

Education

<b>PG DIPLOMA IN BIG DATA ANALYTICS</b> – Centre of Development and Advance Computing, Pune	03/2024 – 09/2024
Majors: Data Collection and DBMS, Python, Data Visualization- Analysis and Reporting, Practical Machine Learning, Advance Analytics, Linux Programming and Cloud Computing	
<b>BACHELOR OF ENGINEERING</b> – RVS College of Engineering and Technology, Jamshedpur	06/2020 – 07/2023
<b>DIPLOMA</b> – Al-Kabir Polytechnic, Jamshedpur	06/2015 – 07/2018
<b>MATRICULATE</b> – Kerela Public School, Jamshedpur	03/2008 – 03/2015

Skills

- SQL (SQL Server, MySQL, PostgreSQL, BigQuery)
- Tableau (Data Visualization, Dashboard creation, Data Blending, Storytelling)
- Python (Pandas, NumPy, SciPy, Scikit-Learn, Matplotlib, Seaborn)
- MLOps, Dagshub
- Git and Github (Version Control, Push, Pull, Branching, Merging, Collaborative Workflow)
- Docker (Containerization, Image Creation, Docker Compose)
- AWS (IAM, ECR, EC2, S3, Basic Cloud Computing Concepts)
- CI/CD Pipelines (GitHub Actions, Workflow Automation)
- Advance Excel (VLOOKUP/XLOOKUP, conditional formatting, complex IF functions, dynamic tables, pivot tables, DAX calculations, Power Pivot)
- Linux (Debian)

Projects

PHISHING WEBSITE DETECTION | END-TO-END ML PIPELINE

| PG Diploma Project | July 2024 | [https://github.com/saifimd1234/ML\\_network\\_security](https://github.com/saifimd1234/ML_network_security)

**Objective:** Built a production-grade ML system to classify phishing vs. legitimate websites using URL-based features (31 columns, e.g., SSL validity, domain age, IP presence).

- ML Pipeline: Designed modular OOP architecture for data ingestion, validation, and transformation workflows using Python, pandas, and scikit-learn.
- Modeling: Trained classifiers (Logistic Regression, XGBoost, Random Forest) with hyperparameter tuning (GridSearchCV) and achieved 95% F1-score.
- MLOps: Containerized the app via Docker, pushed images to AWS ECR, and deployed to EC2 for real-time inference. Automated CI/CD with GitHub Actions.
- Tools: Integrated Dagshub for experiment tracking, AWS S3 for artifact storage, and FastAPI.
- Robustness: Implemented logging, schema validation, and exception handling to ensure pipeline reliability.

**Technical Stack:** Python, Docker, AWS (ECR, EC2, S3), scikit-learn, Git, GitHub Actions, Dagshub, FastAPI, MLOps

Certifications

- HackerRank SQL (Basic and Intermediate)
- HackerRank Python (Basic)
- Essential Business Communication Skills – LinkedIn Learning
- Data Visualisation: Empowering Business with Effective Insights Offered by Tata Group on Forage March 2025

Work Experience

<b>BLUE BELLS HIGH SCHOOL, DIMNA — SENIOR MATHS TEACHER</b>	03/2021 – 03/2022
<ul style="list-style-type: none"><li>Taught Mathematics and Physics to classes 7-10, focusing on conceptual clarity and data interpretation skills.</li><li>Designed interactive lessons, developing students' problem-solving abilities and logical thinking.</li></ul>	