Rahul Thaker

rthaker3@illinois.edu | +1(447)-902-1180 | https://rahulthaker.com/ | https://www.linkedin.com/in/rahul-thaker/

EDUCATION

University of Illinois at Urbana-Champaign   *Expected 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

SKILLS

Languages: Python, SQL, R, C++|Libraries - Pandas, NumPy, boto3, Matplotlib, Scikit-Learn, PyTorch, TensorFlow
Analytics Tools: Tableau (Portfolio), Power BI, Google Analytics, Excel, Dataiku, AWS QuickSight, Streamlit, Plotly
Data Engineering: MongoDB, Django, PostgreSQL, Flask, Snowflake, AWS DynamoDB, Apache Airflow, MLOps
Certifications: IBM Data Science Professional, IBM Artificial Intelligence Engineer, AWS Technical Essentials

EXPERIENCE

Caterpillar Inc.| *Digital & Analytics intern*   *Nov 2022 – Aug 2023*
- Implemented random forest and H2O regressor with 89% accuracy to forecast demand of spare parts for up to 3 months at Caterpillar stores and distribution centers having 800K+ rows of data
- Reduced runtime of Machine learning model by 87% for distribution centers and by 50% for dealer stores
- Led initiative to automate the pipeline by migrating data preparation, integration (with Snowflake) and machine learning to cloud (Dataiku) and implementing AutoML
- Performed feature engineering and streamlined data collection, preventing inaccurate predictions; anomalies

National Center for Supercomputing Applications | *Graduate Research Student*   *Aug 2022 – Dec 2022*
- Deployed an AI-based teaching assistant chatbot via a multimodal question-answering dialogue system leveraging large language models such as GPT-3, OPT 175B and FLAN-T5 in HuggingFace
- Fine-tuned models on custom dataset generated using prompt engineering and data generation through GPT-3
- Performed filtering and ranking of response objects by implementing GPT classification models on NCSA servers

Teach For 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 eliminating the need for Excel sheets, saving 20 hours of equivalent weekly manual effort
- Automated the process of analysis, visualization and developing reports. Published a paper titled 'Statistical analysis, ETL automation and Intelligence Reporting' on IEEE DOI: 10.1109/CONIT55038.2022.9847751
- Derived new insights and statistical correlations, improving accuracy of selection model by 30%

Semantics Solution| *AWS Intern*   *July 2021 – Aug 2021*
- Serviced client needs and implemented IAM roles and policies using JSON notation
- Created and configured VPCs, Subnets, ACLs and route tables based on requirements of the solution
- Deployed solutions for 4 projects based on right set of EC2 compute, storage and database services

Pingzee, ZeeQ | *Data analyst Intern*   *Sept 2020 – Jan 2021*
- Developed a real-time reporting system for recommending changes based on behavior of infrastructure by leveraging analytics libraries Pandas, NumPy and visuals utilizing scripting layer of Matplotlib
- Executed Time-series data analytics and forecasting of backup metadata using SARIMA model with only 5% error
- Instituted the process of creating highlights, alerts and notifying clients by incorporating SMTP

REAL WORLD PROJECTS

Marketing Analysis – Saint Louis University
- Conducted growth analytics on data from social media including attributes like clicks, costs per click, reach etc.
- Saved ~1000 dollars 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 communicated actionable insights 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 build reports, reducing faculty effort by 80%

EXTRA CURRICULAR ACTIVITIES

- Grader for the 'Foundations of Data Curation' course offered by Computer Science department at UIUC
- Organizing head of Computer Society of India at KJSIEIT 2020-21