

Siddhant Thakur

College Station, TX | sidthakur08@tamu.edu | (206) 480-8468 | [siddhantthakur08](https://siddhantthakur08.github.io)

EDUCATION

Texas A&M University, College Station, TX

MS in Computer Science, GPA - 3.67/4.0

August 2022 - May 2024

SRM Institute of Science and Technology, Kattankulathur, India

B.Tech in Computer Science Engineering with Specialization in Big Data Analytics, GPA - 9.55/10.0

June 2018 - May 2022

University of Wisconsin, Madison, WI

Semester Abroad, GPA - 4.0/4.0

January 2021 - May 2021

RELEVANT EXPERIENCE

Data Science Intern, Junior

Airtel, Haryana, India

June 2021 - August 2021

- Facilitated addition of **15+** different **streets** for **6** major **cities** in India by developing an **ETL pipeline** for address extraction using **OSMNx** and **OpenStreetMap** data libraries.
- Extracted location information from **40000 residential addresses** employing natural language models written utilizing the **deepparse** library in Python.
- Optimized **average resolution time** across **2 states** for broadband support tickets by deciphering highly unstructured addresses thus enhancing feasibility of network expansion in rural areas.

STAT433 - Data Science with R Course Project

UW Madison, Madison, Wisconsin

January 2021 - May 2021

- Examined diversity over **263 aphid** species from **49 sites** across **14 years** throughout the Midwestern US landscape visualized using **Shiny in R** web app.
- Established association of aphid diversity count with **cumulative weather over the years** and **crop monocultures** by implementing a **Poisson Regression** model built on environmental features engineered through PCA.

Data Science & Machine Learning Domain Head

Gen-Y, Tamil Nadu, India

November 2020 - July 2021

- Mentored a batch of **12 students** as a domain head & a teaching assistant for a research club named Gen-Y.
- Assisted in subjective assessment correction of **100+** candidates in the Computer Science Department by designing **Manhattan LSTM** network in Python.
- Devised an end-to-end pipeline consisting of an **analytical panel** to help Fantasy Premier League managers to make informed decisions and choose from **560 premier league players**.

Data Science Intern, Sophomore

Bajaj Finserv Health, Remote, India

March 2020 - August 2020

- Proposed a proof of concept **recommendation engine** providing suggestions to more than **5,000** users regarding their health, based on different parameters present in lab reports.
- Developed a **logistic regression** model in Python with a **recall** score of **0.74** to estimate **at-risk** probability of a patient with Diabetes or Hypertension, while identifying important features.
- Orchestrated serverless deployment of the classification model to **10,000** users on **AWS Lambda** for sustainable scaling.

Data Science & Machine Learning Member

Gen-Y, Tamil Nadu, India

August 2019 - November 2020

- Implemented a **random forest** algorithm with **63% accuracy** to predict weekly game winners of 2019 NFL regular season.
- Organized workshops and seminars for **200 students** on applications of machine learning, computer programming and different statistical concepts.

OTHER EXPERIENCE

Senior Grader, CSCE 477/703 Cybersecurity Risk

Texas A&M University, College Station

January 2023 - Present

- Assisted **90+** students in developing an understanding of cybersecurity risk and performing risk assessments on case studies based on **NIST Privacy and Cybersecurity Frameworks**.

Grader, CSCE 402/702 Law & Policy in Cybersecurity

Texas A&M University, College Station

September 2022 - December 2022

- Facilitated evaluation of **6 essays** and **3 exams** on law and policies related to cybersecurity jobs defined by the **NIST NICE framework** in a class of **110** graduate and undergraduate students.
- Maintained a high level of organization by effectively communicating deadlines and expectations to students.

SKILLS & TOOLS

Skills - Statistics, Machine Learning (Supervised & Unsupervised Learning), Natural Language Processing, Information Retrieval, Time Series Analysis, Deep Learning, Dimensionality Reduction, Feature Engineering, Data Visualization

Tools - SQL, Python (Pandas, NumPy, Scikit-Learn, Flask, Plotly, Dash), R (Shiny, CropScapeR), Tensorflow, Keras, Pytorch, C++, C, Java, GIT, Rest APIs, HTML/CSS, Ruby, Rails, Visual Studio Code, Microsoft Excel, Tableau, Figma

ACCOMPLISHMENTS

- Achieved **1st** position in the All-India Smart India Hackathon 2020 Software edition.
- Received a **Scholarship** offer from University of Wisconsin-Madison for the Semester Abroad Program.
- Received a **CSCE Fellowship Scholarship** offer from Texas A&M University for the academic year 2022-23.