Vaibhav Vikas Gaikwad

Syracuse, NY | gaikwadvaibhav114@gmail.com|Linkedin:www.linkedin.com/in/vaibhav-vikas-gaikwad/|
GitHub: https://github.com/vaibhavgaikwad7

EDUCATION

Syracuse University, School of Information Studies, USA — Syracuse, NY
Master of Science in Applied Data Science | **Concentration:** Artificial Intelligence| **GPA:** 3.98 — May 2025
Relevant Courses: Big Data Analytics, Applied Machine Learning, Deep Learning, Visual Analytics Dashboard, NLP

University of Mumbai, India — Mumbai, India
Bachelor of Engineering (BE) in Computer Engineering| **Top 10% of Class** — June 2022
Relevant Courses: Data Structures & Algorithm, Natural Language Processing, Object Oriented Programming

WORK EXPERIENCE

Syracuse University, Syracuse , NY | Research Data Scientist — July 2025 – Present

- **Automated SQL-driven Tableau dashboards**, **reducing reporting delays by 30%** and enabling faster research funding decisions by **25%**, by building optimized data queries and automated refresh processes.
- Implemented enterprise-wide data-quality checks and validation rules, **reducing dataset errors by 20% and increasing trust in institutional reporting**, using SQL and Python-based validation scripts.
- **Designed analytics dashboards for 10k+ records**, **improving visibility** into department performance and supporting budget planning, using Tableau's advanced modeling and KPI visualizations.

NEXIS Labs, Syracuse, NY | Data Science Researcher — January 2025 – May 2025

- **Cleaned, transformed, and analyzed 80+ GB of healthcare data**, **uncovering actionable patient and operational trends** by using PySpark, SQL, and Pandas to engineer features and aggregate large datasets.
- Built automated **Power BI dashboards integrated with Azure SQL**, **eliminating manual refresh cycles** and enabling real-time decision-making by configuring scheduled refresh pipelines.
- **Developed data-validation and anomaly detection scripts**, **reducing manual QA effort by 35%**, by building automated checks using Python and SQL-based rule systems.

Syracuse University, Syracuse, NY|Teaching Assistant — August 2024 – December 2024

- **Supported 50+ students in SQL, R, Python, and data modeling**, **raising overall project success rate to 90%**, by providing structured guidance, office hours, and debugging assistance.
- Reviewed and optimized student data pipelines, **increasing accuracy and efficiency of workflows** by correcting issues in R/Python scripts for data cleaning, feature engineering, and visualization.
- **Facilitated labs on statistical testing, EDA, and BI tools**, **strengthening students' end-to-end analytics capabilities**, by demonstrating applied use of Tableau, Power BI, and R's analytical libraries.

HeadOn, New York, NY|AI Researcher intern — May 2024 – August 2024

- **Designed, trained, and evaluated NLP models (BERT, GPT)** for **automated meeting summarization and sentiment classification**, translating **open-ended product needs** into **measurable modeling objectives** and **user-impact metrics**.
- Built and deployed real-time inference pipelines with **production-grade REST APIs**, enabling **product teams** to **integrate NLP capabilities into user workflows** and **reducing manual note-taking effort by 40%**.
- Ran controlled **A/B testing** across model architectures and inference strategies, analyzing **accuracy, latency, and downstream user-signal metrics** to guide iteration and **achieve >90% sentiment detection accuracy**, improving the **quality of insights delivered to end users**.

PROJECTS

Health Blog Credibility Detection, Python, Scikit-learn, TF-IDF, Embeddings, SHAP

- **Engineered and trained ML classifiers** (Logistic Regression, SVM, Gradient Boosting) using TF-IDF and semantic embeddings to detect misinformation in health blogs, **improving classification accuracy across multiple source domains**.
- **Built an explainability framework** using **SHAP and LIME**, generating token-level and feature-level attributions that **increased model transparency and stakeholder trust**.

Finance Portfolio Optimization, Python, CNN, LSTM, Attention

- **Developed a BiLSTM + Attention model for stock-return forecasting**, **outperforming regression baselines**, by leveraging deep learning architectures for temporal pattern extraction.
- **Engineered feature pipelines and preprocessing workflows**, **improving forecasting accuracy and stability**, by applying normalization, windowing, and lag-feature techniques.

TECHNICAL SKILLS & CERTIFICATIONS

- **Programming:** Python (Pandas, NumPy, Scikit-learn, TensorFlow/PyTorch), R (dplyr, ggplot2), SQL (PostgreSQL, T-SQL), PySpark
- **Statistical & ML Techniques:** Regression, Classification, Time-Series Forecasting, Random Forest, XGBoost, Neural Networks, Natural Language Processing, A/B Testing, Statistical Testing (t-tests, ANOVA)
- **Data Engineering & Wrangling:** ETL/ELT, Data Cleaning, Feature Engineering, Data Modeling, Data Profiling, Azure SQL, Snowflake, BigQuery, Databricks
- **Visualization & BI:** Tableau, Power BI, Matplotlib, Seaborn, Looker,Folium, Shiny
- **Tools:** Git, Jupyter, Databricks, Azure Data Factory, Power Query, Airflow
- **Certifications:** AWS Machine Learning Engineer - Associate, IBM - Data Analytics and Visualization Tools