Abhishek Sharma

abhishariitk@gmail.com | 585-230-2000 | 513 Richardson Road, Rochester, NY 14623

EDUCATION

University of Rochester, College of Arts Sciences and Engineering — Rochester, NY
Masters in Data Science (MS) — December 2024

- Cumulative GPA: 3.98/4.0
- Relevant Coursework: Machine Learning, Statistics, Big Data Analytics, Probability & Stochastic Processes

Indian Institute of Technology Kanpur — Kanpur, India
Bachelors of Technology in Chemical Engineering (B.Tech.) — May 2021

- Recipient of the Merit Based Scholarship awarding 80% tuition waiver — 2018, 2019, 2020

PROFESSIONAL EXPERIENCE

MacroXStudio, Inc. — San Francisco, CA
Data Scientist — September 2024 - December 2024

- Utilized python to develop a data analysis chatbot using ChatGPT and fine-tuned LLaMA 3.3, streamlining data analysis for finance and business professionals. Achieved a 3-second response time with 90% accuracy, enabling faster decisions and reducing manual effort.
- Designed a modular framework and evaluated 4 pipelines using various agents and LLMs for column identification and code generation. Integrated Google Cloud API for data fetching, executed code locally, and built a Streamlit UI for seamless user interaction
- Integrated the best-performing pipeline (fine-tuned LLaMA 3.3 for column identification and GPT-4o for code generation) into Macroxstudio's offerings per the CEO's recommendation, projected to save 200+ person-hours/month for the data analysis team.

University of Rochester — Rochester, NY
Research Assistant, Department of Chemical Engineering — October 2023 - January 2024

- Utilized python and R on high performance computing Linux cluster to develop in-house datasets, bioinformatics tools and pipelines
- Current Project: Developing synthetic nucleotide sequences for bacterial plasmids using Diffusion Models with Transformers (DiT)
- Publication: "Plasmid prevalence is independent of antibiotic resistance in Environmental Enterobacteriaceae." D. Gewurz, **A. Sharma**, J. Harrison, I. Lee, JJ Miranda, B. Meillleux, K. Hamilton, and A.J. Lopatkin (2024). Journal: Plasmid *(Under Review)*
- Publication: "PlasAnn: A curated, highly accurate plasmid annotation program reveals novel associations of plasmid-encoded functions" H. Islam, **A. Sharma**, M. Ahmed, J. Blair, A.J. Lopatkin (2024). Journal: ISME *(Under Review)*

Vedanta Limited — New Delhi, India
Data Consultant, Vedanta Aluminum Business — July 2021 - June 2023

- Tasked with improving plant operations, reducing costs on raw material sourcing, and increasing domestic market share for aluminum products using data analysis and machine learning.
- Optimized the coal procurement strategy for Vedanta's aluminum facility by conducting comprehensive market assessments, identifying cost-effective sources, and leveraging time series forecasting and Power BI, resulting in over $18 Million in savings for FY22-23 and receiving the CEO appreciation award.
- Developed a price forecasting model for the Indian Energy Exchange, employing neural networks and SARIMAX, achieving an R^2 of 0.6 over a period of 15 days, enabling optimized power plant operations during coal shortages, improving boiler efficiency by 5%, with 4.5% reduction in raw material consumption, and reducing carbon emissions.

KPIT Technologies Ltd. — Pune, India
Data Science Intern, Automotive Diagnostics — May 2019 - July 2019

- Developed a machine learning model using OBD-II vehicle data to predict remaining engine oil life, evaluating Linear Regression, XGBoost, and Random Forest, with XGBoost performing best ($R^2 = 0.9$, lowest MSE)

PROJECTS AND LEADERSHIP

Computational Modeling of Human Speech Perception: Exploring Vowel Sound Invariance — Rochester, NY
Independent Research Project - Brain and Cognitive Sciences, University of Rochester — August 2024 - January 2024

- Built a computational model using Naive Bayes to simulate how the human brain perceives vowel sounds and remains invariant to frequency variations, integrating insights from exemplar and normalization theories.
- Publication: "Exemplar models of speech perception do not prevent the need for normalization." **A. Sharma**, TF Jaeger, A Person (2025) Conference paper submission for CogSci 2025 . *Under Review*

Cofounder & Technology Lead — New Delhi, India
PrakrAIti (Vedanta Corporate Social Services backed) — April 2022 - December 2022

- Collaborated with the Odisha state government to develop a mobile application for farmers to detect & provide support for crop diseases
- Curated a dataset of 80k images for five major crops grown in the state with the help of local volunteers and annotated them with the assistance of agricultural scientists. Developed a U-Net-based model for crop disease detection, achieving 92% recall, 90% precision, and 92% accuracy. Deployed the model via a Flutter-based app, impacting over 300k farmers in the region

Real-time Tweets Sentiment Analysis — Rochester, NY
Course Project - Big Data Analytics — January 2024 - April 2024

- Developed a scalable end to end ETL pipeline using Apache Spark to process 200K tweets from Amazon S3 into Delta tables, implementing continuous streaming services for real-time sentiment classification and data processing.
- Deployed a pre-trained Hugging Face transformer model via MLflow on Datacricks, achieving 59% recall, 74% precision, 55% accuracy for real-time sentiment prediction, ensuring scalable and efficient model deployment
- Conducted sentiment aggregation and continuous data analysis, providing high-level insights through real-time visualizations, enabling robust tracking and incorporation of new data into final analysis.

SKILLS

- **Programming Languages:** Python, R, C, C++, HTML, CSS, UNIX
- **Big Data & Machine learning:** Spark, DataBricks, AWS S3, Python (e.g.: scikit-learn, NumPy, pandas, Matplotlib)
- **Data Science & Miscellaneous Technologies:** A/B Testing, ETL, Data Science Pipeline (cleansing, wrangling, visualization, modeling, interpretation, feature engineering), Statistics, Time Series, Experimental design, Hypothesis testing, OOP, OOD, APIs, Excel, Git, Deep learning architectures (GANs, Diffusion Models, Transformers), Natural Language Processing (NLP), Markov Chains, Web scraping and automation (Selenium), Python package development and publication,