Harshith Oberoi

732-322-2993 | oberoiharshith@gmail.com | linkedin.com/in/harshith-oberoi | oberoiharshith.github.io/portfolio

Cut QA cycles 60% & deliver insights from 10 M events/day, ex-Capgemini (Mercedes-Benz), M.S. Data Science – Rutgers '25

EXPERIENCE

Capgemini – Mercedes-Benz Account
Remote

Senior Analyst - Data Science & Engineering — *Jul 2022 – Jul 2023*

- Slashed QA validation time by **60%** (≈ 240 hrs) and saved $**20K** per release by deploying a Random-Forest test-case prioritizer.
- Sustained **99.7%** SLA by orchestrating **10M** telemetry rows/day using PySpark + Airflow pipeline.
- Enabled **20+** stakeholders to monitor system health in real-time through Grafana dashboards.

Analyst - Systems Engineering & Data Automation — *Mar 2021 – Jun 2022*

- Improved CX KPIs across five vehicle lines by mining insights from **50K** reviews.
- Streamlined diagramming effort by **85%** (≈ 2 FTEs/release) with a SysML modeling framework.
- Standardized output and minimized clicks by **76%** (**25** to **6**) via a Python/Flask GUI.

Cogent Infotech
Remote

Data Science Intern — *Jan 2024 – May 2024*

- Achieved **88%** top-3 precision on 35 K tickets with a LangChain chatbot powered by FAISS-RAG.
- Decreased mean response time by **10%** (1.25 s to 1.12 s) through async LLM calls and caching.

Rutgers University
New Brunswick, NJ

Lecturer & TA - Introduction to Data Science — *Jan 2025 – May 2025*

- Elevated assignment scores by **18%** for **50+** students through **40+** hands-on tutorials; nominated Best Lecturer.

Itronix Solutions
Remote

Data Science Intern — *Jan 2020 – Jul 2020*

- Enabled local COVID-19 planning with **87%** accurate forecasts using SVM & regression models.

PROJECTS

Lang2Query | *Python, LoRA, RAG, FAISS*

- Boosted ROUGE-1 by **35%** over LSTM + Mistral baselines by fine-tuning Mistral-7B (LoRA + RAG) on **262K** Q&A pairs indexed in FAISS.

SignalSense | *Python, XGBoost, MLflow, GCP Vertex AI*

- Achieved Sharpe of **0.30** back-test on **29** instruments by engineering an XGBoost ensemble tracked in MLflow and deploying to Vertex AI for forward validation.

CloudBridge | *ADF, ADLS, Databricks, Synapse, Power BI*

- Delivered sub **5-min** Power BI refreshes for **100K** orders by designing a pipeline with ADF, Databricks and Synapse on MySQL & MongoDB data.

TECHNICAL SKILLS

Languages/Query: Python, SQL, Java, R, Excel
ML/AI: scikit-learn, XGBoost, TensorFlow, PyTorch, Keras, LangChain, Hugging Face
Data Eng/Orchestration: Spark, PySpark, Airflow, DBT, Databricks, Docker, MLflow
Cloud/BI: AWS, Azure, GCP (BigQuery), Snowflake, Tableau, Power BI
Databases: PostgreSQL, MySQL, MongoDB, Neo4j
Core Methods: A/B Testing, Predictive Modeling, NLP, ETL, Big-Data Analytics, Deep Learning, Data Visualization

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*