

# Siddhant Thakur

College Station, TX | sidthakur08@tamu.edu | (206) 480-8468 | [siddhantthakur08](#)

## Education

### Texas A&M University, College Station, TX

MS in Computer Science

August 2022 - May 2024

### SRM Institute of Science and Technology, Kattankulathur, India

Bachelors of Technology 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

June 2021 - August 2021 | Data Science Intern

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

### Bajaj Finserv Health, Remote, Smart India Hackathon 2020

March 2020 - August 2020 | Data Scientist

- Formulated a proof of concept **recommendation engine** providing suggestions, based on the different parameters in the 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 the classification models on **AWS Lambda** for **serverless orchestration** in the application.

## Project Experience

### Gen-Y, Tamil Nadu-India, Data Science and Machine Learning Domain

November 2020 - July 2021 | DS & ML Head/ Teaching Assistant

August 2019 - November 2020 | DS & ML Member

- Mentored a batch of **12 students** as the domain head & the 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 Department.
- Implemented a **random forest** algorithm to predict weekly game winners of the National Football League and achieved an **accuracy** of **63%** for the **2019 NFL** regular season.
- Devised an **end-to-end pipeline** consisting of an analytical panel to help the 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**.

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

January 2021 - May 2021 | Semester Abroad Student

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

## Skills and 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

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