

Rahul Thaker

Thakerrahul16@gmail.com | +1(447)-902-1180 | <https://rahulthaker.com/> | <https://www.linkedin.com/in/rahul-thaker/>

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; Information modelling, Business Intelligence; Methods of Data Science

K.J. Somaiya Institute of Engineering and Information Technology

Aug 2018 - Jun 2022

Bachelor of Technology in Computer Engineering | **GPA: 9.25/10**

Coursework: Applied Data Science; Object Oriented Programming; Distributed 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.

Lens Customer | Data Scientist

Nov 2023 – Present

- Engineered a comprehensive scoring system integrating decision tree models to accurately predict and recommend the most suitable sales channels and product recommendations for over 1 million customers
- Performed statistical analysis, generated/collected 200+ behavior-based attributes and developed K-Means clustering models for behavior-based customer segmentation for over 10 million customers globally
- Delivered actionable insights for key digital applications and integrated data into Power BI dashboards for real-time visibility into customer behaviors and trends, aiding strategic decision-making using 5 KPIs

Living Machine Logistics team | Data Science intern

Nov 2022 – Aug 2023

- Implemented machine learning models with 89% accuracy to forecast demand of spare parts for up to 3 months at Caterpillar stores and distribution centers
- Led initiative to automate the pipeline by migrating data preparation, integration (with Snowflake) and machine learning to cloud (Dataiku) and implementing AutoML, reducing runtime

National Center for Supercomputing Applications | AI engineer

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 on GPT-3
- Performed filtering and ranking of response objects resulting in a 27% increase in relevant response detection

Teach For India | Data Engineer Intern

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](https://doi.org/10.1109/CONIT55038.2022.9847751)
- Derived new insights and statistical correlations, improving accuracy of selection model by 30%

Pingzee, ZeeQ | Data analyst Intern

Sep 2020 – Jan 2021

- Developed a real-time reporting system for recommending changes based on machine metadata leading to 37% reduction in response time for critical infrastructure changes
- Executed Time-series data analytics and forecasting of backups using SARIMA model with 95% accuracy
- 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
- 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 Flask based Web API to analyze results, providing customized visuals and reports based on input parameters by directly utilizing the PDF file given by the university, generating reports for over 10000 students
- Deployed by K.J. Somaiya Institute of Engineering and Information Technology faculty to write SQL queries, produce graphs and build reports, resulting in a 60% acceleration in report generation