Rohit Akole

rohitakole5@gmail.com | rohitakole50.github.io | [(959) 223 0185](tel:9592230185) | [LinkedIn: rohitakole](https://linkedin.com/in/rohitakole) | Hartford, CT

DATA SCIENTIST

Data professional who goes beyond delivering reports by translating insights into actionable strategies. Experienced across multi-industry projects and fast-paced startup environments, with a proven ability to adapt, innovate, and drive impact. Placed in the top 10% at the Humana-Mays competition, demonstrating leadership in solving complex challenges.

EDUCATION

University of Connecticut, Hartford, CT Aug 2023 - May 2025
Master of Science (M.S.), Business Analytics and Project Management (Data Science) GPA: 4.00/4.00
- **Coursework:** Intro to Project Management, Statistics Business Analytics, Predictive Modeling, Project Leadership and Communications, Data Mining and Business Intelligence, Data Science using Python, Agile Project Management and Methodologies, Data Management and Business Process Modeling, Business Decision Modeling, Intro to Deep Learning, Adv Business Analytics and Project Management Capstone

University of Bridgeport, Bridgeport, CT Jan 2017 - May 2019
Master of Science (M.S.), Technology Management (IT and Big Data) GPA: 3.89/4.00
- **Coursework:** Statistical Quality Control, Finance & Accounting, Strategic Sourcing & Vendor Management, Intro to Big Data and Data Science, Intro to SQL and R Programming

North Maharashtra University, Jalgaon, India Jun 2012 - May 2015
Bachelor of Business Management
- **Coursework:** Financial and Cost Accounting, Corporate Accounting, Managerial Economics, MS Access, Business Mathematics and Statistics, System Analysis and Designing, Business Management, MIS and ERP, Oracle, Auditing, Information System Audit

SKILLS & CERTIFICATIONS

Programming & Languages: Python, SQL, R, HTML5, CSS3, PySpark, Scala, CUDA, Pandas, NumPy, Scikit-learn
Data Science & Analytics: Machine Learning, Deep Learning, Predictive Modeling, A/B Testing, Time Series Forecasting, Natural Language Processing (NLP), Generative AI, Sentiment Analysis, Feature Engineering, Statistical Modeling, Model Evaluation, Exploratory Data Analysis (EDA)
Tools & Libraries: GitHub, Tableau, Power BI, Django, Excel, Access, SAS Studio, JIRA, SAS Enterprise Miner
Data Engineering & Infrastructure: Extract Transform Load (ETL), Data Pipelines, Data Warehousing, APIs, REST APIs, SQL Server, Relational Databases, Big Data, Docker, GPU Programming, MLOps
LLMs & Agentic AI: Large Language Models, Agentic AI, Semantic Search, Ollama, Mistral, LLaMa3, Gemini AI
Business & Methodologies: Business Intelligence, Data Analysis, Data Mining, Data Visualization, Dashboards, Agile, Scrum, Strategic Recommendations, Cross-Functional Collaboration, SaaS
Certifications: [IBM Data Science, professional certification by IBM on Coursera, 2021](); [IBM Data Analyst, professional certification by IBM on Coursera, 2022]()

EXPERIENCE

Data Scientist Intern, Alo Index Jan 2025 – Apr 2025
- Automated ESG data integration for 6,000+ hotels by extracting and mapping data from 33+ sustainability certifications using APIs and Python. Verified claims across 1.2M+ records, resolved discrepancies through hotel outreach, and pre-filled 300+ ESG assessment criteria, reducing manual effort and improving data accuracy.
- Developed a local LLM-powered assessment engine using Ollama, mistral, and semantic search to map 300+ ESG questions to certification text, enabling automated evidence extraction from PDF documents and reducing manual review time by 80%.
- Collaborated with the founders to leverage business insights, enhance product strategies and boost operational efficiency through data-driven recommendations, gaining leadership exposure in a high-paced startup environment.

Data Science Capstone Consultant – Stanley Black & Decker, University of Connecticut Jan 2025 – May 2025
- Forecasting 12-month warranty claim volumes and costs using time series modeling on 5 years of historical data (~375K claims, 1.6M+ rows) to support strategic planning.
- Leveraging LLM-based (llama3 using Docker) text mining on dealer communications (diagnosis and repair fields) to identify recurring issues and extract 9 actionable features.
- Developed an interactive Tableau dashboard to present historical trends, forecast outputs, and text mining driven visualizations (e.g., word clouds) for cross-functional use.

President and Founder, Modlee AI/ML Student Club, University of Connecticut   Sept 2024 – May 2025
- Founded and led the Modlee AI/ML Student Club (MAIC), creating a collaborative platform for students of all backgrounds to explore AI and machine learning.
- Organized hackathons, competitive events and projects, providing members with hands-on experience and industry insights through events featuring AI/ML professionals.
- Featured twice in Student Highlights by the University for leadership and dedication to fostering a supportive, innovative learning environment in AI/ML.

Graduate Teaching Assistant – Introduction to Deep Learning, University of Connecticut   Jan 2025 – May 2025
- Evaluated graduate-level assignments with a focus on code quality, model performance, and documentation.
- Delivered individualized, constructive feedback to support student learning, model implementation best practices.
- Collaborated with the teaching team to maintain academic standards and ensure timely grading cycles.

Data Scientist Team Lead, Humana-Mays Healthcare Analytics Case Competition, 2024   Aug 2024 – Oct 2024
- Led a team to top 10% in the national competition to address preventive healthcare visit gaps.
- Developed XGBoost model for a dataset of 300+ columns & 1.6 million rows, achieving 77% accuracy (AUC: 0.7686), identifying key features like claims history & chronic conditions.
- Proposed actionable strategies including mobile clinics, telehealth integration, & targeted outreach programs to improve healthcare access & patient engagement.

PROJECTS

Predictive Modeling of Adolescent Digital Overuse (Python, TSFresh, SMOTE, FNN, SHAP, Parquet Processing)
- Processed 986 time-series parquet files to engineer 300+ features from physical activity, heart rate, and sleep data, enabling predictive modeling using deep learning techniques of adolescent internet overuse.
- Achieved R^2 of 0.72 using Feed-Forward Neural Network, with SHAP analysis revealing sedentary time and disrupted sleep as top predictors; applied SMOTE to improve model generalization and fairness.
- Delivered actionable insights via multivariate visualizations and correlation heatmaps, supporting early intervention strategies for pediatric behavioral health.

Insurance Fraud Detection Using Machine Learning (Python, Decision Tree, Random Forest, Logistic Regression)
- Engineered machine learning models to detect insurance fraud, achieving a precision of 92% and a recall of 88%.
- Preprocessed and analyzed 50,000+ insurance claims data, performing feature selection and extraction which improved model accuracy by 15%.
- Implemented the final Random Forest model with hyperparameter tuning, reducing false positives by 20% and overall fraud detection rate by 25%, leading to projected annual savings of $500,000.

AI-Powered Gmail Assistant (Python, Gmail API, Gemini AI, NLP, TextBlob, Pandas, PyMuPDF, LLM, Agentic AI)
- Developed an AI-driven agent that reads, summarizes, and prioritizes unread Gmail messages, parsing both email content and attachments using NLP and LLM-based summarization.
- Automated email replies with Gemini AI by generating sentiment-aware summaries and three personalized response options with dynamic placeholders and in-thread Gmail API integration.
- Presented a live demo at AI Day 2025 organized by Launc[H] in Hartford to 50+ attendees, showcasing the end-to-end intelligent workflow and internal logic behind real-time, human-like email handling.

Sentiment Analysis of 2020 US Presidential Election Tweets (SAS Miner, Sentiment Analysis, Text Mining, TFIDF)
- Analyzed over 1 million tweets related to the 2020 US Presidential Election, aiming to understand public sentiment during key events.
- Engineered and optimized machine learning models (Logistic Regression, Decision Tree, Random Forest) using SAS Enterprise Miner for sentiment classification.
- Achieved 89% accuracy and an F1-score of 0.85, providing actionable insights through sentiment trend analysis and visualization, which informed real-time decision-making for stakeholders.