

SIRITHA CHIDIPOTHU

sc2855@scarletmail.rutgers.edu | New Brunswick, NJ, 08901 | [Linked In](#) | 732-253-2314

EDUCATION:

Rutgers, State University of New Jersey

Master of Science in Data Science (Statistics)

Gandhi Institute of Technology and Management

Bachelor of Technology in Computer Science and Engineering

New Brunswick, NJ, USA

Sep 2024 - May 2026

Vizag, AP, India

Jul 2020 - May 2020

SKILLS:

Programming & Querying: Python, SQL, C++ , Power Query

Data Engineering & ETL: Apache Airflow, Pandas, NumPy, Azure Datafactory, ETL Pipelines

Data Analysis & Modeling: Time-Series Forecasting, Feature Engineering, Root-Cause Analysis

Big Data Processing: PySpark, Hive, HDFS

Visualization & Reporting: Power BI, Tableau, Seaborn, Matplotlib

Deep Learning: Tensorflow, Pytorch

ML & Predictive Modeling: Scikit-learn, Regression Models, Transfer Learning

Dev & Collaboration: Git, Linux, Firebase, REST APIs, Jupyter

EXPERIENCE:

Rutgers | Research Assistant

Feb 2025- Present

- Led controlled experimentation and research in text classification using **supervised ML** (logistic regression, SVM, and hyperparameter optimization), improving F1-score by 18%.
- Presented research findings and visual summaries to supervisors to drive model-based decision-making.
- Collaborated with **cross-functional research teams** to ensure scalable **data handling** for large-scale sentiment analysis experiments.
- Emulated **Airflow-style pipeline orchestration** using modular design for automation.

URZZA | Data Science Intern

Jan 2024 - Apr 2024

- Developed statistical anomaly detection models for EV session logs using time-series forecasting, improving detection accuracy for reliability metrics.
- Conducted root-cause analyses using **time-series forecasting** and **statistical modeling** to improve asset reliability and utilization efficiency.
- Partnered with engineering teams to **validate pipeline outputs** and streamline alerting mechanisms for anomaly detection.

ADONMO | Data Analysis Intern

May 2023 - Jul 2023

- Built regression and clustering models to optimize campaign strategies and predict performance, leading to improved ad targeting and conversion metrics.
- Automated **daily reporting pipelines** using Python scripts and **Power BI dashboards**, reducing manual reporting time by 60%.
- Collaborated with business and client teams to deliver **insightful visualizations** that drove **campaign optimization**.
- Developed automated campaign performance models in Python using regression and clustering to identify high-performing segments.

PROJECTS:

Customer 360° Analytics Platform for Banking

June 2025 - Aug 2025

- Designed and implemented a **full-stack data** engineering project to unify customer data across marketing, transactions, and credit risk sources.
- Developed robust **ETL pipelines** in Python, performed data cleaning and enrichment, and integrated multiple datasets into a centralized Customer 360° view.
- Built interactive dashboards using **Streamlit** to visualize customer demographics, behaviors, and campaign responses. Deployed RESTful APIs using **FastAPI** to expose customer insights and transaction summaries.

End-to-End Credit Risk Scoring System Using XGBoost and SHAP

May 2025 - July 2025

- Experimented with **advanced ML** techniques including transfer learning, **SHAP** explainability, and semi-supervised learning to improve classification outcomes.
- Trained and **validated XGBoost models** to predict loan defaults with an emphasis on handling severe class imbalance using `scale_pos_weight` and SMOTE, achieving a balanced AUC of 0.71.
- Generated insights that improved model-driven decision-making for loan defaults, reducing false positives by 75%.

Transfer Learning for Lung Disease classification using Efficient Net

Jan 2025 - May 2025

- Developed a **modular image preprocessing and ingestion pipeline** for X-ray data covering TB, Pneumonia, and Normal cases.
- Created data pipelines and Power BI dashboards that reduced reporting latency by 60% and informed marketing strategy adjustments, improving campaign ROI.
- Integrated **image augmentation** workflows to rebalance datasets and support downstream deep learning with **EfficientNet-B0**.
- Achieved **95% classification accuracy** with strong recall and F1 metrics (TB: 0.98, Normal: 0.92), simulating scalable health diagnostics.