

Siddhant Thakur

Student



EXPERIENCE

Data Science and Machine Learning Lead

Gen-y

Aug 2019 - Present

- Wrote Python scripts for Exam Correction Automation to process data and generate features like common word ratio, difference between two sentence vectors, etc.
- Selected to attend and work with the GDEs at the Google ExploreML Bootcamp held at the Google campus in Hyderabad.
- Guided and Lead a team of 12 people under the Machine Learning Domain.

Data Scientist

Smart India Hackathon 2020 (Software) - All India Winner

- Worked as a team to solve the Bajaj Finserv problem statement of Digitization of Lab Reports.
- Developed a proof of concept recommendation engine providing suggestions to the user on maintaining a healthy body based on the values of parameters fed in.
- Analyzed the important features influencing a patient's sugar level and designed a classification model providing inference regarding whether a patient is in risk of Diabetes Mellitus or not.
- Deployed these models on AWS Lambda for serverless orchestration.

OTHER PROJECTS

Soccer Analytics

- Created a website and Github repositories consisting of Soccer analytics in Python & R over Event & Tracking data provided by Statsbomb, Metrica & Wyscout.
- Used statistics like xG, npxG, xA and plots like shot plots and pass maps to generate insights on match data.
- Developed a visualization panel using Flask, Plotly & Dash to provide a tool for all Fantasy Premier League Managers to select players based on stats.

Project Darwin by Team Blueprint

- Generated a prototype to automate personality insights for the preliminary stages of recruitment.
- Utilized the Twitter API in order to fetch the tweets from the interviewee's twitter handle and pre-process it further to be used in IBM's PersonalityInsights model.
- The above feature helps the employer to search and select an ideal candidate based on their company profile.

National Football League Game Prediction Model

- Built a Random Forest Classifier in Python to predict weekly game winners of the NFL and got an accuracy of about 63% for the 2019 NFL regular season.
- Wrote ETL jobs to collect stats like rushing yards, passing yards, sacks per game etc., for home and away teams using the nflgame API

siddhantthakur08@gmail.com

+918368103652

[siddhant-thakur-08](#)

[sidthakur08](#)

EDUCATION

SRM Institute of Science and Technology

CGPA - 9.52

Bachelor in Computer Science Engineering

2018 - Present

Apeejay School, Noida

Percentage - 90.8%

Class 12th in Physics, Chemistry and Maths with Comp Sci

2016 - 2018

SKILLS AND TOOLS

Programming: Python, R, C++, C, SQL, Flask, Dash, Shiny, plotly, REST APIs

Modeling: Machine Learning (Regression, Classification and Clustering), Feature Selection, Data Visualization, Feature Engineering, Dimensionality Reduction, Natural Language Processing

Software: Jupyter Notebook, RStudio, Microsoft Excel, Visual Studio Code, Tableau

ADDITIONAL COURSES

University of Wisconsin - Madison

2021

- STAT 601 - Statistical Methods I
- STAT 433 - Data Science with R