

# 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

Relevant Coursework: NLP, ML in QF, Data Analysis, Statistical Learning, Case Study in ML and Finance

### Rajiv Gandhi Proudhyogiki Vishwavidyalaya

Indore, India

Bachelor of Technology in Computer Science Engineering (Data-Science) | GPA: 3.6

Dec 2020 – Jun 2024

Relevant Coursework: Database Management System, Linear Algebra, Probability & Statistics, Data Engineering

## PROFESSIONAL EXPERIENCE

### Algoventor Solutions Pvt. Ltd.

Indore, India

Data Science Intern (Energy Consumption Forecasting for Smart Grids)

Jun 2024 – Aug 2024

- Designed **energy forecasting model** using **ARIMA** and **machine learning algorithms**, enhancing smart grid stability prediction accuracy by 15%.
- Conducted **exploratory data analysis (EDA)** on a dataset with 15+ features, identifying key factors influencing grid performance. Collaborated with **cross-functional teams** to integrate forecasting models into existing system.
- 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.
- Created visual insights using **Seaborn** and **Matplotlib**, facilitating **data-driven decision-making**.

## 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.
- Visualized product relationships using network graphs, presenting findings in a clear and concise manner to facilitate data-driven decision, resulting in a 25% increase in targeted marketing efficiency.

### 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.
- Implemented a feedback mechanism for users to rate the quality of summaries, enabling continuous improvement of the summarization models, leading to a **10%** improvement in summary accuracy.

### Healthcare Analysis using SQL

- Engineered **SQL queries** to analyse healthcare data and performed **data cleaning** and **transformation techniques**, using **CASE statements** that improved accuracy by ensuring 95% quality standards for meaningful analytics.
- Developed complex SQL queries utilizing aggregate functions, subqueries, window functions, and CTEs.

## TECHNICAL SKILLS

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