RAKESH DAS

+1 (647) 821 6228 ⬦ Guelph, On ⬦ rdas05@uoguelph.ca ⬦ linkedin.com/in/rakesh-das-data/

EDUCATION

University of Guelph, Ca   Sep 2024 - Present
Master of Data Science

Odisha University of Technology and Research, In   Sep 2016 - Sep 2020
Bachelor of Technology - Electrical Engineering

Relevant Academic Coursework:
Mathematical & statistical Concepts:
Elementary Functions, Sequences and Series, Differential Calculus, Integral Calculus, Polynomial Interpolation, Numerical Differentiation and Integration, Systems of Linear Equations and Gaussian Elimination, Linear Independence, Basis, and Dimension, Matrix Operations, Eigenvalues and Eigenvectors, Diagonalization, Orthogonalization and Projections, Linear Transformations. Probability, Introductory Statistics, Regression, Correlation, Statistical Learning.

SUMMARY OF SKILLS

- Skilled in data analysis and visualization using R, with experience in tools such as ggplot2, dplyr, and tidyr for manipulating and presenting complex datasets
- Hands-on experience in statistical and machine learning techniques using Python libraries such as Scikit-learn, TensorFlow, NumPy, and Matplotlib.
- Proficient in GCP services: Dataflow,DataProc,BigQuery, CloudSQL, Airflow,Pub/Sub, Transfer services, & Cloud Storage.
- Strong knowledge of Big Data technologies and programming languages including Python, Hadoop, MapReduce, Spark, Kafka, Databricks, PySpark, Pandas, and Shell scripting.
- Experienced with SQL databases, ETLs, and data warehousing, including MySQL, PostgreSQL, Netezza, and Oracle.

EXPERIENCE

Deloitte, Consultant (GCP Data Engineer), In   May 2024 – Sep 2024
Guided clients in utilizing GCP for data ingestion, storage, processing, and visualization. Designed and implemented scalable data pipelines and resolved platform issues for large-scale systems.

Infosys Limited, Digital Specialist Engineer - Data & Analytics, In   June 2021–May 2024
Developed scalable data pipelines using PySpark and GCP (Dataproc, BigQuery). Led cloud migrations and optimized ETL processes, enhancing data sharing with SQL, Unix, and Python.

Wipro Technologies, Project Engineer, In   Sep 2020 – Jun 2021
Built ETL pipelines using PySpark in Databricks for a banking client, transforming financial data and providing insights for business improvement.

Odisha University of Technology and Research, Class Representative**,** In   Aug 2018 - Aug 2019
Managed attendance, faculty notices, and student-department communication in Electrical Engineering. Coordinated seminars and facilitated internship approvals for practical learning enhancement.

Odisha University of Technology and Research, Energy Club, Co-ordinator, In   2018- 2020
Led member inductions and organized sessions on tech advancements. Presented innovative ideas to professors, driving significant club growth. Raised funds for club activities, ensuring ongoing success and initiating new projects.

KEY PROJECTS

COVID-19 Workforce Salary Analysis Using R   Oct 2024
Analyzed salary trends across industries and job roles using messy, text-based data from a dataset. Conducted data cleaning, transformation, and visualization to generate various insights, and salary trends over time and geography. Offered insights into observed trends, addressing limitations & inconsistencies. Applied R for data manipulation, statistical analysis, and visualization.

SDP MATCH: Customer model and Reporting model   Apr 2022 – Dec 2023
Built GCP Dataproc data pipelines using PySpark for customer and reporting models. Key tasks included deduplication, record linking, and DataFrame-based graphs. Set up workflows for clean data on business dashboards. Utilized the Splink module for probabilistic record linkage, improving matching precision and calculating match probabilities. Incorporated graphframes in PySpark for advanced data relationship analysis, optimizing data deduplication and enhancing analytical capabilities.

TRAINING AND CERTIFICATIONS

- Associate Cloud Engineer Certification from Google Cloud
- Data Science Bootcamp: The Data Science Course
- Statistics & Data Analytics For Data Science And Business [Topics:Visualizing, Describing Data,correlation, regression]
- Gold Badge for Problem-Solving on HackerRank (Python)

VOLUNTEER EXPERIENCE

NSS (National Service Scheme) Volunteer, India   2017 - 2019