

# VIRTI JAIN

New York, USA ◇ virtirjain09@gmail.com ◇ +1 (934)-246-7273 ◇ LinkedIn

## EDUCATION

### State University of New York at Stony Brook

New York, USA

Master of Science in Data Science (*with Advanced Certificate in Quantitative Finance*) | GPA: 3.89 / 4.0 08/2024 – Present  
Coursework: Machine Learning in Quantitative Finance, Statistical Learning, NLP, Data Analysis, Case Studies in ML and Finance

### Rajiv Gandhi Proudlyogiki Vishwavidyalaya

Bhopal, India

Bachelor of Technology in Computer Science and Engineering (Data Science) | GPA: 8.8 / 10.0 12/2020 – 06/2024  
Coursework: Machine Learning, Probability, Statistics, Database Management System, Data Engineering, Big Data Analytics

## PROFESSIONAL EXPERIENCE

### Stony Brook University

Stony Brook, NY

#### Data and Operations Intern

05/2025 – Present

- Led client-facing data visualization initiatives by creating interactive **Excel** dashboards for **20+** academic programs, directly informing strategic resource decisions across departments.
- Designed and deployed structured **Qualtrics** surveys with branching logic and sentiment scaling, boosting response rates by **35%** and generating insights that reshaped course offerings for the next academic term.
- Standardized workflows in **CampusCE** and authored onboarding SOPs; cut data entry errors by **60%** and enabled a **15% faster** registration cycle through streamlined collaboration and process clarity.

### Algoventor Solutions Pvt. Ltd.

Indore, India

#### Data Science Intern

06/2024 – 08/2024

- Developed a predictive analytics solution for smart grid load-balancing using a **hybrid Random Forest model** with **time-series features**, improving forecast accuracy by over **90%** compared to a seasonal baseline.
- Identified key consumption drivers, such as circadian and seasonal cycles, via exploratory data analysis (**EDA**) on **over 10 engineered features**, aligning the modeling strategy with real-world energy usage patterns.
- Designed and implemented a robust ETL pipeline to preprocess over 120,000 hourly records, which included correcting out-of-order timestamps (a **non-monotonic index**) to ensure chronological integrity for time-series modeling.
- Delivered interactive visual reports using **Matplotlib** and **Seaborn**, enabling both technical teams and business stakeholders to interpret trends and anomalies, driving informed energy planning decisions.

## TECHNICAL PROJECTS

### LLM-Powered Summarization Tool for Financial Documents

- Led the development of a summarization tool for financial and legal documents by fine-tuning a pre-trained **BART** model on annual filings, cutting document review time by **40%**.
- Designed a scalable summarization pipeline using input chunking, reducing average summary length by **35%** while preserving key insights. Evaluated 500+ summaries using **ROUGE** and manual QA, achieving a **25%** gain in recall and fluency over baseline, improving model trustworthiness.
- Deployed an interactive **Streamlit** UI enabling financial analysts to tune model parameters and preview summaries in real time, validated via iterative user testing and feedback loops.

### User Behavior Analytics Platform: Funnel Drop-Off, Retention, and Churn Prediction

- Led the design and development of a behavioral analytics platform using **Python** and **Streamlit**, enabling product and growth teams to visualize retention, churn, and user journey metrics in real-time.
- Conducted funnel and cohort analysis to identify top 3 drop-off points, directly informing product improvements that increased activation rate in pilot testing.
- Collaborated with cross-functional stakeholders to define KPIs and trained a time-series **LSTM** model in **Keras**, achieving an **88%** F1-score and improving churn prediction recall by **15%**.

## TECHNICAL SKILLS

- **Programming:** Python(NumPy, Pandas, Matplotlib, Seaborn, TensorFlow, PyTorch), R, SQL, Java
- **Stats Analysis & Modeling:** Regression, Time-Series Forecasting, Clustering, Tree Models, A/B Testing
- **Data Platforms and Tools:** Tableau, Power BI, Excel, AWS, Streamlit, Plotly, Looker, Alteryx, Qualtrics, MySQL, Hadoop, PostgreSQL, Snowflake, Visual Studio