Amaan Mansuri

New York, NY +1 (562) 209-6787 amaan.mansuri@nyu.edu amaanmansuri Amaan165

Education

New York University – Center for Data Science New York, USA
Master of Science in Data Science *Expected May 2026*

Nirma University – Institute of Technology Ahmedabad, India
Bachelor of Technology in Computer Science and Engineering *Sept 2020 - Jun 2024*

Experience

Research Assistant New York, USA
Hartley Lab, Department of Psychology, New York University *Mar 2025 – Present*

- Developed a web scraping pipeline to collect and process over **15,000** metadata entries from books and media sources, enhancing stimuli selection efficiency by **60%**.
- Proposed and implemented **knowledge graph-based** models to simulate semantic relationships across age groups, boosting prediction accuracy in reasoning tasks by **25%**.
- Created visualizations and result summaries for IRB reports and manuscript drafts, improving clarity and comprehension for non-technical reviewers by **30%** (based on internal lab evaluation).

Research and Development Intern Kadi, India
Johnson Controls-Hitachi *Jan 2024 - Jun 2024*

- Engineered a Bluetooth-based communication system for Hitachi ACs using UART, boosting NLOS stability by **33%** and expanding range from **7m to 55m**.
- Built an Android app replicating IR remote controls, improving device response time by **40%** through optimized signal handling.

Research Assistant Ahmedabad, India
Centre of Excellence in Data Science, Nirma University *May 2023 - Jul 2023*

- Established a system to segment blister packets for QR code printing, increasing information retention to **100%** by preserving details like expiration date and batch number.
- Executed a segmentation solution using **Mask RCNN** and **VIA**, resulting in a **50% decrease** in manual inspection time, utilizing hyperparameter tuning; the new system streamlined operations for the quality assurance team.

Projects

flight-wx: Flight Delay & Weather Monitoring System (Big Data + API Integration) *Github*

- Developed a real-time flight delay tracking system integrating FAA and NOAA APIs, supporting flexible city-to-airport queries and resolving metadata for **400+ U.S. airports** with high accuracy and reliability.
- Engineered modular CLI tools with batch-mode ingestion, intelligent caching, and multi-airport resolution logic, reducing redundant API hits by **70%** and improving end-to-end ingestion speed by **3x**.

Multi-Horizon Stock Price Prediction (LSTM + FinBERT)) *Github Report*

- Built a deep learning pipeline using **LSTM networks** to forecast 1-day and 3-day prices for **26 major stocks**, integrating **FinBERT-based sentiment scores** to boost accuracy by up to **18.57%** and directional accuracy by **12%**.
- Evaluated four feature sets (price, technicals, sentiment, combined) using **RMSE**, **MAE**, **MAPE**, and $\mathbf{R^2}$; achieved best-case $\mathbf{R^2 = 0.969}$ (AAPL, 1-day); optimized training with `ReduceLROnPlateau`, `EarlyStopping`, and `TensorBoard`.

Simulated CCAR Stress Test (Machine Learning + API Integration) *Github*

- Achieved a CCAR-style stress testing pipeline to simulate **macroeconomic shock impact** (unemployment, inflation, GDP decline) on **retail loan defaults**, modeling forward-looking risk exposure.
- Trained a borrower-level **Probability of Default (PD)** model and computed **Expected Loss = PD × EAD × LGD)**, estimating capital reserve needs under adverse conditions.

Humor Detection (Machine Learning + NLP) *Github*

- Designed an NLP-based humor classification pipeline using **BoW, TF-IDF, and Word2Vec**, trained on a labeled corpus of **200k texts**, achieving **88%** accuracy and cutting misclassification by **20%**.
- Extended the pipeline by fine-tuning an LLM on the **Humicroedit** dataset to generate humorous versions of non-humorous text; incorporated automatic un-funny word detection and transformation using generative decoding.

Skills

Languages: Python, SQL, Java, C++, C, Bash, HTML, CSS, TypeScript, Solidity.

Frameworks & Libraries: NumPy, Pandas, scikit-learn, Matplotlib, XGBoost, TensorFlow, Keras, PyTorch, Transformers, NLTK, spaCy, LangChain, Hugging Face, Django, Flask, Hadoop, Spark, Dask, Git, JAX, OpenCV.

Tools & Platforms: Tableau, Power BI, Docker, TensorBoard, Airflow, AWS, GCP, Unix/Linux, Anaconda, FastAPI, Selenium.

Coursework

Machine Learning, Deep Learning, Big Data, Programming for Scientific Computing, Data Mining, Linear Algebra, Probability and Statistics, Database Management Systems, Computer Architecture, Operating Systems.