Mayur Kumar Sourav

hey5134mayur18@gmail.com · +91 9928964252 · linkedin.com/in/mayur-kumar-sourav · github.com/mayurks23

Summary

Data Science graduate (M.Sc., VIT Vellore) with hands-on experience across the full analytical workflow: data cleaning and EDA, feature engineering, hypothesis testing, and building, validating, and deploying classification, forecasting, and NLP/LLM-based models. Comfortable working from a clearly defined problem statement, following structured development and documentation practices, and turning data into validated, interpretable results.

Experience

Data Science Intern January 2024 – March 2024
Innomatics Research Labs

- Collected and cleaned operational data from multiple sources, applying enrichment routines to produce analysis-ready datasets; ran hypothesis testing and SQL-driven analysis on the AMCAT dataset (10,000+ records) to validate assumptions.
- Built end-to-end ETL pipelines to transform and stage data, enabling repeatable exploratory and confirmatory analysis workflows.
- Deployed a Flask-based model service on AWS EC2 and added runtime monitoring to track model performance.
- Shared analysis results with the team to support delivery-performance and revenue-management decisions.

Projects

GNN + LLM Portfolio Optimization | *M.Sc. Thesis; GraphSAGE, FinBERT, Mistral-7B, PyTorch*

- Built a PyTorch pipeline combining GraphSAGE stock-relationship embeddings with FinBERT news sentiment, alongside a separate LLM-based scorer (Mistral-7B) that ranked stocks from prompted news summaries.
- Designed and ran backtests comparing four strategies – Markowitz baseline, GNN-only, GNN + FinBERT, and the LLM direct scorer – and found news-coverage quality, not model architecture, was the main constraint on the LLM scorer's returns.

Sentiment Analysis with Deep Learning and MLOps | *TensorFlow, PyTorch, Flask API, AWS EC2, Prefect, MLflow*

- Built and compared LSTM and CNN classifiers on 8,518 Flipkart reviews, selecting the final model by F1-score after feature engineering and hyperparameter tuning; reached 85%+ accuracy.
- Served the model through a Flask API on AWS EC2 with monitoring, and automated training and deployment with Prefect, tracking experiments and versions in MLflow.

Time Series Forecasting for the Aviation Industry | *TensorFlow, ARMA, ARIMA, LSTM, SatScan*

- Prepared India civil-aviation financial data (handling autocorrelation, heteroscedasticity, non-stationarity) and compared ARMA, ARIMA, and LSTM forecasting models through structured evaluation.
- Ran SatScan spatial analysis to identify hotspots supporting infrastructure-planning recommendations; guided a team of three on model selection.

Retail Sales Analytics Dashboard | *SQL, Pandas*

- Cleaned and validated 10,000+ Blinkit sales records; engineered derived KPI metrics for segmentation and performance analysis.
- Ran exploratory analysis in SQL and Pandas to identify sales patterns, trends, and anomalies; built an interactive dashboard with drill-downs tracking Total Sales, Average Sales, Item Count, and Average Rating.

Technical Skills

Programming: Python (Advanced), R, SQL, Bash
ML & Analytics: Regression, Classification, Time Series Forecasting, Hypothesis Testing, Feature Engineering, Exploratory Data Analysis, Model Validation (F1-score and related metrics)
NLP & GenAI: FinBERT, LLM-based scoring/prompting (Mistral-7B), Sentiment Classification (LSTM, CNN)
Deep Learning: TensorFlow, PyTorch, Keras, scikit-learn
Data & Tools: Pandas, NumPy, PostgreSQL, Git, Jupyter, VS Code, MLflow, Prefect
Deployment: Flask, REST APIs, AWS EC2

Education

Vellore Institute of Technology July 2024 – May 2026
M.Sc. Data Science *GPA 8.47/10*

Christ University, Delhi NCR May 2021 – May 2024
B.Sc. Data Science & Artificial Intelligence Honours *GPA 8.5/10*

Certifications & Achievements

Python For Everybody, Coursera, December 2021 – February 2022
Second Rank, Fanalytics 2022 – Intra-Data Analysis Competition, Christ University
Languages: English (Fluent), Hindi (Native) **Work Authorization:** Indian Citizen **Availability:** Immediate