

Shreyash Sahare

Boulder, CO | work.shreyash1001@gmail.com | 720-255-4636 | LinkedIn | GitHub

SKILLS

Languages and Frameworks: Python, R, SQL, Linux, TensorFlow, PyTorch, Scikit-Learn, Snowflake, Tableau, Docker, Airflow

Orchestration & Cloud: MLflow, AWS (Glue, S3, EC2, Lambda, Athena), Azure (Databricks), Data Analytics

EXPERIENCE

Void Robotics

Robotics Engineering Intern

Marathon, FL (Remote)

08/2025 - Present

- Engineered ROS2-based perception modules containerized in Docker, to support **robotic automation**
- Developed a **computer vision pipeline** for pill counting in 50+ images using **K-Means clustering** and CV techniques

Mercor

Data Science Expert (Independent Contractor)

San Francisco, CA (Remote)

08/2025 - 09/2025

- Completed 20+ ML/analytics projects applying **SQL**, **Python**, and **statistical modeling** to solve real-world business challenges
- Conducted A/B testing, regression, and clustering analyses to optimize average model performance by **36%**
- Designed evaluation rubrics for LLMs, translating results into product insights and **improving model usability by 27%**

Parlay Finance

Data Scientist

Boulder, CO (US)

01/2025 - 05/2025

- Built loan classifiers using **XGBoost**, **Random Forests** and **TabTransformers** with **87% accuracy** on 200k+ **financial records**
- Reduced false approvals by **22%**, directly improving user trust and conversion strategies
- Conducted **exploratory data analysis** and identified variables that increased model interpretability by **35%** through **SHAP**

National Institute of Technology, Tiruchirappalli

Undergraduate Research Intern

Tiruchirappalli (India)

06/2022 - 08/2022

- Devised **ResCoWNet CNN** to denoise OCT scans, boosting **SSIM** by **12%** over baseline models for clearer retinal imaging
- Contributed to peer-reviewed research **cited 8+ times**, showcasing applied expertise in **medical AI and computer vision**

TECHNICAL PROJECTS

Semantic Book Recommender | [GitHub]

05/2025 - 06/2025

- Crafted a **RAG-based** semantic retrieval pipeline using **HuggingFace sentence transformers + ChromaDB** for **5K+ books**
- Designed **context orchestration logic** for emotion tagging and genre clustering, **improving personalization by 22%**
- Achieved **78% genre** (BART-MNLI) and **66% emotion** (DistilRoBERTa) accuracy, resolving cold-start issues for **1.2K+ books**

Privacy Enhancing ML for Job Interview Analysis | [GitHub]

11/2024 - 12/2024

- Classified job interview responses into four explanation levels for **38 interviewees**, analyzing linguistic patterns.
- Delivered **79% balanced accuracy** for **Random Forest** (Over-explained vs. Comprehensive) and **66.67%** for **Conv1D** (Under-explained vs. Succinct).
- Engineered **1500+ linguistic features** (TF-IDF, POS tagging, embeddings), reducing speaker bias by **30%**

TextSum | [GitHub]

09/2024 - 12/2024

- Developed an **NLP-based** text summarizer for **187K+ compliance/legal records**, cutting manual review time by **40%**
- Customized a local attention mechanism using **Bi-LSTMs** to optimize summarization quality while **lowering GPU usage by 20%**
- Showcased practical NLP deployment for **document intelligence**, demonstrating skills in **sequence modeling and text analytics**

Wine Quality Prediction with MLOps | [GitHub]

10/2023 - 11/2023

- Designed a **modular MLOps pipeline** with **user-triggered predictions** for beverage manufacturers
- Logged **30+** experiments via **MLflow UI** for **ElasticNet tuning**, reducing RMSE by **18%** and training time by **27%**
- Deployed via **Docker + Streamlit** on **AWS EC2** with GitHub actions, delivering predictions in **under 3s** per input

EDUCATION

University of Colorado Boulder

08/2023 - 05/2025

Master of Science in Data Science

GPA: 3.73

- Key Coursework:** Neural Networks and Deep Learning, Machine Learning, Big Data Analytics, Statistical Methods

National Institute of Technology, Tiruchirappalli

07/2019 - 05/2023

Bachelor of Technology in Electronics and Communication Engineering

GPA: 3.22

- Key Coursework:** Linear Algebra and Calculus, Statistical Theory of Communication, Deep Learning

CERTIFICATION

- SnowPro Core Certification, Snowflake - Aug 2025**