Varnica Sharma

+91-779-304-0404 varnicasharma1515@gmail.com LinkedIn GitHub

Summary

Data Science Master's student with specialized expertise in **Machine Learning Engineering**, **Time Series Forecasting**, and **LLM-Driven Workflows**. Experienced in building high-performance data pipelines (FastAPI/Docker) and optimizing predictive models (XGBoost/CatBoost) for complex, high-volatility datasets. Passionate about leveraging quantitative analysis to solve scalable challenges in energy trading and market research.

Education

- **Master of Science in Data Science** *Aug 2024 – May 2026*
  Manipal Academy of Higher Education Manipal, Karnataka

- **Bachelor of Science in Computer Science, Statistics, and Mathematics** *Sep 2021 – May 2024*
  Chandigarh University Kharar, Punjab

Technical Skills

Languages & Frameworks: Python, SQL, PyTorch, Scikit-learn, FastAPI, Docker, AsyncIO, Git/GitHub
Machine Learning & Forecasting: Time Series Analysis (ARIMA, Exponential Smoothing), Boost (Methods), CNN-LSTM, Graph Neural Networks
Generative AI & NLP: Transformers, LLM APIs, RAG, spaCy, Prompt Engineering
Data Analytics & Visualization: Pandas, NumPy, Tableau, Power BI, Matplotlib, Seaborn, Hypothesis Testing, A/B Testing
Cloud & Infrastructure: AWS (S3, Redshift, Athena)

Key Projects

- **Intermittent Demand Forecasting & Inventory Optimization** *Academic Research*
  Python, XGBoost, CatBoost, CNN-LSTM, Time Series Analysis 2025

  – Developed a comparative forecasting framework for **intermittent, high-volatility demand** (analogous to renewable generation) using ensemble tree models and deep learning.
  – Engineered features from exogenous variables (oil prices, holidays) and rolling window statistics to capture non-linear market drivers.
  – Achieved **92% Spike Recall** using **CatBoost**, outperforming classical baselines (Croston, SBA, ARIMA) and reducing **MASE (Mean Absolute Scaled Error)** to **4.53**.
  – Identified critical failure points in Deep Learning models (CNN-LSTM) on sparse data, recommending Gradient Boosting for robust production deployment.

- **Graph Neural Network for Alzheimer's Disease Diagnosis** *Personal Project*
  PyTorch Geometric, Python, Scikit-learn, Streamlit 2025

  – Built multi-modal Patient Similarity Graph (PSG) from ADNI dataset integrating cognitive scores, MRI volumetrics, CSF biomarkers, and APOE 4 genetics using **square-root weighted fusion** to balance feature scales across modalities.
  – Implemented **GCN with weighted message-passing** on k-NN graph using Heat Kernel edge weighting, achieving robust CN/MCI/AD classification with GroupShuffleSplit validation to prevent patient-level data leakage.
  – Deployed **Streamlit dashboard** with prediction probabilities, gradient-based feature importance, and neighbor analysis for clinical explainability and decision support.

Experience

- **Data Science Intern** *Oct 2023 – Nov 2023*
  CodeClause Remote

  – Built automated data cleaning pipelines using Python and Pandas to preprocess raw datasets for downstream modeling.

- Implemented supervised learning algorithms (Random Forest, SVM) to identify trends and derive actionable business insights.
- Created dynamic dashboards to visualize model performance metrics, facilitating stakeholder decision-making.

Achievements & Publications

- **Participant - Novo Nordisk Hackathon 2025**: Developed "Automated eCRF Tool," recognized for engineering excellence and 15x speed optimization.
- **IEEE Publication (CIEES 2025)**: Co-authored "Benchmarking Machine Learning Techniques in Under-Resourced Contexts," presenting findings on NLP model performance for policy analysis.
- **Statistical Consultant**: Advised on experimental design and sample size calculation for 10+ academic studies, ensuring rigorous statistical validity.