ROHIT KAUSHIK

(+1) 217-953-1352 | kaushikrohit2024@gmail.com | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Results-driven Data Analyst with expertise in end-to-end data solutions. Proficient in **Python, R, SQL, and Machine Learning** for comprehensive analysis and modeling. Applied advanced techniques like clustering, **StyleGAN, and GenAI models** with database design, including MySQL and MongoDB. Experienced in **MLOps**, deploying **CI/CD** pipelines, and cloud solutions such as **Databricks and AWS**. Strong command of Big Data tools such as **Hadoop, Spark, and Hive**.

SKILLS & CERTIFICATIONS

- **Programming:** Python, R, MATLAB, Hadoop, Spark, Hive, Java, SQL, PSQL, Shell Scripting
- **Machine Learning and Data Science:** NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, MLlib, PyTorch, OpenCV, Keras, TensorFlow, SciPy, ggplot2, Plotly, Neural Networks, Data Wrangling, Data Warehouse, Data Visualization, A/B Testing
- **Technologies:** Jupyter, Google Colab, Visual Studio Code, Microsoft Excel, Microsoft Access, Tableau, Power BI, AWS S3, AWS Redshift, AWS EC2, AWS CloudFront, AWS DynamoDB, Jenkins, Snowflake, Docker, Git, GitHub, SAS, SPSS, Alteryx
- **Certifications:** Machine Learning with Python: IBM, Data Analysis Using Python: IBM, Evidence and Data Collection for Problem Solving: University of Leeds and Institute of Coding.

EXPERIENCE

Motor Vehicle Data Linkage Graduate Intern, Illinois Department of Public Health (IDPH), Illinois, USA *Feb 2023 - May 2024*

- Strengthened analysis of healthcare (IDPH) and transportation (IDOT) datasets with **200+ attributes** using Python **(Pandas, Seaborn, Scikit-learn, TensorFlow, and Keras)**. Improved regional and state-wide safety by 40% with preemptive safety measures based on machine learning models.
- Collaborated with cross-functional teams to develop **30+ interactive dashboards using Tableau** and Statistical Techniques, such as **descriptive statistics, correlation, and regression**, increasing usability and effectiveness by 23%.
- Led data integration using SAS, enhancing precision by 15%. **Automated routine tasks**, such as data ingestion, storage, and processing, boosting team productivity by **25% to reduce fatalities**, injuries, and healthcare costs related to accidents.

Consultant, Birlasoft Limited, Noida, India *Jan 2021 - July 2022*

- Spearheaded a team of **3 data analysts**, harnessing the power of Python and Machine Learning algorithms such as **Regression, Decision Trees, and Neural Networks** in deploying over **15+ ETL** pipelines and **20+ ML** pipelines guided by the principles of **GenAI**.
- Pioneered the development of a global, real-time reporting system using **Tableau** providing comprehensive insights to stakeholders, catalyzing a 30% improvement in organizational decision-making.
- Optimized SQL queries and orchestrated a large-scale migration from a legacy system to **AWS** enhancing system performance. **Performed A/B testing**, reduce data segmentation and analysis delays by 13%.

Web Scraping Lead, Indiamart Intermesh Limited, Noida, India *Aug 2020 - Oct 2020*

- Utilized **R (dplyr, tidyr, ggplot2) and Excel** to analyze large datasets, applying statistical modeling to enhance data processing efficiency. Processed over 500,000 data points daily, implementing data validation checks to ensure high data accuracy and integrity.
- Established robust **CI/CD pipelines** using Jenkins for continuous integration, **Git for version control**, and Selenium for automated testing improving code quality and reducing deployment failures. Introduced **MLOps** practices to streamline the machine learning workflow, from data preparation and model training to deployment.
- Championed data-driven strategies, increasing decision-making and reducing system bugs by 30%. Enhancing data processing speed by **40%.**

EDUCATION

University of Illinois, *Springfield, USA*
Master of Science in Data Analytics, GPA 3.9/4.0 *Aug 2022 - May 2024*
Coursework: Advanced Statistical Methods, Big Data Analytics, Machine Learning, Deep Learning, Data Visualization, Database Systems, Data Mining, Statistical Computing.
Award/Scholarships: Graduate Public Service Internship Program (top 3%), Graduate honors award in Data Analytics

Amity University, *Noida, India* *July 2016 - Aug 2020*
Bachelors in Technology, Computer Science and Engineering, GPA 8.1/10.0

PROJECTS

Real-Time Sign Language Recognition using Convolutional Neural Networks *Jan 2024*
Engineered a real-time sign language recognition system utilizing advanced CNNs (LeNet, AlexNet, VGG, ResNet) in Python with Keras and TensorFlow. Attained an impressive 97.52% accuracy on a dataset featuring 78,000 grayscale images, leveraging a meticulously fine-tuned ResNet-50 model. Employed rigorous data preprocessing, incorporating PIL and augmentation techniques. The resulting user-friendly Streamlit web application stands as a technological stride, fostering enhanced communication for the deaf and hard-of-hearing community with a focus on inclusivity.

Predicting Customer Behavior for British Airways with Data Scraping and Machine Learning *June 2023*
Utilized Selenium and Beautiful Soup to meticulously scrape and analyze 10,000 diverse customer feedback from British Airways, emphasizing the critical role of staff service in satisfaction. Achieved a commendable 90% accuracy through implementation of Random Forest and Logistic Regression models, addressing bias-variance challenges. Deployment on the Google Cloud Platform ensured scalability and efficiency. Results highlight strong predictive performance, especially in precision, recall, F1-score, and ROC AUC metrics, showcasing the model's nuanced understanding of sentiments and factors influencing customer ratings.

Visual Question Answering System with Deep Learning *Dec 2022*
Architected a Visual Question Answering (VQA) system, integrating sophisticated deep learning paradigms BERT+VGG19 and LSTM+VGGNet in Python with Keras. Utilized the MSCOCO training dataset, housing 443,757 intricately annotated images spanning 7 attributes. Attained an exemplary 88.24% accuracy on the test set, bolstered by meticulous evaluation metrics encompassing F1-score, and BLEU score. Illuminated the model's dynamic evolution, performance across diverse question types, and intricate attention mechanisms through insightful visualizations.

PUBLICATIONS & LEADERSHIP EXPERIENCE

- Cardiovascular Disease Prediction Using Machine Learning Techniques. (2023). *AIP Conference Proceedings.*
- Postmortem Concentrations: Distributed Privacy-Preserving Blockchain Authentication Framework in Cloud Forensics. (2023). In *CRC Press* (pp. 161–184).
- Dynamics of EHR in M-Healthcare Application. (2023). In *Wiley-Scrivener Publishing*, Chapter 11, *Wiley*.
- Sensory Perception of Haptic Rendering in Surgical Simulation. (2023), 14th International Conference (ICCCNT) (pp. 1-7). *IEEE*.
- YouTube Channel: Guiding students on “Studying in the USA” with valuable insights.
- IEEE ADGITM Speaker (2021): Delivered impactful sessions on advanced data science and ML techniques.