

SUKRITI SEHGAL

Boulder, CO | 720-939-7944 | sukritisehgal2800@gmail.com
linkedin.com/in/sukriti-sehgal-b49827209 | github.com/sukritisehgal-28

EDUCATION

University of Colorado Boulder

Master of Science in Data Science (GPA: 3.8)

2024 – Present

- Coursework: Machine Learning, Statistical Modeling, Data Mining, Datacenter Scale Computing, Computer Vision

Maharaja Surajmal Institute of Technology, Delhi, India

B.Tech in Electronics and Communication Engineering (GPA: 3.9)

2019 – 2023

TECHNICAL SKILLS

Languages: Python (Pandas, NumPy, SciPy), SQL (T-SQL, SparkSQL, HiveQL), R, Dart

ML/AI: Scikit-learn, TensorFlow, PyTorch, Hugging Face, LangChain, Regression, Classification, Clustering, NLP, Deep Learning, Time Series

Data/Cloud: PostgreSQL, MySQL, MongoDB, Spark, Hadoop, AWS, Azure, GCP, Docker, FastAPI, ETL

Analytics: Statistical Modeling, A/B Testing, EDA, Feature Engineering, Data Visualization (Tableau, Power BI, Matplotlib, Seaborn), Git

PROFESSIONAL EXPERIENCE

GoodieBag Food Co.

Jun 2025 – Present

Boulder, CO

Data Science Intern

- Built a LangChain-based marketing agent to automate caption and story generation, improving content creation speed by 55%.
- Developed SQL pipelines computing 20+ KPIs across sales and marketing data, powering weekly reporting for 10 stakeholders to optimize campaigns.
- Built Tableau dashboards tracking ROI, engagement, and retention across 7+ campaigns; insights supported marketing decisions impacting 20k+ active app users.
- Integrated an NLP-powered support chatbot into the app by connecting Flutter UI with a FastAPI backend, enabling intent-aware responses and structured ticket data capture.
- Orchestrated n8n automation pipelines to route chatbot outputs through AI agents for tagging, summarization, and workflow handoff.

C5 Consultare / University of Denver

Sept 2025 – Present

Arvada, CO

Data Analytics & Application Development Intern

- Developed predictive regression models for ROI and debt-to-income forecasting using Python (Pandas, NumPy, scikit-learn), enabling program-level repayment estimation.
- Built stakeholder-facing dashboards for repayment, ROI, and sensitivity analysis across 4.99%–8.05% interest-rate scenarios.
- Engineered a multi-state analytics dataset by standardizing College Scorecard + FREOPP into a unified schema for cross-state ROI and debt-burden comparisons.
- Automated data cleaning and feature engineering pipelines for 100k+ records; reduced runtime by 20% and improved reproducibility for iterative research deliverables.

Omdena

Nov 2023 – Jun 2024

Remote - Global

Machine Learning Engineer Intern

- Improved tuberculosis risk prediction pipeline using TensorFlow by optimizing preprocessing, training, and evaluation against project benchmarks (e.g., accuracy/AUC).
- Collaborated in a global Agile team to ship an end-to-end TB risk prototype (model + Streamlit app), integrating preprocessing, training, and evaluation by deadline.

Defence Research and Development Organization (DRDO)

Oct 2022 – Jan 2023

New Delhi, India

Machine Learning Intern

- Implemented secure NLP workflows using one-time/two-time pad encryption techniques for privacy-preserving text processing.
- Built a Transformer-based summarization pipeline using Hugging Face + spaCy on 10k+ documents; achieved 0.85+ BERTScore while optimizing for high compression and entity preservation.

PROJECTS

Secure Medical AI | React, TypeScript, FastAPI, PostgreSQL, GCP, Docker, LangChain, GPT-4, FAISS, Redis

- Built a secure clinical documentation platform with role-based dashboards, note templates, and scheduling workflows using React + FastAPI.
- Implemented summarization + semantic search with LangChain (GPT-4) and FAISS embeddings for retrieval over clinical notes.
- Deployed on GCP using Cloud Run + Cloud SQL (PostgreSQL) + Memorystore (Redis) with async job orchestration via Cloud Tasks.

OmniShelf AI – Retail Shelf Intelligence | YOLOv11, FastAPI, React, TypeScript, PostgreSQL, GPT-4

- Built an end-to-end shelf intelligence system for product detection, planogram comparison, and out-of-stock alerting.
- Developed a 3-stage CV pipeline: SKU-110K (localization), YOLOv11s fine-tuned on Grozi-120 (classification), GPT-4o Vision (verification).
- Achieved mAP@50 95.51%, precision 84.89%, recall 88.52% on shelf detection and classification tasks.

Wikipedia Clickstream Analysis | Python, Pandas, Scikit-learn, NetworkX

- Processed 70GB+ clickstream data (1M+ records) to analyze navigation patterns and retention signals at scale.
- Applied clustering + association rule mining and visualized community structure using modularity-based network graphs.

LEADERSHIP

University of Colorado Boulder

2024 – Present

Boulder, CO

Chair – Data Science Student Association (DaSSA)

- Leading a student board delivering workshops, speaker events, and community programming for the MS-DS cohort.

AWARDS & ACHIEVEMENTS

CalHacks 12.0

San Francisco, CA

Oct 2025

Winner – Best ASI Personality Agent (Fetch.ai Track)

- Won for InterViewAR, a VR-based AI interview experience using Fetch.ai agent verse, Groq Whisper, FastAPI, and WebXR.