Rohit Akole

rohitakole5@gmail.com | rohitakole50.github.io | (959) 223 0185 | LinkedIn: rohitakole | Hartford, CT

DATA SCIENTIST

Curious and data-driven, with hands-on experience in fast-paced startup environments, machine learning, and predictive modeling. Placed in the top 10% at the Humana-Mays competition, demonstrating leadership in solving complex data challenges. Skilled at collaborating with leaders to drive strategy, boost efficiency, and fuel business growth.

EDUCATION

University of Connecticut, Graduate School of Business, Hartford, CT May 2025
Master of Science (M.S.), Business Analytics and Project Management (Data Science) GPA: 4.00/4.00
University of Bridgeport, School of Engineering, Bridgeport, CT May 2019
Master of Science (M.S.), Technology Management (IT and Big Data) GPA: 3.89/4.00
North Maharashtra University, School of Management, Jalgaon, India May 2015
Bachelor of Business Management

TECHNICAL SKILLS & CERTIFICATIONS

Programming: Python, SQL, Machine Learning, R Programming, HTML5, CSS3
Tools: PyCharm, Jupyter Notebook, Tableau, GitHub, Docker, SAS Studio & Miner, SQL Server, Visio, Excel, Access
Certifications: IBM Data Science, professional certification by IBM on Coursera, 2021; IBM Data Analyst, professional certification by IBM on Coursera, 2022

PROFESSIONAL EXPERIENCE

Data Science Associate Intern, Alo Index Jan 2025 – Present
- Automated data integration from 10+ sustainability certifications using Python, enriching ESG profiles for 6,000+ hotels and improving data accuracy through verified matches out of 1.2M+ dataset.
- Mapped certification criteria to 200+ evaluation questions, enabling dynamic pre-filling that saved hours of manual work per hotel and enhanced platform usability.
- Collaborated with the founders to leverage business insights, enhancing product strategies and boosting 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 – Present
- 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 actionable insights.
- Developed an interactive Tableau dashboard to present historical trends, forecast outputs, and text-driven visualizations (e.g., word clouds) for cross-functional use.

President and Founder, Modlee AI/ML Student Club, University of Connecticut Sept 2024 – Present
- 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.

ACADEMIC EXPERIENCE

Graduate Teaching Assistant – Introduction to Deep Learning, University of Connecticut Jan 2025 – Present
- 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.

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, XGBoost, SHAP, Parquet Processing)

- Processed 986 time-series parquet files to engineer 300+ features from physical activity, heart rate, and sleep data, enabling predictive modeling of adolescent internet overuse.
- Achieved R^2 of 0.72 using XGBoost, 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)

- Built an intelligent agent that reads unread Gmail messages, ranks them by urgency, and summarizes both email body and attachments (PDF, DOCX, XLSX, etc.) using NLP and LLMs for contextual understanding.
- Engineered an end-to-end automation pipeline with Gemini AI to detect sentiment, extract key insights, and generate 3 professional, editable reply options with placeholder inputs.
- Enabled real-time email responses via Gmail API, enhancing communication efficiency and preserving conversation continuity for professionals managing high-volume inboxes.

Sentiment Analysis of 2020 US Presidential Election Tweets (SAS Miner, Sentiment Analysis, Text Mining)

- 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.

Walmart Sales Forecasting (Time Series Analysis, SARIMAX, SAS Studio, Machine Learning)

- Executed weekly sales forecasts for 45+ Walmart departments to optimize inventory management and minimize financial losses.
- Developed and refined SARIMAX and Seasonal Decomposition models, incorporating features like promotional campaigns, holidays, and economic indicators to improve prediction accuracy.
- Achieved 90% forecasting accuracy, leading to a 20% reduction in stockouts and a 15% decrease in excess inventory, resulting in an estimated $2.5 million in annual cost savings.