Virti Jain

New York | virtirjain09@gmail.com | (934) 246 – 7273 | https://www.linkedin.com/in/virti-jain-550020265/

EDUCATION

Stony Brook University   **Stony Brook, New York**
Master of Science in Data Science | GPA: 3.89/4   *Expected Graduation: Dec 2025*
Relevant Coursework: Data Analysis, NLP, ML in Quant Finance, Statistical Learning, Case Study in ML and Finance

Rajiv Gandhi Proudyogiki Vishwavidyalaya   **Indore, India**
Bachelor of Technology in Computer Science Engineering (Data-Science) | GPA: 8.8/10   *Dec 2020 – Jun 2024*
Relevant Coursework: ML, DL, Probability, Statistics, Linear Algebra, DBMS, Data Engineering, Big Data Analytics

PROFESSIONAL EXPERIENCE

Stony Brook University   **Stony Brook, New York**
Research Project Assistant for Dr. Sharon Nachman's Lab   *Feb 2025- Present*

- **Beyfortus Project**: Analyzed 52 patient to assess vaccine uptake trends and ICU admission factors, identifying disparities in vaccination. Evaluated the impact of hospital admission rates, and birth location on vaccine administration. Applied **Python** for data processing, statistical testing, and visualization.
- **Lyme Disease Project**: Investigated 127 patient cases to examine how different factors affected treatment success. Discovered that certain blood markers were linked to needing extra treatment, while NSAID use and joint inflammation had little impact. Utilized **Python** for statistical analysis, hypothesis testing and data visualization.

Stony Brook University   **Stony Brook, New York**
Research Assistant for Dr. Shennan A Weiss's Epilepsy Lab   *Jan 2025- Present*

- Assessed machine learning algorithms in MATLAB to automate the detection and classification of epileptiform events in large EEG datasets.
- Engineered data pipelines using MongoDB and MATLAB to manage and query large-scale EEG datasets.

Algoventor Solutions Pvt. Ltd.   **Indore, India**
Data Science Intern (Energy Consumption Forecasting for Smart Grids)   *Jun 2024 – Aug 2024*

- Designed forecasting model using **ARIMA** and **ML**, improving grid stability prediction accuracy by **15%.**
- Led exploratory data analysis on 15+ key features to identify factors affecting grid performance.
- Implemented data preprocessing to handle 100+ missing values and remove outliers for accurate forecasting.

ACADEMIC RESEARCH & PROJECTS

Real Time Financial Forecasting and Dashboard

- Authored data pipeline using Python, achieving 40% faster processing speed with real-time data integration.
- Proposed a hybrid ML/DL model with 92% accuracy for short-term financial predictions.
- Introduced interactive Streamlit dashboards, improving data accessibility and enhancing decision-making.

Advanced Multilingual Document Summarization

- Developed an NLP pipeline using Streamlit, BART & PEGASUS, automating multilingual text summarization and translation for faster document processing.
- Conducted a workflow that cut document processing time by 40% while enhancing readability by 30% through optimized summarization techniques.
- Planned a text summarization model with key phrase extraction to improve text quality.

SKILLS

Programming: Python (NumPy, Pandas, Matplotlib, TensorFlow, Seaborn, PyTorch), SQL, Java, R
Data Frameworks: Regression, SVM, kNN, K-Means, RF, Statistical analysis, time-series forecasting, CNN, RNN
Data Tools: Power BI, MATLAB, Tableau, Excel, Looker, MySQL, MongoDB, PostgreSQL, Oracle, Streamlit, Google Cloud Platform (GCP), Azure, Hadoop, Snowflake