

# VAIBHAV VIKAS GAIKWAD

(315) -952-2931 • [gaikwadvaibhav114@gmail.com](mailto: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
<ul style="list-style-type: none"><li>Applied <b>Generative AI (LangChain, Hugging Face)</b> to accelerate workflows, increasing <b>model development speed</b> by <b>25%</b> and improving <b>reproducibility</b>.</li><li>Built <b>prompt engineering scripts</b> and <b>LLM pipelines</b> for <b>text classification</b> and <b>summarization</b>, reducing <b>iteration cycles</b> for research teams by <b>20%</b>.</li><li>Documented <b>best practices</b> on <b>responsible AI use</b>, ensuring <b>ethical deployment</b> and alignment with <b>real-world product environments</b>.</li></ul>	
Data Science Researcher, Nexis Student Labs, Syracuse, NY	January 2025 – May 2025
<ul style="list-style-type: none"><li>Designed <b>scalable data pipelines</b> for <b>medical datasets</b> using <b>Python/SQL</b>, reducing <b>processing time</b> by <b>35%</b> on <b>multi-gigabyte datasets</b>.</li><li>Deployed predictive models and automated workflows for <b>health risk classification</b>, integrating outputs via modular APIs to deliver end-to-end solutions.</li><li>Established and <b>managed comprehensive ML performance monitoring frameworks</b>, including automated alerts for model drift and data anomalies using custom evaluation scripts, ensuring continuous model reliability.</li></ul>	
AI Researcher, HeadOn, Remote	June 2024 – August 2024
<ul style="list-style-type: none"><li><b>Engineered and fine-tuned advanced Generative AI models (BERT, GPT)</b> to autonomously generate concise meeting summaries and accurately classify discussion sentiment (positive/negative), enhancing operational efficiency.</li><li>Architected and deployed real-time inference pipelines via robust <b>REST APIs</b>, <b>seamlessly integrating AI capabilities into the core product</b> and <b>reducing manual notetaking by 40%</b>, demonstrating proficiency in building production-ready systems.</li><li><b>Achieved over 90% accuracy in sentiment detection through iterative model optimization, A/B testing</b> of different architectures, and continuous feedback loop integration, significantly improving actionable insights.</li></ul>	
Data Scientist, NeoSOFT, India	June 2022 – July 2023
<ul style="list-style-type: none"><li><b>Performed large-scale data mining and statistical analysis</b> on customer behavior datasets using SQL and Python, uncovering key patterns that drove <b>15% improvement in product engagement metrics</b>.</li><li><b>Built and validated predictive models</b> (clustering, regression, risk scoring) to support data-driven business decisions, achieving <b>20% higher forecasting accuracy</b> compared to baseline methods.</li><li><b>Developed automated ETL pipelines and visualization dashboards</b> in Python and Power BI to streamline reporting on multi-terabyte datasets, reducing analytics turnaround time by <b>30%</b>.</li></ul>	

## PROJECTS

Credibility Detection of Health Web Blogs Using Explainable AI
<ul style="list-style-type: none"><li><b>Developed an Explainable AI (XAI) model</b> leveraging <b>LIME</b> and <b>SHAP</b> for transparent insights into model predictions, assessing the credibility of health web blogs and enhancing trust in filtering.</li><li><b>Utilized Natural Language Processing (NLP) techniques</b>, including <b>TF-IDF</b> and <b>Word Embeddings</b>, combined with machine learning <b>algorithms</b> like <b>Support Vector Machines (SVM)</b> and <b>Gradient Boosting</b>, to <b>analyze textual content</b>.</li><li><b>Implemented a robust classification pipeline</b> to <b>identify deceptive patterns</b> and <b>accurately classify blog credibility</b>, demonstrating expertise in real-world problem-solving for misinformation detection.</li></ul>

## 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