Shreya Jain

Ann Arbor, MI, USA
+1 (734) 450 9984 - shreyadj@umich.edu - linkedin.com/in/shreyadjain/ - https://github.com/JainShreya26

EDUCATION

University of Michigan
Ann Arbor, MI, USA

Master's in Data Science, Statistics *Aug 2023 - May 2025*

Courses: Probability, Statistical Inference, Regression, Machine Learning, Causal Inference, Data Analysis & Visualization, Time Series
Achievements: TA (Data Analysis, Data Visualization, Tableau), DataThon (Ross) 2024 and 2025, TAMU Health Hackathon

SKILLS

Programming Languages: Python, R, C++, Java, SQL, JavaScript, HTML, CSS
Machine Learning Frameworks and Libraries: PyTorch, TensorFlow, Scikit-learn, Keras, NumPy, Pandas, Seaborn
Data Analysis and Visualization: Power BI, Tableau, Matplotlib, Plotly, Altair, Gephi
DevOps and Collaboration Tools: Git, Jira

WORK EXPERIENCE

Data Scientist
Michigan Medicine, Ann Arbor, MI, USA *April 2025 - Present*

- Implementing the DECI model (Microsoft Causica) to generate and analyze causal graphs and compare produced adjacency matrices with those generated using the DoWhy library to evaluate model accuracy and consistency
- Assessing the reliability of synthetic data by analyzing and comparing adjacency matrices of original and synthetic datasets through quantitative techniques such as network centrality and graph similarity

Data Analyst
UM Precision Health, Ann Arbor, MI, USA *June 2024 - Dec 2024*

- Increased survey participation rates by 50% by curating surveys and developing SQL queries and Python-based analytics to process and extract insights from response data, identifying key behavioral patterns and trends
- Developed real-time Power BI dashboards for survey monitoring
- Implemented reinforcement learning-based interactive games to assess patient attention span, collecting structured behavioral data for deeper analysis

Research Assistant
School of Information, Umich, Ann Arbor , MI, USA *June 2024 - Nov 2024*

- Achieved 98% word retrieval accuracy by developing a web scraping model for structured data extraction from American Historical textbooks
- Increased classification accuracy by 10% by expanding spaCy NER annotations with 7 new historical categories
- Cut manual validation time by 2 hours per dataset by creating an algorithm for data annotation verification
- Pre-processed and trained LLMs model on OntoNotes 5 and a custom data, fine-tuning GPT-NER and Meta-Llama-3-8B-Instruct using LoRA and 4-bit quantization, achieving 66% NER accuracy, improving historical NER

Graduate Student Instructor: SI649
University of Michigan, Ann Arbor , MI, USA *June 2024 - Dec 2024*

- Instructed students in advanced data visualization, covering Tableau, Python libraries (Altair, Plotly), network analysis, D3.js, and cutting-edge generative AI tools for interactive and automated data visualizations.
- Designed and led a hands-on design lab featuring real-world case studies, where students tackled diverse tasks in data analysis, color theory, and visualization best practices.

Data Science Intern
Ministry of Defence, Government of India , Pune, MH, India *Oct 2022 - May 2023*

- Enabled real-time skeletal tracking for army drills by developing an AI-based pose estimation model using OpenCV, Mediapipe, and YOLOv8, tailored for military movement analysis
- Deployed a real-time feedback system generating automated pose correction reports and annotated video snippets, assisting trainees with immediate visual guidance. Deployed the algorithm using Flask
- Designed and Created UI/UX using Figma and user-friendly web application using Flask. Paper submitted to IEEE

Data Scientist
Beyond Business Travel, TUS, Ireland *Aug 2022 - Dec 2022*

- Built a data extraction pipeline for Enterprise and Hertz invoices using OCR, NLP, and machine learning, achieving 98% accuracy and reduced manual entry time
- Outperformed RPA tools such as BluePrism and UIpath with improved anomaly detection and 75% faster processing

PROJECTS

VAE Variants for Causal Effect Estimation

Causal Inference - Machine Learning, Pytorch, Tensorflow *Nov 2024 - Dec 2024*

- Extended the CEVAE model to improve causal inference under unmeasured confounders using PyTorch and TensorFlow
- Implemented VAE variations, including Correlated-V AE, Beta-VAE, HVAE, and VQ-VAE, for enhanced latent variable modeling
- Conducted experiments on real-world (IHDP, JOBS, TWINS) and synthetic datasets using Pyro and Adamax optimizer achieving the best PEHE of 1.47 ± 0.18 and ATE error of 1.26 ± 0.75 on the JOBS dataset

Soccer Analysis (Try it!)

Tableau Dashboard, Data Analysis, Data Visualization, Statistics *Apr 2024 - Apr 2024*

- Analyzed and processed over 3 million soccer records from 11+ datasets using SQL and Python, performing data cleaning, feature engineering, and statistical evaluation
- Developed an interactive Tableau dashboard to visualize key soccer analytics

Forecasting Tuberculosis Cases with ARIMA and POMP

Time Series, ARIMA Modeling, POMP Modeling (SEIRS framework), R Programming, Data Visualization *Nov 2023 - Dec 2023*

- Improved TB incidence forecasting accuracy by comparing ARIMA(0,1,5) and SEIRS-based POMP models
- Implemented a stochastic POMP model with overdispersion and time-varying transmission, achieving a more realistic fit to TB decline trends over time

Offer and Deal Quality Prediction - Shark Tank India

Python, R, Machine Learning, Artificial Neural Networks *Oct 2022 - Dec 2022*

- Identified key trends in startup features (industry, revenue, valuation) by performing exploratory data analysis using Python (Pandas, Matplotlib, Seaborn) and R
- Achieved an F1-score of 87.09% by building a machine learning pipeline with ANN, outperforming SVM and Random Forest in predicting startup funding offers
- Improved post-investment evaluation by implementing a fuzzy logic system in MATLAB with 22 rule sets to categorize deal quality

PUBLICATIONS

- Shreya Jain, Atharva Parikh, et al. **Offer and Deal-Quality Prediction using Machine learning and Fuzzy approach: A Shark Tank India Case Study** *Proceedings of the ACM Web Conference 2023*. doi:10.1145/3590837.3590891
- Shreya Jain *et al.* **Empowering India's Climate Action: Harnessing Blockchain for Carbon Trading.** *2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)*. doi:10.1109/XYZ.2024.10837382