

Siddharth Gosawi

+1 (930) 333-4031 | siddharth.gosawi@gmail.com | <https://www.linkedin.com/in/siddharth-gosawi> |
[Portfolio- https://sid12153.github.io/](https://sid12153.github.io/)

Graduate Data Scientist with experience in machine learning, NLP, and computer vision, delivering end-to-end AI solutions from research to production. Skilled in Python, SQL, R, TensorFlow, PyTorch, PowerBI & Tableau, with proven success in building models, scalable pipelines, and stakeholder dashboards that drive measurable outcomes.

EDUCATION

Indiana University - Bloomington

August 2023 - May 2025

Master of Science in Data Science; GPA: 3.92

Bloomington, Indiana

Medi-Caps University

August 2019 - July 2023

Bachelor of Technology in Computer Science and Engineering

Indore, India

WORK EXPERIENCE

Data Scientist

June 2025 - Present

Project 990 Inc.

Bloomington, Indiana

- Built **GPT**, **RoBERTa**, **BERT**, and **Llama** embeddings for **120K+** mission statements; engineered features in **Python** and **SQL**, improving KNN clustering accuracy by **22%**.
- Applied **statistical and causal inference** on grants and demographic data to estimate impact drivers and recommended funding strategies that improved targeting precision by **18%**.
- Designed interactive **Tableau** dashboards highlighting nonprofit density and civic infrastructure gaps, reducing executive manual reporting and ad-hoc analysis requests by **35%**.
- Cleaned and standardized **120K+** organizational records using **Python** and **SQL**, improving data quality by **40%** and creating modeling-ready datasets for downstream pipelines.

Teaching Assistant(TA)

August 2024 - May 2025

Indiana University

Bloomington, Indiana

- Taught **NLP** and **time-series analysis** to **40+** graduate students using **Python**, **R**, **SQL** and **transformer models**.
- Led weekly **hands-on labs** covering **NER**, **NLP models**, **R Shiny**, **ARIMA**, and **LSTM**, strengthening students' applied **ML** and **forecasting** skills.

Research Assistant(RA) Data Scientist

May 2024 - August 2024

Indiana University

Bloomington, Indiana

- Trained **CNN models** in **PyTorch** to identify elephants from wrinkle patterns with **85%** accuracy.
- Expanded dataset **200%** using **OpenCV** augmentation, improving wrinkle detection precision **18%**.

Data Science Intern

January 2023 - May 2023

Space Application Centre, ISRO

Ahmedabad, India

- Trained **CNN models** in **TensorFlow** to classify elephant wrinkle patterns, achieving **85%** identification accuracy across diverse lighting and angle variations.
- Expanded dataset size by **200%** through augmentation using **OpenCV** and **Keras**, improving wrinkle detection accuracy by **18%**.

TECHNICAL SKILLS

Machine Learning: Supervised/Unsupervised Learning, Neural Networks, CNNs, Deep Learning (TensorFlow, Keras, PyTorch), Transformers, LLMs, Generative AI, Clustering, Classification, Regression, Random Forest, XGBoost, SVM, NLP
Languages: Python, R, C, C++, C#, Java, Shell Scripting(UNIX & LINUX)

Databases and Tools: SQL, MySQL, OracleDB, PostgreSQL, MongoDB, Snowflake, Neo4j, DB2, Microsoft SQL Server, SQLite, Git, GitHub, GitLab, Docker

Data Analysis and Visualization: Pandas, NumPy, Scikit-learn, Seaborn, Matplotlib, Plotly, ggplot2, Tableau, Power BI, Streamlit, Looker Studio, ArcGIS, D3.js, Jupyter Notebook, dplyr

Others: MS Office, ETL Pipelines, Data Modeling, Statistical Analysis, Data Wrangling, Predictive Modeling, Communication, Collaboration

PROJECTS

Bridge Pledge | *Python, JSON, Excel*

September 2024 - December 2024

- Designed a **machine learning pipeline** to predict bipartisan Bridge Scores for U.S. legislators, achieving an **R²** of **0.85** by leveraging a **Random Forest Regressor** and **hyperparameter tuning**.
- Improved decision-making through a flexible scoring system with **real-time integration of new metrics** and **actionable visualizations**, including **score distributions** and **learning curves**, tailored for non-technical stakeholders

Interactive Global News Explorer | *Power BI*

March 2024 - May 2024

- Developed an interactive **Power BI dashboard**, enhancing reporting capabilities, data accessibility, and user analysis by **25%**, by creating **maps**, **timelines**, and **scatter plots** for the GDELT Project, transforming over 100 global TV channels into skimmable formats.