Siddhant Thakur

sidthakur08@tamu.edu | 206-480-8468 | LinkedIn:siddhant-thakur-08 | Github:sidthakur08

EDUCATION

Texas A&M University, College Station
M.S. in Computer Science, GPA - 3.87 — *August 2022 - May 2024*

SRM University, Chennai
B.Tech. in Computer Science Engineering, GPA - 3.87 — *June 2018 - May 2022*

University of Wisconsin, Madison
Semester Abroad, GPA - 4.00 — *January 2021 - May 2021*

SKILLS & TOOLS

Skills - Statistics, Natural Language Processing, Computer Vision, Time Series Analysis, Deep Learning, Quantitative Analysis, Dimensionality Reduction, A/B Testing
Tools - Python (Pandas, NumPy, Matplotlib, Scikit-Learn, Flask, Dash), SQL, R (Shiny, Lahman, pitchRx, CropScapeR), PySpark, Tableau, PowerBI, TensorFlow, Pytorch, C++, GIT, AWS, Docker
Courses - Machine Learning, Information Storage & Retrieval, Statistical Methods, Data Mining & Analysis, Data Visualization, Software Engineering, Data Science with R, Analysis of Algorithms

WORK EXPERIENCE

Bharti Airtel Ltd. — **Haryana, India**
Data Science Intern, Engineering — *June 2021 – August 2021*

- Enhanced **customer satisfaction score** by **30%** through implementation of an ETL pipeline targeting address resolution within Airtel's support system
- Employed natural language methods to extract addresses from OpenStreetMaps, utilizing **fuzzy matching** with **Levenshtein similarity** to enrich address dictionary
- Integrated **~40000** previously undiscovered residential addresses nationwide while collaborating with **Logistics** team

Bajaj — **Remote**
Data Science Intern, Finserv Health — *March 2020 – August 2020*

- Engineered a **full-stack recommendation system** providing suggestions to more than **5000 users** based on the users' health assessments and lab reports
- Modeled a **logistic regression** algorithm to estimate at-risk probability of a patient with Diabetes or Hypertension having recall score of **0.74**
- Deployed model in a serverless scaled environment on **AWS Lambda** for a peak capacity of **10000 user**s under the Bajaj health plan

PERSONAL PROJECTS

Sports Analytics — *August 2019 - Present*

- Applied **computer vision** to track player and ball movement for Arsenal's 2022/23 Premier League season using **YOLOv7**, enhancing pass analysis through **Kernel Density Estimation**
- Developed a sequence model using **LSTM** to generate pitch sequences optimizing probability of a strikeout, enhancing in-game strategic decision-making
- Formulated measure of catcher framing using **mixed effect modeling**, incorporating pitching data from PITCHf/x and Statcast tools in **R**
- Devised a **random forest** algorithm with **64.1%** accuracy for **2019 NFL** regular season and published articles on Medium summarizing weekly game winners

Group Recommender System — *December 2022 – May 2023*

- Designed a group recommender system utilizing LinkedIn profiles and Slack bios for optimal **pairings of 3-4** students
- Leveraged advanced techniques including **GloVe** embeddings, **TF-IDF** scores, and Hugging Face **Zero-Shot transformer** to calculate user-pair similarities and category-based scores
- Achieved **63% preference rate** for the model-generated optimal pairings over randomly selected pairings through independent observer evaluations

ESPN March Madness Competition — *March 2021 - March 2023*

- Attained **2nd** position in 2022 ESPN Bracketology coding an **XGBoost** model in **Python**, leveraging key statistics like FG%, ELO and +/- score, with brier score of **0.26**
- Optimized 2022 model incorporating features like Rebound Differential and Assist to Turnover Ratio and an ensemble of models fine-tuned via **GridSearchCV** resulting in **23%** improvement in brier score

ACCOMPLISHMENTS

- Received **CSCE Fellowship Scholarship** offer from Texas A&M University for both academic years 2022 & 2023
- Received **Scholarship** offer from University of Wisconsin-Madison for Semester Abroad Program
- Achieved **1st** position in All-India Smart India Hackathon 2020 Software Edition