Siddharth Gosawi

siddharth.gosawi@gmail.com | linkedin.com/in/siddharth-gosawi | [Portfolio- sid12153.github.io/](https://sid12153.github.io/)

EDUCATION

Indiana University - Bloomington   **August 2023 – May 2025**
Master of Science in Data Science; GPA: 3.91   *Bloomington, Indiana*

Medi-Caps University   **August 2019 – July 2023**
Bachelor of Technology in Computer Science and Engineering   *Indore, India*

WORK EXPERIENCE

Teaching Assistant(TA)   **August 2024 – Present**
Indiana University   *Bloomington, Indiana*

- Delivered **NLP** and **Time Series Analysis** lectures to **40+** graduate students, covering **transformer models**, **data wrangling**, and **forecasting techniques**.
- Led collaborative sessions and simplified complex topics for diverse learners, improving comprehension and engagement.
- Conducted weekly office hours and review sessions, earning a **90%** recommendation rate based on student response distribution.

Research Assistant(RA) Data Scientist   **May 2024 – August 2024**
Indiana University   *Bloomington, Indiana*

- Developed and fine-tuned **deep learning models** to classify Asian elephants based on unique eye wrinkle patterns, achieving **85%** identification accuracy.
- Collaborated with a multidisciplinary team, including non-technical members, to integrate traditional **computer vision** methods with CNN-based models, improving wrinkle pattern recognition by **18%**.
- Expanded dataset size by **200%** through advanced **image preprocessing** and **augmentation techniques**, leading to more robust and reliable models.

Data Science Intern   **January 2023 – May 2023**
Space Application Centre, Indian Space Research Organization   *Ahmedabad, India*

- Enhanced deep learning models for object detection and tracking in infrared conditions, achieving a **10%** increase in Mean Average Precision (mAP).
- Curated a dataset of over **50,000** images and implemented a custom loss function, driving a **15%** improvement in precision and recall metrics.

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 R^2 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

Impact of Philanthropic Grants on Educational Outcomes   **September 2024 – December 2024**

- Discovered a **1.7%** increase in high school graduation rates by developing **causal inference models** with **DoWhy**, analyzing 160+ variables across grants, demographic, and economic data from 2011–2022.
- Refined **funding insights** by quantifying binary and continuous treatment effects, validating results with robust refutation tests (p-value = 1.0), and recommending targeted educational investments for under-served counties.

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.

TECHNICAL SKILLS

Languages: Python, R, C, C++, C#, Java, Shell Scripting
Databases: SQL, MySQL, PostgreSQL, Microsoft SQL Server, SQLite, MongoDB, Snowflake, Neo4j
Data Analysis and Visualization: Statistical Analysis, Data Wrangling, Power BI, Tableau, Streamlit, Plotly, ggplot, ArcGIS, Matplotlib, Looker Studio, Seaborn, D3.js, Jupyter Notebook, dplyr, scikit-learn, pandas, numpy
Machine Learning: CNN, PCA, Clustering, Neural Networks, Deep Learning (TensorFlow, Keras, PyTorch), Linear Regression, Logistic Regression, Decision Trees, Random Forest, XGBoost, SVM, Naive Bayes, K-Means, DBSCAN, Gaussian Mixture, Arimax, Sarimax
Tools/Technologies: Microsoft Office, GitLab, GitHub, Git, Docker
Others: Version Control, Problem-Solving, Communication, Collaboration, ETL Pipelines, Data Wrangling, Predictive Modeling