

# RISHABH KUMAR KANDOI

## Data Engineer / Data Science Enthusiast

Master's in Data Science | Bachelor's in Computer Science & Engineering

Redmond, Washington, USA - 98052 | [+1\(585\) 410-8739](tel:+15854108739) | [rishabhk8@gmail.com](mailto:rishabhk8@gmail.com) | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

### CAREER OBJECTIVE

Results-oriented Data Engineer with 4+ years of experience specializing in data processing, ETL, and data analytics. Proficient in Python, SQL, AWS, and ML. Seeking opportunities to leverage expertise in drive data-driven decision-making and enhancing organizational efficiency.

### WORK EXPERIENCE

- **Microsoft**, Redmond, WA, USA | **Data Engineer** | Contract Mar 2024 – Present
  - Managing M365 sales data in GDC, enabling sales team to take informed decisions for customer retention & conversion optimization.
  - Migration to Microsoft Fabric from OnPrem & Azure SQL Server, ADF Pipelines, ADLS Gen2 & PowerBI Reports.
- **People Tech Group**, Redmond, WA, USA | **Data Engineer** Dec 2023 – Present
  - Performed PoC for Microsoft, to build Azure Pipeline for comparing PowerBI Reports.
- **Freelancer**, NY, USA | **Data Engineer / Data Scientist** May 2023 – Nov 2023
  - Employed Python libraries (NumPy, Pandas, Seaborn, Matplotlib) for data cleaning, scaling, and engineering, enhancing data quality.
  - Optimized SQL procedures, achieving a 20% reduction in database update time, improving data processing efficiency.
  - Collaborated on NoSQL solutions like MongoDB, enabling rapid data retrieval and analysis for better insights.
- **Paytm Payments Bank**, Noida, India | **Senior Software Engineer - Data** Sep 2021 – Aug 2022
  - Implemented Java and SQL query optimizations, handling 20M+ daily transactions with intuitive error handling.
  - Utilized AWS Glue ETL to boost data analytic throughput by 25% through S3 and Redshift integration.
  - Led Money Transfer team projects, simplifying user interactions, and achieving a remarkable 30% increase in customer retention.
  - Enhanced collect transaction notifications, reducing errors by 25% and improving accuracy by 20%.
- **BigBasket**, Bangalore, India | **Software Engineer - Data** Jan 2019 – Aug 2021
  - Managed end-to-end projects, optimizing Docker, Kubernetes, and Helm for production releases, improving project efficiency by 30%.
  - Introduced real-time data processing solutions, reducing data latency by 25%, enabling informed decision-making.
  - Spearheaded a 500% speed increase in cron job execution, saving 20 hours weekly and enhancing operational efficiency.

### EDUCATION

- University of Rochester, NY, USA – 3.7/4.0, Master of Science, Data Science Aug 2022 – May 2023
- NIIT University, India – 3.9/4.0, 1st Rank Holder, Bachelor of Science, Computer Science Aug 2015 – Jul 2019

### LEADERSHIP AND TECHNICAL SKILLS

- Managed and mentored a team of interns, fostering their professional growth and contributing to successful project completion.
- Proficient in programming languages: Python, R, SQL, and experienced with C/C++, Java.
- Strong knowledge of database systems including MySQL, PostgreSQL, MongoDB, Elasticsearch.
- Expertise in ETL & Infrastructure tools: Databricks, Snowflake, Airflow, Kafka, Docker, Kubernetes.
- Skilled in the Big Data ecosystem: Hadoop, MapReduce, Hive, Apache Spark, Pig.
- Cloud platform experience: AWS and Azure (ADF, Synapse Analytics, Fabric, CoPilot).
- Expertise in data visualization tools: Tableau, PowerBI (PowerApps, DAX), SSRS, Plotly, Matplotlib, Excel
- Statistical modeling proficiency: A/B Testing, Generalized Linear Models, Clustering, Time Series Forecasting, Association Rules and Pattern Mining, Ensemble Models, Neural Network Models, Deep Learning
- Data Science / Machine Learning packages: SciPy, Scikit, TensorFlow, Keras, PyTorch.
- Management tools: Github, JIRA, Grafana, Kibana, NewRelic, Confluence, Datadog.

### PROJECTS

- **Trauma Detection (Healthcare)** – Achieved under 5% FNR and 25% FPR for classifying Trauma level of the patients, with 90% accuracy, as opposed to metrics for manual classification (65% FNR, 16% FPR, 72% accuracy), by utilizing EDA, sampling, and ML modelling techniques (Ensemble Model). Performed statistical tests to show demographic based influence.
- **Spam Mass Detection (Big Data)** – Detection of spam pages in search engines using Page Rank algorithm, exploiting MapReduce Framework in Hadoop (HDFS) for parallel processing of the data.
- **Crime Rate Prediction (Research Project)** – Utilizing Twitter, demographics & Google Searches related to mental health, multiple factors showed high impact on crime rate, enabling prediction of crime across multiple cities in the USA, with MSE of ~0.04, better than any existing studies at this scale. Identified crime categories with high count but needs awareness to raise voice against.
- **Group Chat Text Segmentation Using Topic Modeling (NLP)** – Segmented Slack Dataset using hierarchical Bayesian unsupervised topic segmentation model. Process involved Data cleaning, Tokenization, identifying reply objects, calculating similarity distance and naming the topic. Most prominent use-cases are decision auditing and dynamic responsibility allocation.

### CERTIFICATIONS

- Microsoft Certified: Fabric Analytics Engineer Associate – Microsoft Fabric Delta Lake, Pipelines, PowerBI, Security & many more.
- Microsoft Learn Cloud Skills Challenge at Build 2024 – GenAI with Microsoft Azure AI services like OpenAI, AI Search, Vision & many more.