Rahul Thaker

rthaker3@illinois.edu | +1(447)-902-1180 | Champaign, IL | Rahulthaker.com

EDUCATION

University of Illinois at Urbana-Champaign *Aug 2022 - May 2024*
Master of Science in Information Management | **GPA: 4.0/4.0**
Coursework: Data, Statistical Models; Data Warehousing, Business Intelligence; Methods of data science

K.J. Somaiya Institute of Engineering and Information Technology *Aug 2018 - June 2022*
Bachelor of Technology in Computer Engineering | **GPA: 9.25/10**
Coursework: Applied Data Science; Data Structures; Database Management Systems; Machine Learning

EXPERIENCE

Caterpillar Inc.| Champaign, Illinois| Data Science intern *November 2022-Present*
- Built a random forest regressor to predict demand for CAT dealers, dealer stores, and distribution centers with 800K+ rows of data for each month to minimize problems of low inventory turns and dead stock using python
- Reduced runtime of machine learning model and segmented code into functions for better reusability and automation, faster debugging, promoting collaboration, low-risk updates if required

National Center for Supercomputing Applications |UIUC, Illinois **|**Graduate Research Student *Aug 2022-Dec 2022*
- Implementing and fine-tuning Natural language processing models, text-DaVinci-002 API from Open AI and Bert, Bart hugging face models for information retrieval and QA generation with a team of 11 students
- Generating custom Dataset using prompt engineering and performing data pre-processing and cleaning
- Performing filtering and ranking of response objects by implementing GPT classification models on HAL cluster

Teach For India |Mumbai, India **|** Data Engineer Intern| IEEE paper *Aug 2021- May 2022*
- Streamlined data collection process by directly connecting the Salesforce database to on-prem analytics solution which eliminated the need for Excel sheets, saving 20 hours of equivalent weekly manual effort
- Automated the process of analysis, visualization and developing reports and published a paper titled 'Statistical analysis, ETL automation and Intelligence Reporting' on IEEE DOI: 10.1109/CONIT55038.2022.9847751
- Derived new insights and statistical co-relations, improving the accuracy of selection model by 30%

Pingzee, ZeeQ| Noida, India **|** Data analyst Intern *Sept 2020 – Jan 2021*
- Developed a real-time reporting system that can recommend changes based on behavior of infrastructure by leveraging analytics libraries Pandas, NumPy and visuals utilizing scripting layer of Matplotlib
- Performed Time-series data analytics and forecasting of backup metadata using SARIMA model with MAPE of 5%
- Automated the process of creating highlights, alerts and notifying clients by incorporating SMTP.

PROJECTS

Neutralit – Twitter Sentiment Analysis
- Conducted Twitter sentiment analysis using the Twitter API and Python NLTK libraries with 83% accuracy
- Created visuals using Matplotlib and Seaborn and producing reports for trending topics of the day
- Built a website and deployed it on Heroku by applying web development frameworks Django, CSS, JS and HTML

Marketing Analysis – Saint Louis University
- Processed and gathered data from social media including attributes like clicks, costs per click, reach etc.
- Saved 100 dollars monthly of a small-scale company called Globalshala by analyzing more than 20 different marketing campaigns and performing Exploratory and explanatory data analysis
- Led a group of 5 members and presented research and findings to shareholders

Result analysis – Data Analytics Pipeline for Automation
- Developed a Web API based on Flask to analyze semester results, provides customized visuals and reports based on input parameters by directly utilizing the PDF file given by Mumbai university
- Deployed by K.J. Somaiya Institute of Engineering and Information Technology faculty to write SQL queries, produce graphs and create reports, reducing faculty effort by 80%

SKILLS

Languages: Python, SQL, R, C, C++, HTML, Javascript
Frameworks: Pandas, Numpy, Matplotlib, Flask, OpenCV, TensorFlow, PyTorch, Keras, Scikit-learn, MongoDB
Tools: Tableau, Git, AWS, Heroku, Postman, Google Data Studio, Jupyter Notebook, Excel, Talend, MS Office
Certifications: IBM Data Science Professional, IBM Artificial Intelligence Engineer, AWS Technical essentials

EXTRA CURRICULAR ACTIVITIES

- Grader for the 'Foundations of Data curation' course offered by Computer Science department at UIUC
- Chief Executive Officer of American Chemical Society at KJSIEIT 2021-22
- Organizing head of Computer Society of India at KJSIEIT 2020-21