotlib, RAG

Soft Skills: Communication, Collaboration, Problem-Solving, Leadership, Public Speaking, Technical Communication

Completed Courses: Data Structures & Algorithms, Digital Design Fundamentals, Computer Org/Assembly Lang Prog

Technologies & Tools: Kafka, Snowflake, Angular, Firebase, Git, Claude Code, Google Antigravity

WORK

Software Engineer Intern, Five9, California, United States

Jan 2026 - Present

- Engineered a memory-safe, high-throughput pipeline utilizing DuckDB PyRelations to process 1M+ row datasets; slashed RAM overhead by 90% (to under 250MB) while achieving processing speeds of ~7 seconds compared to legacy Pandas/Polars methods.
- Architected a file-agnostic validation engine that dynamically reads Snowflake schemas to process CSV sources without hardcoded configurations; integrated vectorized SQL sanitization and Python UDFs with LRU caching to execute real-time data validation.
- Orchestrated the workflow into a production-grade Apache Airflow DAG; eliminated scheduler bottlenecks by encapsulating I/O within task contexts, executed capacity planning for infrastructure stability, and automated file-state routing to prevent collisions.

Software Engineer Trainee, Acqueon, Texas, United States

Mar 2025 - Dec 2025

- Spearheaded the architectural redesign of a core data enrichment pipeline, integrating TransUnion data into the Acqueon Data Cloud ([Link](#)) to resolve critical data gaps & empower clients with ~129 new demographic attributes for hyper-targeted segmentation.
- Architected a decoupled data flow by separating data fetching from filtering logic using Go channels, resulting in a 76% reduction in code complexity for a core ingestion function; concurrently engineered a reusable, format-compliant CSV generation module with a unique timestamp-based naming convention to ensure data integrity and prevent file collisions in a high-concurrency environment.
- Strengthened system resilience by engineering robust error handling, including 10-second query timeouts and row-level error skipping, ensuring continuous operation during partial data source failures.

PROJECTS

Collaborative Q&A Platform - Backend & Admin Module (JavaFX, MySQL), Arizona State University

Jan 2025 - Apr 2025

- Engineered a secure, multi-role Q&A platform in JavaFX and MySQL, spanning 4.5K Java LOC, 50+ JavaFX screens, and 13 normalized tables, aligning UML (Astar) design and implementing FSM-based input validation, SHA-256 password hashing, and OTP onboarding for six user roles.
- Integrated trust-based moderation and reviewer scoring, delivering feature-rich admin controls and raising unit test coverage above 90% with 60+ JUnit tests, leveraging CI/CD for timely, high-quality delivery.

Image Classification with CNN (CIFAR-10 Dataset), ([Link](#))

Jan 2025

- Designed and implemented a CNN using TensorFlow and Keras for image classification; applied data preprocessing techniques, including normalization, to improve training efficiency across 30 epochs.
- Utilized hyperparameter tuning and model optimization to classify 10 image categories effectively, achieving 67.23% test accuracy.

IMDB Sentiment Analysis, ([Link](#))

Dec 2024

- Built an NLP pipeline to classify IMDB movie reviews; preprocessed raw text data (removing HTML tags, stopwords, and punctuation), improving model performance by 20%.
- Implemented the Bag of Words model, achieving 60%+ accuracy using a Random Forest classifier, evaluated via ROC AUC score.

SFO Crime Clustering Project, ([Link](#))

Dec 2024

- Preprocessed and cleaned large datasets for effective use in clustering algorithms, achieving high data accuracy; analyzed San Francisco crime data (2003-2014) using K-Means clustering to identify 5 major crime hotspots.
- Developed geospatial visualizations and insights to help law enforcement identify patterns, improving decision-making efficiency.

OTHER EXPERIENCES

Data Science Virtual Experience Program, BCGX

Dec 2024

- Extracted and analyzed client data, generating visual insights that highlighted key trends and potential churn drivers.
- Engineered & optimized Random Forest model, achieving 85% accuracy in predicting churn, enhancing predictive capabilities.