

Siddharth Gosawi

+1 (930) 333-4031 | siddharth.gosawi@gmail.com | <https://www.linkedin.com/in/siddharth-gosawi> |

Portfolio- <https://sid12153.github.io/>

Data Scientist with experience in statistical analysis, predictive modeling, machine learning, NLP, and computer vision, moving models from experimentation into production. Skilled in Python, SQL, R, TensorFlow, PyTorch, Power BI, and Tableau, building pipelines, monitoring models, and delivering dashboards with measurable impact.

EDUCATION

Indiana University - Bloomington

Master of Science in Data Science; GPA: 3.92

August 2023 - May 2025

Bloomington, Indiana

Medi-Caps University

Bachelor of Technology in Computer Science and Engineering

August 2019 - July 2023

Indore, India

WORK EXPERIENCE

Data Scientist

Project 990 Inc.

June 2025 - Present

Bloomington, Indiana

- Built GPT, RoBERTa, BERT, and Llama embeddings for 120K+ mission statements; engineered ETL feature pipelines in Python, pandas, and SQL.
- Applied statistical and causal inference on grants and demographic data to identify impact drivers, improving funding targeting precision by 18%.
- Implemented KNN clustering with scikit-learn on embedding features, enabling 20K+ data-driven funder recommendations and improving segmentation accuracy by 22%.
- Designed Tableau dashboards visualizing nonprofit density, impact metrics, and coverage gaps, reducing manual reporting and ad-hoc analysis requests by 35%.

Teaching Assistant(TA)

Indiana University

August 2024 - May 2025

Bloomington, Indiana

- Taught NLP and time-series analysis to 40+ graduate students, delivering hands-on labs in Python, R, SQL, NumPy, R Shiny, transformers, NER, ARIMA, and LSTM.

Research Assistant(RA) Data Scientist

Indiana University

May 2024 - August 2024

Bloomington, Indiana

- Trained CNN models in PyTorch on elephant wrinkle images, expanding data 200% with OpenCV augmentation to achieve 85% accuracy and improve precision 18%.

Data Science Intern

Space Application Centre, ISRO

January 2023 - May 2023

Ahmedabad, India

- Developed CNN-based infrared object detection and tracking models in TensorFlow, improving detection precision 10% and reducing inference latency 25% on GPU pipelines.

PROJECTS

Finance Disclosure Intelligence | Python, FastAPI, RAG, FAISS, LLM, Docker, AWS

2025 - 2026

- Engineered strict RAG pipeline with Mistral and Llama-3 agents, achieving zero hallucinations across 3,800+ SEC chunks using FAISS evidence verification.
- Improved inference reliability to 9.9s P50 latency using Open Telemetry and Arize Phoenix, distributed tracing, and a 10% refusal rate for unsupported queries.
- Built and versioned AWS EC2 MLOps pipelines using Docker, FastAPI, and Streamlit, enabling repeatable inference experiments tracked via GitHub.

Customer Analytics and Revenue Risk Intelligence | Excel, SQL, Python, PostgreSQL, Power BI, ML

2025 - 2026

- Built SQL feature pipelines in PostgreSQL and applied RFM clustering, BG/NBD, and Gamma-Gamma models to estimate CLV across \$2M+ revenue.
- Trained time-aware XGBoost churn models with class imbalance handling and SHAP, identifying recency, frequency, and monetary value as key churn drivers.
- Delivered insights via SQL views, Excel, and Power BI dashboards, prioritizing 298 high-risk customers and quantifying \$56.9K revenue at risk.

TECHNICAL SKILLS

Machine Learning & AI: Supervised and Unsupervised Learning, Statistical Modeling, Neural Networks, CNNs, Deep Learning, Transformers, LLMs, Clustering, Classification, Regression, Random Forest, XGBoost, SVM, NLP

Programming Languages: Python, R, SQL, Java, C, C++, C#, Shell Scripting (UNIX, Linux)

ML Frameworks & Libraries: TensorFlow, PyTorch, Keras, Hugging Face, scikit-learn, Pandas, NumPy

Data Analysis & Visualization: Tableau, Power BI (DAX), Matplotlib, Seaborn, Plotly, ggplot2, Jupyter Notebook

Databases & Data Tools: PostgreSQL, MySQL, Snowflake, MongoDB, SQLite, SQL Server, Alteryx

Data Engineering & Workflow: ETL and ELT Pipelines, Data Modeling, Feature Engineering, Data Wrangling, Git, GitHub, Docker

Analytics & Experimentation: Statistical Analysis, A/B Testing, Predictive Modeling, Data Storytelling