

Harshith Oberoi

732-322-2993 | oberoiharshith@gmail.com | [linkedin.com/in/harshith-oberoi](https://www.linkedin.com/in/harshith-oberoi) | oberoiharshith.github.io/portfolio

Impact at scale: 60% faster validation, 10M-row 99.7% SLA dashboards, 88% precision Gen-AI chatbot — ex-Capgemini (Mercedes-Benz), M.S.DS Rutgers

TECHNICAL SKILLS

Languages & Query: Python, SQL, Java, R, Excel

ML / AI: XGBoost, scikit-learn, TensorFlow, PyTorch, Keras, LangChain, Hugging Face

Data Eng & Orchestration: Spark, PySpark, Airflow, DBT, Docker, MLflow, Databricks

Cloud, Warehouses & BI: AWS, Azure, GCP (BigQuery), Snowflake, Tableau, Power BI

Databases: PostgreSQL, MySQL, MongoDB, Neo4j

Core Methods: A/B Testing, Predictive Modeling, Big Data Analytics, NLP, Deep Learning, ETL, Data Visualization

EXPERIENCE

Rutgers University

New Brunswick, NJ

Lecturer & TA - Introduction to Data Science

Jan 2025 – May 2025

- Mentored **50+** students by creating **40+** Python/data-science tutorials and mini-projects, improving assignment scores by **18%**; nominated for Best Lecturer award.

Cogent Infotech

Remote

Data Science Intern

Jan 2024 – May 2024

- Built a LangChain-based chatbot on 35K support tickets, achieving **88%** top-3 answer precision.
- Integrated FAISS-based RAG, cutting mean response time by **10%** (1.25 s to 1.12 s).

Capgemini– Mercedes-Benz Account

Remote

Senior Analyst - Data Science & Engineering

Jul 2022 – Jul 2023

- Implemented Random Forest-based test-case prioritizer, cutting validation time by **60%** (\approx 240 hours / release).
- Orchestrated PySpark / Airflow ingestion of **10M** telemetry rows per day; sustained **99.7%** SLA.
- Delivered real-time Grafana dashboards for **20+** stakeholders to monitor system health.

Analyst - Systems Engineering & Data Automation

Mar 2021 – Jun 2022

- Led sentiment analysis on **50K** reviews; insights helped improve CX KPIs across five vehicle lines.
- Designed SysML modeling framework, cutting diagramming effort by **85%** (\approx 2 FTEs / release).
- Automated diagram generation via Python & Flask GUI, shrinking clicks from **25** to **6** and standardizing output.

Itronix Solutions

Remote

Data Science Intern

Jan 2020 – Jul 2020

- Developed COVID-19 forecasting models (SVM, Regression) with **87%** accuracy, adopted by local authorities.

PROJECTS

Lang2Query | Python, LoRA, RAG, FAISS

- Fine-tuned Mistral-7B (LoRA + RAG) on **262K** Q & A pairs with FAISS for similarity search.
- Boosted ROUGE-1 by **35%** over LSTM + Mistral baselines in offline evaluation.

SignalSense | Python, XGBoost, MLflow, GCP Vertex AI

- Engineered XGBoost ensemble for 29 financial instruments; back-tested with Sharpe 0.3, tracked in MLflow.
- Forward-tested strategy on unseen data; deployed to GCP Vertex AI and validated long-term stability.

OptiGrade | Python, XGBoost, AWS, Docker, Tableau

- Built end-to-end student performance predictor using EDA + ensemble regressors, achieving **92%** accuracy.
- Deployed model to AWS Beanstalk with Docker; enabled CI/CD via CodePipeline and visualized insights in Tableau.

EDUCATION

Rutgers University

New Brunswick, NJ

M.S. Data Science (Statistics, GPA 3.85 / 4.0)

May 2025

Manipal University Jaipur

Jaipur, India

B.Tech. Information Technology

Jun 2020