Siddharth Nishtala

9971011460 siddharth.nishtala@gmail.com
siddharthnishtala.github.io linkedin.com/siddharth-nishtala github.com/siddharthnishtala

Education

Indian Institute of Technology Madras 2020 - 2024
M.S. By Research, Computer Science & Engineering, CGPA: 9.14 *Chennai*

Shiv Nadar University 2015 - 2019
B.Tech, Computer Science & Engineering, CGPA: 8.17 *Delhi NCR*

Experience

Accertify *(formerly a subsidiary of American Express)* Mar'24 – Present
Senior Data Scientist / Data Scientist *Gurgaon*

- Led Accertify's graph-based fraud detection initiative end-to-end — built transaction communities using GraphFrames in PySpark with a rolling-window approach, engineered graph-derived features, and partnered with ML engineering to validate production communities against offline benchmarks. Delivered 2–14% detection improvements across clients.
- Contributed to fraud prediction models for the amusement industry, improving detection by 8–22% through feature engineering and model iteration.
- Built an internal tool to automate feature transformation pipeline implementations, cutting client onboarding time from 12 hours to 2 hours (6x faster) and enabling the team to scale high-profile merchant integrations.

Robert Bosch Centre for Data Science and Artificial Intelligence Jun'19 - Aug'23
MS Research Scholar / Project Associate *Chennai*

- Built deep learning models (CNN- and LSTM-based) on 70M+ anonymised call records from 300,000 women enrolled in ARMMAN's mMitra maternal health program to predict short- and long-term dropout risk, achieving 13% and 7% improvements over competitive baselines respectively.
- With Google and ARMMAN, we ran a randomised controlled trial across four intervention arms (control, SMS, call, hybrid), demonstrating that targeted phone interventions boosted engagement by 61% relative to no intervention.
- Formulated intervention resource allocation as a Restless Multi-Armed Bandit (RMAB) problem, using Whittle Index policies to prioritise which at-risk women were most likely to respond to a health worker call.

Key Projects

Cost-Sensitive Trees for Interpretable Reinforcement Learning | *MS Thesis, Published at CODS-COMAD 2024*

- Identified a blind spot in tree-based imitation learning: standard methods treat all misclassifications equally, ignoring the fact that some wrong actions are far more costly than others in RL. Proposed using the teacher policy's Q-function to assign per-action, per-state misclassification costs during tree construction.
- Developed CS-VIPER and CS-MoET — cost-sensitive variants of two established interpretable RL methods — by replacing standard decision trees with Example-Dependent Cost-Sensitive Decision Trees, producing policies that make cheaper mistakes even when they make more of them.
- Demonstrated consistent outperformance across four environments (LunarLander, Taxi, FourRooms, highway-fast), with gains ranging from 3% to over 150% at various tree depths.

Jharokha — Self-Hosted Photo Library | *Personal Project*

- Built a full-stack, self-hosted photo management application to replace cloud subscriptions, with a FastAPI + SQLAlchemy backend and a Svelte 5 / Tailwind frontend. Uses SHA256 content hashing for duplicate detection and a "no-move" indexing philosophy — the library references photos in place rather than copying them.
- Integrated CLIP (via HuggingFace) and ChromaDB for semantic search over 5,000+ photos, enabling natural-language queries like "scuba diving" or "cricket match under stadium lights" alongside EXIF metadata, reverse geocoding, and Live Photo pairing for HEIC/MOV files.

Technical Skills

Skills: machine learning, deep learning, reinforcement learning, data visualization and analysis
Programming Languages: Python, SQL, Julia, Matlab, Java
Libraries: Pandas, NumPy, Matplotlib, PySpark, Sci-kit Learn, TensorFlow, PyTorch