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*
Rajiv Gandhi Proudyogiki Vishwavidyalaya   **Indore, India**
Bachelor of Technology in Computer Science Engineering (Data-Science) | GPA: 8.8/10   *Dec 2020 – Jun 2024*

PROFESSIONAL EXPERIENCE

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

- **Beyfortus (Nirsevimab) Project**: Analyzed data from **52 patients** to understand patterns in vaccine uptake and hospital admissions. Identified key factors affecting ICU admissions, vaccine access based on birth location, and outpatient care influence. Used **Python (Pandas, SciPy, Matplotlib)** for data analysis and visualization.
- **Lyme Disease Project**: Studied **127 patient cases** to examine how different factors affected treatment success. Found that certain blood markers were linked to needing extra treatment, while NSAID use and joint inflammation had little impact. Used **Python** for statistical analysis and data visualization.

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

- Developed **machine learning algorithms** in MATLAB to automate the detection and classification of epileptiform events in large EEG datasets, contributing to the refinement of epileptogenic zone localization techniques.
- Utilized **MongoDB** and **MATLAB** to design efficient data pipelines for managing and querying large-scale intracranial EEG datasets, enabling high-throughput analysis of epileptiform spike onset and fast ripple activity.

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

- Created an **energy forecasting model** using **ARIMA and machine learning**, improving smart grid stability prediction accuracy by **15%** through **exploratory data analysis (EDA) on 15+** key features.
- Executed an efficient **data preprocessing** strategy addressing over 1000 missing values while removing outliers, ensuring robust analytical outcomes critical for precise forecasting of smart grid performance metrics.

ACADEMIC RESEARCH & PROJECTS

Market Basket Analysis for Grocery Products

- Conducted market basket analysis using **Python (pandas, NumPy, mlxtend)** to identify key product associations in a dataset, uncovering 15% more cross-selling opportunities and informing marketing strategies.
- Implemented the **Apriori algorithm** to identify frequently purchased items, optimizing support to uncover 30% more relevant customer purchasing patterns, resulting in a 20% increase in insights for management decisions.

Advanced Multilingual Document Summarization

- Architected a Streamlit based processing application, integrating translator for **multilingual translation**, and **transformers (BART & PEGASUS)** for text summarization, reducing document processing time by 40%.
- Designed a comprehensive text summarization pipeline, leveraging models to generate abstractive summaries, incorporating key phrase, and improving readability by 30% through length parameter adjustments.

TECHNICAL SKILLS

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