AASTHA SINGH

aasthasingh2329@gmail.com | +91 82182 74722 | linkedin.com/in/aastha-singh05 | github.com/Aastha0523

Profile

Data Scientist with 3 years of experience building forecasting and machine learning solutions end-to-end. I'm curious by nature, committed to continuous learning, and enjoy turning messy data into clear, decision-ready insights. Recently, I have deepened hands-on work on RAG and agentic systems to bring reliable AI into enterprise workflows.

Professional Experience

Data Scientist — Accenture   Sep 2024 – Present
Bengaluru, India

Built an **enterprise-grade forecasting platform** for a **leading European industrial technology conglomerate** to predict core business KPIs such as **Orders, Revenue, operating profit)** across **businesses and management hierarchies**.

The system enabled rolling planning on an **EUR 84.1B annual order intake**, providing **confidence bands** and **probability of hitting target** insights.

- Built a **Data Preparation and Outlier Treatment module** handling business sign anomalies, outlier detection using Isolation Forest, Local Outlier Factor (LOF), and One-Class SVM (OCSVM), plus imputation.

- Implemented an automated **EDA (Exploratory Data Analysis) framework** for univariate and multivariate diagnostics (seasonality, stationarity, lag effects, driver relationships)published Streamlit dashboards for self-serve analysis.

- Designed a **Feature Selection and Drift Monitoring pipeline** combining Pearson/Spearman correlation, Mutual Information, and Predictive Power Score (PPS) with optional drift checks.

- Developed a **Univariate Forecasting framework** (Holt-Winters, SARIMA, Prophet, Orbit (DLT/LGT)) to forecast external drivers (GDP, PMI, interest rates) with prediction intervals.

- Architected a **Multivariate Modeling module on Azure Machine Learning (Azure ML)** with walk-forward cross-validation (CV), Optuna hyperparameter tuning, and a unified model factory (LightGBM, XGBoost, Random Forest, SARIMAX, SVR, Linear Models, Prophet)improved median MAPE by **18–25%** vs. baselines.

- Built a **Post-modeling validation layer** including residual diagnostics, variance and level sanity checks, STL/Prophet decomposition review, and SHAP explainability.

- Delivered a **Distribution and Probability module** using residual bootstrapping plus KDE to generate sigma-bounds and probability of hitting target metrics.

Reporting and Data Analyst — HCL   Feb 2023 – Apr 2024
Noida, India

- Developed ARIMA/ETS/Prophet forecasting and Isolation Forest anomaly detection, cutting inventory variance by **15%**.

- Built SQL ETL from ServiceNow and Snowflakeachieved **3x** runtime improvement via indexing, partitioning, and optimized joins.

- Automated Python dashboards (Matplotlib/Plotly) with MAPE, RMSE, and sMAPE validationreduced reporting time by **60%**.

- Co-designed risk monitoring combining predictive models and rule enginesexpanded issue-detection coverage by **25%**.

Recent Learning and Hands-on Projects

- Built a **Finance RAG assistant** using **SharePoint as the knowledge base** (policies, FP&A decks, KPI definitions, FAQs). Implemented chunking, embeddings, and vector search for grounded answers with traceable citations.

- Orchestrated **tool-using agent workflows** on top of RAG for multi-step tasks (document retrieval, summarization, Q&A, action routing), including memory and guardrails for safe enterprise usage.

Skills

Machine Learning: Regression and Classification, Time-Series Forecasting (ARIMA/ETS/Prophet/SARIMAX), Tree Models (Random Forest, GBDT, XGBoost, LightGBM), Model Selection, Cross-Validation, Feature Engineering, Explainability (SHAP)
GenAI: RAG pipelines, embeddings and vector search, agentic workflows, evaluation for groundedness/factuality
Programming: Python, SQL
Data Engineering: SQL optimization (indexing, partitioning, query tuning), ETL/ELT, data quality checks
Cloud and MLOps: Azure Machine Learning (Azure ML), GCP (BigQuery, Cloud Storage, Vertex AI basics), MLflow, Docker, CI/CD
Visualization: Matplotlib, Plotly, Streamlit
Databases and Tools: Snowflake, MySQL   **Version Control:** Git, GitHub

Education

B.Tech, Computer Science Engineering, College of Engineering Roorkee   2017 – 2021

Awards

Winner — EY HackPions 4.0 Hackathon (2022); Zonal Winner — Smart India Hackathon (2020)