RISHABH KUMAR KANDOI

Data Engineer

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

Rochester, New York, USA - 14623 | +1(401) 213-9091 | rishabh.k@myjobmails.com | LinkedIn | GitHub

CAREER OBJECTIVE

Results-oriented Data Engineer with 3.5 years of experience specializing in data processing, ETL, and data analytics. Proficient in Python, SQL, AWS, and ML. Seeking opportunities to leverage my expertise to drive data-driven decision-making and enhance organizational efficiency.

WORK EXPERIENCE

- **Charles Schwab,** NY**,** USA | **Data Engineer** | Contract *Jan 2023 – Current*
  - Designed and implemented Python-based regular expression projects, optimizing data extraction processes, resulting in a 20% reduction in processing time.
  - Developed SSRS reports and SSIS integration packages, enhancing data visualization and streamlining data workflows, leading to a 15% improvement in data accessibility.
  - Utilized AWS Glue ETL for loading application call logs to S3 and Redshift, increasing data analytic throughput by 25%.
  - Engineered and optimized SQL procedures and triggers, achieving a remarkable 20% reduction in database update time, thereby enhancing data processing efficiency and ensuring timely updates, which significantly contributed to the team's overall performance.
  - Collaborated in the design and deployment of NoSQL implementations like MongoDB, enabling faster data retrieval and analysis.
- **Vivma Software Inc**, India | **Data Engineer** *Aug 2019 – Jul 2022*
  - Oversaw the complete Software Development Life Cycle (SDLC) following the Waterfall methodology.
  - Developed complex MapReduce programs in Hive, Pig, and Python for Data Analysis, reducing processing time by 30%.
  - Implemented Microsoft Power BI Power Query to extract and format external data, improving data quality and reducing errors.
  - Led multiple projects, providing mentorship and project guidance, fostering a culture of teamwork and knowledge sharing.
  - Conducted data cleaning, feature scaling, and feature engineering using Python libraries (NumPy, Pandas, Seaborn, Matplotlib).
  - Prepared and uploaded SSRS reports, managed databases, and optimized database performance.

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, Azure, GCP.
- Expertise in data visualization tools: Tableau, PowerBI, 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.
- Strong problem-solving and analytical skills.

PROJECTS

- **Trauma Detection** – 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** – Detection of spam pages in search engines using Page Rank algorithm, exploiting MapReduce Framework in Hadoop (HDFS) for parallel processing of the big 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.

RELATED COURSES

- Time Series Analysis
- Intro to Statistical Machine Learning
- Data Mining
- Tools for Data Science
- Data at Scale
- Pricing Analytics