

Aman Singh Thakur

Aspiring Data Scientist | Amherst, US | +1 (413) 472-6405 | amansinghtha@umass.edu | linkedin.com/in/singh96aman

EDUCATION

University of Massachusetts Amherst Expected May 2024
Masters in Computer Science with a Concentration in Data Science - (CGPA: 3.8/4)

Manipal Institute of Technology, Udupi, India Jun 2019
Bachelor of Technology in Computer Science & Engineering - (CGPA: 8.41/10)

RELEVANT COURSEWORK

- **Graduate** Machine Learning, Responsible AI, Systems for Data Science
- **Undergraduate** Natural Language Processing, Social Network Analysis, Data Structures, Algorithms

TECHNICAL SKILLS

- **Data Science** TensorFlow, PyTorch, Pandas, Scikit-learn, Matplotlib, SQL, Excel, Tableau, Keras
- **Software Engineering** Python, Java, C/C++, JavaScript, Spring, Flask, Angular, React, Git, Bash, Android

KEY PROJECTS

Compression of Deep Neural Networks Aug 2017
University of Massachusetts Amherst, Prof. Hui Guan

- Evaluated compression techniques like Network Pruning, Quantization and Knowledge Distillation to compress ResNet-18 model using PyTorch framework.
- Outperformed the basic flavors of knowledge distillation using hyper-parameter tuning on a network **12% the size of raw model**.

Credit Scoring Model for Loan Applicants Aug 2021
Goldman Sachs, Research and Development Lab

- Experimented with **Regression** and **Classification** models that classifies credit applicants into good/bad credit classes based on applicant's probability of reimbursing their financial obligations or defaulting on them.
- Focused on Dimensionality reduction and Achieved **80% accuracy (Top 5 in batch)** with simple Logistic Regression by removing redundant features using Recursive Feature Elimination technique.

PROFESSIONAL EXPERIENCE

Goldman Sachs, Bengaluru, India Oct 2020 — Jul 2022
Associate Software Engineer, Liquidity Risk Applications

- Led a team of developers to uplift the US Federal Liquidity Report (FR 2025a) and Analytics system by building efficient data pipelines and calculation models to robustly handle more than **20 billion data points** daily
- Consistently recognized as an amiable and **diligent** leader capable of coordinating with **cross-functional teams** in a fast-paced setting.

Morgan Stanley, Mumbai, India Jul 2019 — Sep 2020
Software Consultant, Prime Brokerage Technology

- Built a service to tokenize incoming trade details (keywords - buy/sell/ticker/etc) present in email bodies using **NLTK** and **OpenNLP**.
- Integrated the tokenizer service with a Spring backend and Angular UI to allow the business to book these trades autonomously.

Google Summer of Code May 2019 — Aug 2019
Student developer, CERN-HSF

- Constructed an analytical tool that allows scientists at AWAKE (CERN) to index, query and visualize HDF image files effectively.
- Successfully compressed **12TB** of HDF files to 150GB metadata csvs using **Spark and Map-Reduce Programming**. Wrote python routines using **pandas, scit-learn and matplotlib** to visualize uncorrupted data **8x faster** by searching efficiently in the metadata index.

Indian Institute of Technology, Kharagpur, India May 2018 — Jul 2018
Research Intern, Visual Information Processing Lab

- Learnt the decision-making process for the diagnosis of Acute Lymphoblastic Leukemia (ALL) to automate medical prescriptions.
- Researched best machine learning white box models to accurately capture and visualize decisions taken by doctors. Implemented various **Decision Trees** and improved precision for the CART model to **91% accuracy**, by splitting tree nodes using Gini Impurity index.

HONORS & PUBLICATION

- Authored & published article on **Analyzing Race/Gender/Job Diversity in the US** for the Nerd for Tech Medium Publication.
- Published open source library **pyawake** and article **GSoC 2019|CERN-HSF|PYAWAKE** as part of the GSoC programme.
- Secured **94% percentile** in All India Graduate Aptitude Test of Engineering for Computer Science in 2019.