

SIDDHI NALAWADE

Buffalo, NY 14214

716-465-1313 | siddhisu@buffalo.edu | LinkedIn | GitHub

Summary

Data Science and Software Engineering graduate student with experience in biometric authentication, cryptographic protocols, and machine learning. Skilled in building scalable ML pipelines, secure APIs, and privacy-preserving systems through work at Retica Technologies and Accenture. Passionate about developing secure, data-driven solutions for real-world applications.

Education

University at Buffalo, Buffalo, NY

Aug 2024 – present

Master of Science in Data Science

Related courses: Machine Learning, Statistical Learning, Data Intensive Computing, Data Models and Query Language, Probability Theory, Statistical Learning & Data Mining, Predictive Modeling

University of Mumbai, Mumbai, India

June 2021

Bachelor of Engineering in Electronics Engineering

Technical Skills

- **ML & Statistical Modeling:** Predictive modeling, classification, clustering, neural networks, time series analysis, feature engineering, PCA, performance evaluation, model validation, biometric data analysis
- **Programming & ML Frameworks:** Python, SQL, R, Java, PyTorch, TensorFlow, Keras, Scikit-learn, Pandas, NumPy, SciPy, MLFlow, Flask, Streamlit, exploratory data analysis, model deployment, Docker containers, metric-based optimization
- **Data Engineering & Infrastructure:** Apache Spark, Hadoop, Google BigQuery, MySQL, PostgreSQL, Oracle, SQLite3, MongoDB, data pipelines, data fragmentation, privacy-preserving systems, biometric API integration
- **Security & Identity Technologies:** Biometric authentication, cryptographic protocols, secure API development, data distribution methods, privacy-preserving technologies, identity verification systems
- **Visualization & Reporting:** Tableau, Power BI, Matplotlib, Seaborn, executive dashboards, model performance tracking, stakeholder communication, technical reporting
- **Collaboration & Delivery:** Agile workflows, code reviews, cross-functional collaboration, remote work coordination, DevOps tools, version control (GitHub), reproducibility standards

Work Experience

Retica Technologies (CosmosGenerate, Inc.)

May 2025 – Present

Software Engineer/ Data Science Intern

Remote/ New York, USA

- Developed and optimized biometric authentication APIs using secure, scalable prototypes, enhancing accuracy and performance for high-security applications.
- Analyzed biometric data and supported cryptographic protocol research to improve match precision and data security.
- Engineered privacy-preserving techniques including data fragmentation and distribution to prevent unauthorized data reconstruction.
- Collaborated cross-functionally on API integrations, participated in code reviews, and contributed to R&D in identity verification technologies.

Accenture Solution PVT.LTD

Dec 2021 – Jul 2024

Software Analyst

Mumbai, India

- Analyzed large datasets using SQL in Oracle SQL Developer to extract actionable insights, supporting data-driven business decisions through comprehensive business intelligence reports.
- Optimized data workflows with Shell scripting and UNIX commands, improving data processing efficiency by 15% and enhancing data pipeline reliability.
- Improved data quality and integrity by 20% through rigorous validation checks, while automating incident tracking in ServiceNow, reducing resolution time by 25% for data-related issues.

Internship Program

Jun 2020 – Aug 2020

Data Science Intern

Mumbai, India

- Developed and optimized classification and regression models in Python, focusing on predictive analytics, model accuracy enhancement, and feature engineering techniques.
- Engineered a K-NN-based Heart Disease Classifier, applying supervised learning and advanced data preprocessing on large datasets for precise disease risk prediction.

Projects

Handwritten Digit Classification Pipeline | *Logistic Regression, Random Forest, XGBoost, MLFlow, Docker, Streamlit*

- Developed an end-to-end machine learning pipeline for handwritten digit recognition using the MNIST dataset, attaining 99% accuracy with XGBoost and Random Forest, leveraging SQL-based data transformation, feature engineering, and PCA for dimensionality reduction.
- Deployed a real-time Streamlit app with scalable deployment via Docker and integrated MLFlow for performance tracking, enhancing accessibility for automation in education and document processing.

Short-Term Temperature Forecasting (ARIMA-LSTM Hybrid Model) | *R, Time Series Forecasting*

- Created and optimized a hybrid ARIMA-LSTM model to forecast short-term temperature fluctuations across 35 U.S. cities, achieving RMSE \downarrow 0.03 through advanced time series analysis
- Supported data-driven decision-making in agriculture, energy, and urban development by providing climate forecasts for proactive planning.

Caption Bot with Assistive Voice | *CNN, LSTM, Flask, API Integration*

- Designed and deployed an interactive image captioning system combining CNN (ResNet-50) and LSTM models, achieving 95% caption accuracy through effective model training and evaluation.
- Integrated an assistive voice API to enhance accessibility for visually impaired users, showcasing a practical application of deep learning models.

Personalized Tourism Recommender using Chatbot | *Django, Dialogflow, API calls*

- Constructed a rule-based recommendation engine with AI-driven chatbot integration via Dialogflow, delivering personalized travel suggestions.
- Increased user engagement by 30% through dynamic UI design, API-based data integration, and cloud functionalities.