

Shubham Bhagat

+91 9653142137 | shubhamb836@gmail.com | [LinkedIn](#) | [GitHub](#)

PROFESSIONAL SUMMARY

Data Analyst and Junior Data Scientist with 2 years of experience in machine learning, predictive analytics, SQL, Python, RAG/LLM apps, statistical modeling, and data visualization across finance, non-profit, and other domains. Skilled in building end-to-end ML pipelines, forecasting systems, and AI-powered applications. Strong foundations in EDA, feature engineering, data cleaning, BI dashboards, statistical analysis, and model evaluation. My vision is to continuously learn and grow while contributing my skills and knowledge to help the organization achieve its goals. Seeking Data Analyst roles.

TECHNICAL SKILLS

Programming Languages: Python, SQL, R

ML & AI: Classification, Regression, Clustering, Time-Series, NLP, RAG, LLMs, CNNs, Airflow

Generative AI: LLM Applications, Prompt Engineering, Vector Databases, FAISS, Ollama, LangChain

Libraries: Pandas, NumPy, scikit-learn, TensorFlow, Keras, XGBoost, imbalanced-learn

Tools: FastAPI, Docker, Power BI, Tableau, Streamlit, Git, SSIS, Google Apps Script

Cloud: AWS (S3, EC2), GCP (BigQuery, Vertex AI)

Core Skills: EDA, Feature Engineering, ETL, Data Pipelines, Visualization, Reporting

Others : Teamwork, Responsibility, Sincerity, Adaptability, Problem Solving

Languages: English (Professional), Hindi (Native), Marathi (Native)

PROFESSIONAL EXPERIENCE

Data Analyst

Peakline Business Consultancy - July 2026 Current

- Maintain Power BI dashboards across stakeholders, ensuring data accuracy and timely refreshes.
- Supported data extraction tasks using SQL and BigQuery for stakeholder reporting.
- Documenting data processes and workflows to support standardization across the team.
- Designing and developing a forecasting dashboard to support business planning.

Data Analyst – Senior Executive

Utkarsh Global Foundation, Mumbai - May 2025 – June 2026

- Cleaned, validated, and analyzed datasets using Python and SQL to support strategic decisions.
- Automated donor and hospital processes using Google Apps Script and API workflows
- Built dashboards to track Surgery count, treatments count, donor trends across locations enabling leadership to optimize the ground operations

Junior Data Scientist

Freelance - Oct 2024 to May 2025

- Built a sales forecasting model (MAPE < 30%) using time-series and regression techniques.
- Performed econometric modelling and hypothesis testing using Python and R.
- Conducted EDA, feature engineering, visualizations, and insights for research/industry clients.
- Delivered structured reports summarizing models, assumptions, and outcomes.

Data Science Intern

L&T Finance, Mumbai - May 2024 – Oct 2024

- Conducted risk analytics across two-wheeler, farm loan, and microfinance portfolios.
- Built category-tagging logic for bank transaction data using ML + rule-based methods.

- Improved preprocessing workflows using Python and SQL, reducing manual intervention.
- Evaluated models using ROC-AUC, F1-Score, Precision, and Recall for credit risk scoring.

PROJECTS

MarketMetric – AI Stock with Multi-Agent AI Analysis System

- Built an end-to-end AI-powered stock analysis platform for Indian equities, enabling company selection via a Streamlit dashboard that triggers a LangGraph multiagent pipeline.
- Designed a Supervisor-Planner-Agent architecture with dedicated Market, Researcher, Analyst, and Reporter agents for coordinated, structured analysis workflows.
- Stack: Python, LangGraph, FastAPI, Streamlit, FinBERT, Groq AI, Yahoo Finance API, MongoDB Atlas.

RAG + LLM Chatbot (Document Q&A)

- Built a Retrieval-Augmented Generation (RAG) chatbot that answers queries from uploaded PDFs with accurate, context-aware responses.
- Developed a Streamlit web app integrating local LLMs (Llama 3.1, Mistral, Gemma 2) via Ollama for privacy-first offline inference.
- Implemented FAISS-based similarity search with nomic-embed-text embeddings, enabling fast and efficient document retrieval.

Credit Card Fraud Detection

- Built a machine learning pipeline to detect fraudulent credit card transactions on a highly imbalanced dataset (<1% fraud cases)
- Performed data preprocessing, feature scaling, and exploratory analysis to identify patterns in anonymised transaction data.
- Handled class imbalance using SMOTE, under sampling, and ensemble methods, improving minority class detection.
- Trained and compared models including Logistic Regression, Random Forest, and XGBoost

EDUCATION

- **M.Sc. In Data Science and Big Data Analytics 2022 – 2024**
MIT World Peace University, Pune
- **B.Sc. In Computer Science 2018 – 2021**
SIES College of Arts, Science & Commerce, Mumbai

CERTIFICATIONS

- AWS Academy Cloud Foundations
- Google Cloud Program

PUBLICATION

Trend Survey: Neural Network for Underwater Image Enhancement (2025)