

VAIBHAV VIKAS GAIKWAD

(315) -952-2931 • gaikwadvaibhav114@gmail.com • [LinkedIn URL](#) • [Portfolio](#)

EDUCATION

Syracuse University, School of Information Studies, Syracuse, NY	May 2025
Master of Science Applied Data Science	
University of Mumbai, India	June 2022
Bachelor of Engineering Computer Engineering	

EXPERIENCE

Research Analyst, Syracuse University, Syracuse, NY	July 2025 – present
• Applied Generative AI (LangChain, Hugging Face) to accelerate workflows, increasing model development speed by 25% and improving reproducibility .	
• Built prompt engineering scripts and LLM pipelines for text classification and summarization , reducing iteration cycles for research teams by 20%.	
• Documented best practices on responsible AI use , ensuring ethical deployment and alignment with real-world product environments .	
Data Science Researcher, Nexis Student Labs, Syracuse, NY	January 2025 – May 2025
• Designed scalable data pipelines for medical datasets using Python/SQL , reducing processing time by 35% on multi-gigabyte datasets .	
• Deployed predictive models and automated workflows for health risk classification , integrating outputs via modular APIs to deliver end-to-end solutions.	
• Established and managed comprehensive ML performance monitoring frameworks , including automated alerts for model drift and data anomalies using custom evaluation scripts, ensuring continuous model reliability.	
AI Researcher, HeadOn, Remote	June 2024 – August 2024
• Engineered and fine-tuned advanced Generative AI models (BERT, GPT) to autonomously generate concise meeting summaries and accurately classify discussion sentiment (positive/negative), enhancing operational efficiency.	
• Architected and deployed real-time inference pipelines via robust REST APIs , seamlessly integrating AI capabilities into the core product and reducing manual notetaking by 40%, demonstrating proficiency in building production-ready systems.	
• Achieved over 90% accuracy in sentiment detection through iterative model optimization, A/B testing of different architectures, and continuous feedback loop integration, significantly improving actionable insights.	
Data Scientist, NeoSOFT, India	June 2022 – July 2023
• Performed large-scale data mining and statistical analysis on customer behavior datasets using SQL and Python, uncovering key patterns that drove 15% improvement in product engagement metrics .	
• Built and validated predictive models (clustering, regression, risk scoring) to support data-driven business decisions, achieving 20% higher forecasting accuracy compared to baseline methods.	
• Developed automated ETL pipelines and visualization dashboards in Python and Power BI to streamline reporting on multi-terabyte datasets, reducing analytics turnaround time by 30%.	

PROJECTS

Credibility Detection of Health Web Blogs Using Explainable AI

- Developed an **Explainable AI (XAI) model** leveraging **LIME** and **SHAP** for transparent insights into model predictions, assessing the credibility of health web blogs and enhancing trust in filtering.
- Utilized **Natural Language Processing (NLP) techniques**, including **TF-IDF** and **Word Embeddings**, combined with machine learning algorithms like **Support Vector Machines (SVM)** and **Gradient Boosting**, to analyze textual content.
- Implemented a robust **classification pipeline** to identify deceptive patterns and accurately classify blog credibility, demonstrating expertise in real-world problem-solving for misinformation detection.

SKILLS & ACHIEVEMENTS

Scripting Languages & Databases: Python, SQL, R, SAS, Matlab, PySpark, Pandas, NumPy, MySQL, PostgreSQL, Snowflake, SQL Server

Machine Learning & Statistics: Regression, Classification, Quantitative analysis, Neural Networks (ANN, CNN, RNN), LSTM, Gradient Boosting, Random Forests, Clustering (K-Means, DBSCAN), Pattern Recognition, Inferential Statistics (ANOVA, Hypothesis Testing), Explainable AI (LIME, SHAP)

Data Engineering & ETL: ETL (Extract, Transform, Load) pipeline development, Data Cleaning, Feature Engineering, Workflow Automation, Apache Spark, AWS S3, Data Lake Architectures

Visualization & BI: Tableau, Power BI, Matplotlib, Seaborn, Plotly, Google Data Studio

MLOps & Cloud: AWS, Azure, Docker, Kubernetes, Git, CI/CD, REST APIs