

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

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

June 2018 - May 2022

University of Wisconsin, Madison, WI

Semester Abroad (Statistics), GPA - 4.0

January 2021 - May 2021

PROFESSIONAL EXPERIENCE

Airtel X Labs, Haryana-India, Engineering Team

Data Science Intern

June 2021 - August 2021

- Enriched the pre-existing dictionary by developing an **address extraction system** using **OSMNx** and **OpenStreetMap** data package, adding **15+** different **localities** and **streets** for **6** major **cities** in India.
- Refined previous parsing architecture by building a **parsing model** for residential addresses over an open-source architecture, **deepparse**, extracting plot information from approximately **40000** residential addresses.
- Optimized **broadband feasibility** across **2 states** by deciphering highly unstructured addresses to estimate cost details.

Bajaj Finserv Health, Remote, Smart India Hackathon 2020

Data Scientist

March 2020 - August 2020

- Formulated a proof of concept **recommendation engine** providing suggestions, based on different parameters present in lab reports, to **more than 5,000** users seeking to maintain a healthy body.
- Identified important features influencing a patient's sugar level and designed a **logistic regression** model providing an inference regarding whether a patient is **at risk** of Diabetes or Hypertension with a **recall** score of **0.74**.
- Deployed classification models on **AWS Lambda** for **serverless orchestration** to scale the application for **10,000 users**.

RELEVANT EXPERIENCE

Gen-Y, Tamil Nadu-India, Technical Team

Data Science & Machine Learning Domain Head

November 2020 - July 2021

- Mentored a batch of **12 students** as a domain head & a teaching assistant for a research club named Gen-Y.
- Designed a **Manhattan LSTM** (MaLSTM) network for **automated subjective assessment correction** of **100+** candidates in the Computer Science Engineering Department.
- 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**.
- Utilized web scraping to get data from the Football Reference website and **Flask & Dash** to build a web application with necessary visualization, and deployed the web app on **Heroku**.

Data Science & Machine Learning Member

August 2019 - November 2020

- Implemented a **random forest** algorithm to predict weekly game winners of the National Football League and achieved an **accuracy** of **63%** for 2019 NFL regular season.
- Wrote **ETL** jobs to collect NFL game stats including rushing yards, passing yards and sacks per game, for home and away teams through **nflgame** API.
- Organized workshops and seminars on applications of machine learning, introducing programming and statistical concepts to around **200 freshmen and sophomores**.

UW Madison, Madison-Wisconsin, STAT 433 - Data Science with R

Semester Abroad Student

January 2021 - May 2021

- Analyzed **263 aphid species** from **49 sites** throughout the Midwestern US landscape and conveyed related information by creating an interactive web app using **Shiny in R**, displaying aphid diversity across **14 years**.
- Improved accounting of environmental factors responsible for aphid population dynamics by using unsupervised learning, namely **Varimax rotation of Principal Component Analysis** (PCA).
- Predicted association of cumulative **weather with a higher** and **crop monocultures with a lower** aphid diversity by implementing a **Poisson Regression** model.

SKILLS & TOOLS

Skills - Machine Learning (Supervised and Unsupervised Learning), Natural Language Processing, Time Series Analysis, Generalized Linear Models, Deep Learning, Data Visualization, Dimensionality Reduction, Feature Engineering, Web Scraping

Tools - Python (Pandas, NumPy, Scikit-Learn, Flask, Plotly, Dash), Jupyter Notebook, RStudio (Shiny, CropScapeR), Tensorflow, Keras, C++, C, Java, SQL, GIT, Rest APIs, Visual Studio Code, Microsoft Excel, Tableau, Figma, HTML

ACCOMPLISHMENTS

- Achieved **1st** position in the All-India Smart India Hackathon 2020 Software edition.
- Placed in **top 1%** of teams participating in the IBM Hack Challenge 2020 working on a platform automating preliminary stages of recruitment making use of Data Science and AI.
- Received a **Scholarship** offer from University of Wisconsin-Madison for the Semester Abroad Program.
- Received a **Fellowship Scholarship** offer from Texas A&M University for the academic year 2022-23.