Mutian He

San Jose, CA
mhe3@scu.edu, 408-548-1060, www.linkedin.com/in/mutian-he, github.com/austin10231
Portfolio: www.mutianhe.com

EDUCATION

Santa Clara University, Leavey School of Business — Santa Clara, US
Master of Science in Information Systems, GPA 3.59 — Dec 2026

- Data Analytics with Python, Cloud Computing, Database Management Systems, Machine Learning, Deep Learning

University of Glasgow, School of Computing Science — Glasgow, UK
Bachelor of Science in Computer Science (BSc CS), GPA 3.66 — Jun 2024

- Data Structures & Algorithms, Database Systems, Software Engineering, Human-Computer Interaction

TECHNICAL SKILLS

Programming Languages: Python, Java, SQL (MySQL)
Libraries & Frameworks: Pandas, NumPy, Scikit-learn, CatBoost, Streamlit, Flask, Matplotlib, Seaborn Data
Engineering: ETL Pipelines, Data Cleaning, Exploratory Data Analysis (EDA)
Tools: Git, GitHub, Linux, VS Code, IntelliJ IDEA, Jupyter Notebook, Jenkins, Render, AWS(EC2, S3, RDS)

PERSONAL PROJECTS

End-to-End Instacart Reorder Prediction System – Machine Learning(2026)

- **Engineered a modular ML pipeline** (ETL, Training, Inference) driven by centralized YAML configuration for reproducibility.
- **Eliminated data leakage** by implementing strict temporal data splitting, correcting inflated metrics to a realistic 0.62 F1-score.
- **Trained a Random Forest model** using aggregated user-product features, achieving a 93% Recall on reorder events.
- **Deployed an inference script** (predict.py) to load serialized models (.joblib) and simulate real-time predictions.

Intelligent Job Description Analyzer – NLP (2026)

- **Developed an NLP parsing engine** to transform unstructured job text into structured insights (Skills, Education, Seniority).
- **Architected a modular Python system** with specialized extractors (e.g., skill_extractor) to ensure precise information retrieval and code maintainability.
- **Built and deployed a Flask REST API** on Render, creating a live full-stack application that processes both raw text and URL inputs in real-time.

Real-time Flight Delay Prediction System –ML Engineering (2026)

- **Developed an end-to-end ML prototype** using **CatBoost** to process high-cardinality categorical features (e.g., airline carriers, origin/destination), establishing a functional baseline for delay forecasting.
- **Engineered a scalable model deployment pipeline** that bypasses Git file size limits by dynamically fetching and caching serialized models from GitHub Releases upon application startup.
- **Deployed an interactive web dashboard** via **Streamlit**, enabling users to perform real-time inference through a user-friendly interface without local environment configuration.

Jenkins-as-a-Service (JaaS) OSS Platform – DevOps (2025)

- Designed the **high-level architecture** for an internal Jenkins-as-a-Service CI/CD platform.
- Created system workflows, user stories, and pipeline diagrams for build and deployment processes.
- Proposed an **RBAC security model** defining admin, maintainer, and developer access levels.
- Collaborated in an Agile environment to document platform components and present technical deliverables.

ADDITIONAL INFORMATION

Interests: Data analytics, Machine learning projects, Fitness, Cooking
Leadership Experience:

- Vice President, University of Glasgow AI Club (2023)
- Member, University of Glasgow AI Club (2022)