Chinmay Arora

Data Scientist | Data Analyst | Machine Learning Engineer | Gen AI Engineer

Minneapolis, USA 737-340-7827 chinmayarora2001@gmail.com

Education

Master of Science (M.S.) — University of Minnesota, Twin Cities, MN, USA
Data Science, 3.59 GPA — 09/2023-12/2025

Relevant Coursework: Data Mining, Machine Learning Fundamentals, Categorical Data Analysis, Multivariate Analysis, Spatial Data Science, Visualization with AI, Principles and Architecture of Database Systems, AI using Bayesian Optimization and RL for Decision Making

Graduate Research Volunteer: Data Analysis and Management Research Group, Dr. Jaideep Srivastava – work on Gen AI, RAG agentic systems and SLMs

Bachelor of Technology (B.Tech.) — SRM Institute of Science and Technology, Kattankulathur, India
Computer Science and Engineering, 8.98/10 GPA — 06/2019-05/2023

Activities: Project Head – Enactus SRM (led student social innovation projects); Product Manager – Enactus Marketplace (campus startup initiative)

Relevant Work Experience

Graduate Research Assistant (Consultant), CURA, University of Minnesota — **Minneapolis, MN, USA; 05/2025-12/2025**

- **Developed ML-driven resident segmentation and AI integration frameworks** with Hennepin County's Digital Experience and Innovation & Data Analytics teams, applying **demographic clustering, behavioral modeling, and automated ETL pipelines** to enhance digital accessibility and Power BI reporting. Designed a **responsible AI evaluation system** incorporating **fairness metrics, human-in-loop review, and bias mitigation strategies** for scalable, ethical AI deployment in county operations.

Data Science/ ML Intern, MarketMakerCRE (Real Estate Analytics) — **Remote/ Fl, USA; 07/2024-09/2024**

- Built and orchestrated end-to-end **machine learning pipelines** using Python and Azure cloud services. Streamlined data ingestion from web sources (**BeautifulSoup, Selenium**) into databases (**MongoDB, Azure SQL**), reducing data latency and manual effort by **75%**.
- Trained and validated ML models (e.g., **XGBoost, Random Forest**) for real estate property valuation; set up continuous training and evaluation loops. Implemented model **explainability (SHAP)** and performance monitoring to ensure robust deployment in production environments.

Data Intern, Confetti AI (EdTech/Careers Startup) — **Remote/ Fl, USA; 05/2024-07/2024**

- Collaborated with software engineers and DevOps to **integrate machine learning features into a live product environment**. Tuned **recommendation algorithms** and A/B tested model updates, resulting in a **40%** boost in user engagement on the platform.
- Assisted in containerizing ML services using Docker and deploying them via FastAPI endpoints, accelerating the team's model deployment cycle by 20% and establishing a basic **CI/CD workflow** for ML experiments.

IT Intern, Daikin India (Manufacturing Division) — **Gurugram, India; 06/2022-07/2022**

- Developed an **NLP-driven FAQ** chatbot (Flask backend with spaCy NLP) and deployed it on the company intranet to handle common customer inquiries. Ensured the solution utilized efficient vector search (Pinecone) for quick information retrieval.
- Created data visualization dashboards and automated data workflows to demonstrate how **machine learning and data automation** can improve decision-making in sales and support teams.

Projects/ Publications

Payment System for University (*IJSER*, Volume 11 Issue 8); Face Mask Detection Using CNNs (*IJSER,* Vol. 10 Issue 12)

Financial Balance Sheet RAG System – Engineered a **hybrid agentic RAG pipeline** for automated financial Q&A over balance sheets and filings using **LangChain, HuggingFace, ChromaDB, and Supabase**, deployed on a **private server** with containerized **FastAPI microservices**. Integrated **Text-to-SQL** for tabular reasoning and **multi-hop retrieval**. Implemented **MLOps workflows** with **Docker**, CI/CD automation, **MLflow**, and real-time evaluation via **RAGAS** and **Prometheus**, ensuring scalable, secure, and explainable retrieval performance.

AutoML Hyperparameter Tuning – Implemented Bayesian hyperparameter optimization for various models (scikit-learn, TensorFlow) to systematically improve model accuracy. Achieved up to 15% better performance on benchmark datasets and deployed the best models via a Flask API for demonstration.

Relevant Skills

Programming: Python (expert), C++/Java (familiar), Bash; proficient in data structures and algorithms.

Machine Learning: Model Development (Regression, Classification, Clustering), Feature Engineering, Deep Learning (CNNs, basic RNNs), NLP techniques, Reinforcement Learning fundamentals.

ML Frameworks: scikit-learn, TensorFlow, Keras, PyTorch; Libraries: pandas, NumPy, SciPy.

MLOps & Deployment: Docker, Kubernetes (basic), FastAPI/Flask for model serving, Git/GitHub (version control), CI/CD pipelines (GitHub Actions); Model tracking and evaluation.

Data Engineering: ETL processes, Apache Spark & Hadoop ecosystem basics, Airflow (familiar); Relational DBs (SQL) and Big Data stores.

Cloud: Azure (Functions, Data Factory, ML Studio), AWS (S3, EC2, SageMaker basics), GCP (BigQuery); experience deploying and scaling apps on cloud platforms.

Tools: Linux/Unix, Jupyter Notebooks, REST APIs, JSON; Others: Agile/Scrum experience, Technical documentation, Team collaboration tools (JIRA, Confluence).