ROHIT DAS – DATA SCIENTIST

CA, USA | LinkedIn | GitHub | +1 (562) 939-4190 | rohitd4956@gmail.com

Data professional with 3+ years of experience across Data Science, Data Analysis, and Data Engineering. Proven ability to leverage advanced analytics and machine learning to solve complex business problems and drive strategic decisions. Proficient in developing machine learning models, managing ETL pipelines, and delivering actionable insights through data visualization. Skilled in Python, SQL, AWS, Google Cloud, and big data technologies such as Hadoop and Spark.

WORK EXPERIENCE

Kofax   **United States**
Data Scientist   *Jan 2024 – Current*

- Developed and deployed predictive models using Python and Scikit-learn that identified at-risk customers, leading to a 15% increase in retention through targeted marketing strategies.
- Built and optimized ETL pipelines using Apache Airflow and AWS (S3, Redshift), which reduced data processing time by 30% and improved data freshness for analytical reports.
- Analyzed large datasets with SQL and PostgreSQL, providing insights that resulted in a 10% increase in operational efficiency by identifying bottlenecks in supply chain processes.
- Managed big data infrastructure using Hadoop and Apache Spark, enabling real-time analytics on high-volume datasets, which supported timely decision-making for marketing campaigns.
- Created interactive dashboards using Tableau and Power BI, allowing stakeholders to visualize key metrics and trends, ultimately improving strategic planning sessions.
- Implemented machine learning models for time series forecasting with TensorFlow, achieving a 20% improvement in sales forecasting accuracy, which enhanced inventory management strategies.
- Collaborated with cross-functional teams to automate data flows, streamlining data access and reporting processes, which reduced manual effort by 40%.

Magna Infotech   **India**
Data Engineer   *May 2020 – July 2022*

- Conducted in-depth data analysis using Python (Pandas, NumPy) and SQL, identifying critical business metrics that supported a 12% increase in revenue through data-driven decision-making.
- Designed data pipelines using Apache NiFi and Talend, facilitating seamless data integration across various platforms and ensuring data accuracy and consistency.
- Developed and maintained data warehouse solutions in Google Cloud (BigQuery), enhancing data accessibility and reducing query performance time by 25%, resulting in faster report generation.
- Performed data cleaning and feature engineering to ensure high-quality datasets for machine learning models, improving model performance by 15%.
- Created visualizations using Tableau and Matplotlib, simplifying complex data into actionable insights for non-technical stakeholders, improving team communication and project alignment.
- Assisted in implementing A/B testing and hypothesis testing with statistical techniques, leading to a 12% increase in customer engagement through targeted strategies based on data insights.
- Collaborated with engineering teams to design scalable database architectures, optimizing query performance on NoSQL systems and improving application response times by 30%.

SKILLS

- **Programming:** Python, R, SQL
- **Data Manipulation:** Pandas, NumPy, SAS
- **Machine Learning:** Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost
- **Generative AI**: OpenAI GPT, LLaMa, Stable Diffusion, LangChain, LLaMaIndex, Fine-tuning
- **Deep Learning:** Neural Networks, CNN, RNN, LSTM, GANs, NLP (NLTK, SpaCy, BERT)
- **Statistical Analysis:** Hypothesis Testing, A/B Testing, Time Series Analysis
- **Big Data:** Hadoop, Apache Spark
- **Cloud Technologies:** AWS (S3, Redshift), Google Cloud (BigQuery)
- **Data Visualization:** Tableau, Power BI, Matplotlib, Seaborn
- **ETL & Data Pipelines:** Apache Airflow, Apache NiFi, Talend
- **Database Management:** SQL, PostgreSQL, MongoDB

EDUCATION

California State University   Long Beach, CA
Master of Science: Computer Science   **Aug 2022 – May 2024**

University of Mumbai   Mumbai, India
Bachelor of Engineering in Information Technology   **Aug 2018 – May 2022**

CERTIFICATION
Microsoft Certified: Azure Data Science Associates, Azure AI Fundamentals, Azure Data Fundamentals

PROJECTS

Generative 3D model (text-to-3D model)
- Employed Stable Zero and Convolution Reconstruction Model to generate 3D assets for game development.
- Enhanced the functionality of the Stable Zero model by fine-tuning it for text-to-3D asset generation.

Automated Career Guidance System
- Developed an Automated Career Guidance System using NLP to analyze essays and suggest career paths.
- Implemented textCNN, RNN, XGBoost, and Random Forest in the system's algorithmic framework, achieving 80-90% accuracy in career path recommendations.

AI-Based ChatBot for Providing Medical Information
- Designed an AI-based ChatBot in Flutter to diagnose over 150 diseases from symptoms.
- Utilized CNN for accurate classification and developed a user-friendly platform for accessible medical guidance, providing AI-driven home remedy.