GAURAV SINGH

+91 7668950731 | gauravbisht2803@gmail.com | linkedin.com/in/contact-gauravsingh | github.com/gaurav280303 | Dehradun, India

SUMMARY

Entry-level Data Scientist and Data Engineer with 1 year of experience in AML/KYC analytics, cloud data pipelines, and machine learning. Proficient in Python, SQL, Databricks, Power BI, and Spark. Built end-to-end ELT pipelines on AWS S3 and Databricks Delta Lake. Developed NLP classification models for financial fraud detection. Pursuing MBA in AI & Data Science (CGPA: 8.7/10). Certified by Oracle OCI, NVIDIA, and McKinsey.

WORK EXPERIENCE

Anaptyss — Financial Services   Aug 2024 – Jul 2025
Data Analyst   Gurugram, India

- **Boosted** AML risk detection accuracy by **25%** by analyzing customer transaction patterns using SQL, Power BI, and Advanced Excel.
- **Automated** compliance reporting workflows using Generative AI tools, reducing manual effort by **20%** and ensuring **100%** audit readiness.
- **Achieved 100% KYC/AML compliance** for **M&T Bank** by delivering data-driven insights to senior stakeholders and coordinating cross-functional teams.

PROJECTS

End-to-End FMCG Data Pipeline | Databricks & AWS S3   Feb 2026
[Github Repository]
Tech Stack: Python, SQL, PySpark, Databricks, Delta Lake, AWS S3, Auto Loader, Spark Structured Streaming, Unity Catalog

- **Architected** Medallion Architecture lakehouse **(Bronze → Silver → Gold)** on **Databricks** Delta Lake & **AWS S3**, converting raw **FMCG** sales data into business-ready analytics.
- **Engineered** real-time **ELT** pipelines using Auto Loader & Spark Structured Streaming, automating incremental data ingestion across 3 pipeline stages.
- **Orchestrated** Databricks Workflows with Unity Catalog for data lineage, governance, and end-to-end pipeline automation.

Chargeback Dispute Intelligence System | NLP & Risk Scoring   Jan 2026
[Github Repository]
Tech Stack: Python, NLP, TF-IDF, Logistic Regression, Scikit-learn, Pandas, NumPy, Feature Engineering, Cross-Validation

- **Developed** NLP-powered complaint classification pipeline **(TF-IDF + Logistic Regression)** to auto-route chargeback disputes, reducing manual intervention in financial fraud resolution by **40%**.
- **Optimized** model via class-weighting and cross-validation, improving accuracy and enabling real-time risk scoring for high-impact bank dispute categories.

SKILLS

Programming & Databases: Python (Pandas, NumPy, Scikit-learn, TensorFlow), SQL, PostgreSQL, MySQL, Git & GitHub.
AI & Machine Learning: Machine Learning, Deep Learning, NLP, Generative AI, Agentic AI, Predictive Modelling, Statistical Analysis.
Data Engineering & Cloud: Databricks, Delta Lake, PySpark, AWS S3, Azure, Auto Loader, Unity Catalog, ETL/ELT Pipelines.
Data Visualization: Power BI, Tableau, Advanced Excel (Macros, VBA, Pivot Tables).
Accounting & Finance: AML/KYC Compliance, Risk Analytics, Data Governance, Financial Modelling, CDD, Tally Prime.

EDUCATION

Graphic Era University, Dehradun   2025 – Present
Master of Business Administration (MBA) — Artificial Intelligence & Data Science
CGPA: 8.7/10 | Coursework: Machine Learning, Deep Learning, Data Engineering, Statistical Modelling, Database Systems

LEADERSHIP & ACTIVITIES

AIESEC in Dehradun — UN SDG Partner | World's Largest Youth NGO   Aug 2025 – Present
Business Development Manager   Dehradun, India

- **Spearheaded** business development for **AIESEC's World's Largest Lesson (UN & UNICEF-backed),** raising **INR 1 Lakh** in sponsorships and securing brand partnerships.
- **Led** cross-functional team , driving stakeholder negotiations and delivering **SDG** awareness to **500+ students**.

National Service Scheme (NSS)   Jul 2022 – May 2024
Content Team Member   Dehradun, India

- **Produced** social awareness content and outreach campaigns for NSS, engaging **1,000+ students** across 2 years of community-driven initiatives.

CERTIFICATIONS

Oracle Cloud Infrastructure 2025 Data Science Professional (1Z0-1110-25)	*Oracle University*
Oracle Cloud Infrastructure 2025 Multicloud Architect Professional (1Z0-1151-25)	*Oracle University*
NVIDIA Generative AI with Diffusion Models	*NVIDIA Deep Learning Institute*
McKinsey Forward Program Certification 2025	*McKinsey & Company*
Microsoft & EY AI Skill Passport	*Microsoft & EY*