Ishmeen Garewal

(929) 996-8705 | ig2548@columbia.edu | https://www.linkedin.com/in/ishmeen-garewal/ | https://ishmeen-11.github.io/Ishmeen-Portfolio/

EDUCATION

Columbia University — New York City, NY
Master of Science in Data Science, GPA: 3.6/4 — Expected Dec 2026
Courses: Machine Learning, Statistical Inference, Causal Inference, Forecasting, Applied Deep Learning, Data Visualization

Fr. Conceicao Rodrigues Institute of Technology — Mumbai, India
Bachelor of Engineering in Computer Science, CGPA: 9.66/10
Jun 2025

TECHNICAL SKILLS

Programming: Python, SQL, R, C++, Java, JavaScript
ML & Deep Learning: Transformers, Graph Neural Networks (GCN, GAT), Convolutional Neural Networks (CNNs), NLP, Time Series Modeling
Data & Infrastructure: ETL Pipelines, Feature Engineering, Data Validation, REST APIs (FastAPI, Django), AWS (S3, EC2), BigQuery, MySQL, MongoDB, Pinecone, Chroma, Git, Docker, Unix/Linux, CI/CD

EXPERIENCE

Northeast Big Data Innovation Hub — New York City, NY
Graduate Student Assistant — Jan 2026 - Present
- Developing healthcare digital twin prototype integrating multi-source physiological datasets for patient-level treatment simulation
- Constructing structured and time-series feature pipelines to support predictive modeling experiments
- Delivered Advanced Excel workshop to 25+ MS Data Science students; currently developing an Advanced SQL workshop

Indian Institute of Technology, Roorkee — Roorkee, India
Project Intern — May 2025 - Aug 2025
- Analyzed 5,000+ vehicular trajectories using spatiotemporal modeling, validating data quality and consistency with <2% error
- Engineered behavioral and efficiency metrics (gap acceptance, yielding, speed compliance) across 10+ maneuver types, improving model classification accuracy by 18%
- Designed and evaluated ML-based decision systems using fine-tuned BERT and retrieval-augmented pipelines, achieving a 22% accuracy lift over rule-based baselines

Deloitte Touche Tohmatsu India LLP - Financial Advisory — Mumbai, India
Intern — Jun 2024 - Jul 2024
- Automated FX deal valuation pipeline using QuantLib with REST API integrations, reducing manual valuation processing time by 75%
- Built Django-based visualization dashboard with CI/CD deployment adopted by 40+ consultants across practice, enhancing reporting efficiency
- Analyzed RBI vs. NBFC climate risk policies, identifying 3 major gaps with actionable mitigation proposals
- Presented automation framework and policy recommendations to 3 Partners and 15+ senior leaders

PROJECTS

FraudGT: Financial Fraud Detection with Graph Transformers — Sep 2025 - Dec 2025
- Implemented baseline Graph Convolutional Network on 100K sampled transactions from 5M-transaction IBM AML dataset (0.1% fraud rate), applying 1048× fraud class weighting to address extreme imbalance
- Achieved 1.60% F1 (90.9% recall, 0.81% precision), highlighting precision-recall tradeoffs under severe class skew
- Benchmarked against PE-FraudGT (76.4% F1) and analyzed edge-centric gating and multigraph enhancements driving large performance gains in graph transformers

Fashion Trend Analysis & Prediction — Sep 2025 - Dec 2025
- Built multimodal AI engine to quantify fashion style signals using 150,000+ Instagram images, leveraging Google Vertex AI multimodal embeddings for real-time similarity search.
- Engineered automated embedding pipeline with PyTorch autoencoder and vector storage (Chroma), enabling scalable retrieval and ranking of similar items for design validation.
- Designed and validated *Enhanced Popularity Score* integrating log-scaled engagement metrics and sentiment weighting to surface market-validated trend recommendations.
- Deployed high-performance pipeline querying BigQuery partitions with normalized similarity thresholds, achieving sub-second trend retrieval across partitioned BigQuery datasets.

Heat's Toll on Health — Nov 2025 - Dec 2025
- Integrated 4 CDC datasets (87K+ records, 1,806 counties, 50 states, 2000–2022), resolving schema mismatches and geographic resolution gaps to build a unified analytical dataset queried in SQL and analyzed in R
- Engineered a clean complete-case schema by handling 52% suppressed mortality records, splitting composite variables, and standardizing across datasets with inconsistent state reporting
- Applied z-score normalization and correlation analysis to distinguish data suppression artifacts from true signal, revealing hospitalizations grew 2x faster than heat exposure (32% vs. 16%)

PUBLICATIONS

- **Garewal, I.K.**, Mahamuni, C.V., & Jha, S. "Emerging Applications and Challenges in Quantum Computing: A Literature Survey." *2024 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, Port Louis, Mauritius.
- **Garewal, I.K.**, Jha, S., & Mahamuni, C.V. "Topic Modeling for Identifying Emerging Trends on Instagram Using Latent Dirichlet Allocation and Non-Negative Matrix Factorization." *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*
- Jha, S., **Garewal, I.K.**, Aathisaya, L., Alphonso, L., & Aher, B. "FoodMO: A Food Nutrient Analysis Application Using Optical Character Recognition and Machine Learning." *Lecture Notes in Networks and Systems, vol. 1159*, Springer, Singapore, 2025.