

MAYUR PATIL

Falls Church, VA | (571) 251-6813 | mayurpatil4001@gmail.com | [linkedin.com/in/mayurpatil26](https://www.linkedin.com/in/mayurpatil26) | github.com/mayur212626

PROFESSIONAL SUMMARY

M.S. Data Science student at GWU with hands-on experience building large-scale ML and data systems. Processed 500K+ log records with PySpark, built ensemble anomaly detectors with 540x critical error lift, and trained clinical ML models with full governance pipelines and live API deployment. Strong in Python, SQL, statistical modeling, and end-to-end ML system design. Published researcher open to relocation for a Summer 2026 Data Science internship.

EDUCATION

The George Washington University

Washington, DC | Expected Dec 2026

M.S. in Data Science | Coursework: Machine Learning, Deep Learning, Neural Networks, Data Mining, Trustworthy AI, Algorithm Design, Data Warehousing, Big Data Analytics

D.Y. Patil University | B.Tech in Computer Science

Pune, India | Jul 2024

TECHNICAL SKILLS

Languages: Python, SQL (Joins, Window Functions, CTEs, Data Validation), R, C++, JavaScript, Bash

Machine Learning and Deep Learning: PyTorch, TensorFlow, Scikit-learn, XGBoost, LSTM, CNN, Random Forest, Gradient Boosting, Deep Learning

Techniques: Feature Engineering, Model Governance, Bias and Fairness Checks, Hypothesis Testing, A/B Testing, Experimental Design, SHAP, NLP, CI/CD, Reproducibility

Data Engineering: Pandas, NumPy, PySpark, Apache Spark, Hadoop, SQL Pipelines, Data Curation, Docker, Containerization

Cloud and Tools: AWS SageMaker, AWS EC2, AWS S3, Google Cloud Vertex AI, Git, Docker, Linux, Jupyter, MLflow

MACHINE LEARNING AND DATA SCIENCE PROJECTS

Large Scale Log Anomaly Detection | PySpark, Scikit-learn, Isolation Forest, MLflow, FastAPI Spring 2026

- Built a large-scale telemetry monitoring pipeline processing 500K+ HTTP server log records using PySpark window functions for per-IP behavioral feature engineering and approxQuantile for heavy-IP thresholding; automated anomaly detection to identify and resolve system errors and distribution shifts in real time; partitioned Parquet output to S3 for EMR-ready querying.
- Engineered 20 behavioral features per IP including error rate, critical error rate, admin endpoint access count, response size z-score, and DoS signal composites; ensemble of Isolation Forest, Local Outlier Factor, and rule engine with 2-of-3 majority vote achieved 540x critical error lift and Precision@100 of 0.78.
- SHAP explainability, MLflow experiment tracking, and KS and PSI drift detection for continuous monitoring; delivered actionable anomaly insights via FastAPI with full model governance documentation. github.com/mayur212626/anomaly-detection

Clinical Lab Abnormality Predictor | PyTorch, Scikit-learn, FastAPI, Docker, Render.com Spring 2026

- Built an end-to-end clinical ML system predicting diabetes risk from lab values; applied group-median imputation, SMOTE class balancing, Random Forest and PyTorch neural network training, and bias audits across patient age groups; deployed a live FastAPI service on Render.com.
- Achieved AUC-ROC 0.9618 (Random Forest) and 0.9460 (PyTorch), 89% accuracy, 92% sensitivity, and zero fairness disparity across demographics; delivered actionable clinical decision-making insights applicable to millions of patients. Model governance status: APPROVED.
- Implemented SHAP explainability, MLflow tracking, Streamlit dashboard, Docker, and GitHub Actions CI/CD. Live API: clinical-lab-predictor.onrender.com | github.com/mayur212626/clinical-lab-predictor

SIGNAL: Sales Intelligence and Gap Analysis Layer | Python, Multi-Agent LLM, NLP, Streamlit Spring 2026

- Built a 5-agent LLM pipeline automating qualitative thematic analysis, sentiment tracking, and cross-document pattern synthesis across enterprise B2B sales transcripts; structured qualitative coding of pain points, objections, unspoken fears, and buying triggers; achieved 91.2% thematic recall vs. manual coding baseline (kappa=0.81).
- Buyer engagement scoring correlated $r=0.87$ with human raters (ICC=0.92); strategic advisory agent produces prioritized P1/P2/P3 recommendations rated 4.1/5 actionability by domain experts; deployed as interactive Streamlit dashboard with automated PDF executive reports. github.com/mayur212626/signal-ai

Multi-Modal Stock Price Prediction Using LSTM | PyTorch, AWS SageMaker, NLP, PySpark Fall 2025

- Designed and deployed an LSTM-based stock prediction pipeline on AWS SageMaker achieving MAPE of 7.19%; engineered 44 features combining 39 technical indicators with NLP sentiment scores derived from 6,700+ financial news articles; ran controlled A/B experiments achieving 7.1% accuracy improvement over baseline.
- Automated model versioning and retraining pipeline on AWS S3 with continuous performance monitoring and audit-ready documentation for reproducible and compliant experiment tracking.

PROFESSIONAL EXPERIENCE

Cybersecurity Engineer Intern, Academor

Bengaluru, India | Nov 2023 - Dec 2023 (8 weeks)

- Built Python automation scripts for data monitoring workflows with real-time alerting and observability dashboards; reduced manual effort by 40% and enabled 2x faster issue detection across a 50+ server environment.
- Investigated and resolved 15+ data integrity issues across web and database layers; documented findings and communicated prioritized remediation plans to cross-functional engineering teams; implemented compliance checks reducing open vulnerabilities by 30% within 8 weeks.

PUBLICATIONS

- Data Security Management*, International Journal of Innovative Research in Engineering and Multidisciplinary Physical Sciences, Sep 2024. [ijirmps.org/research-paper.php?id=231231](https://www.ijirmps.org/research-paper.php?id=231231)
- Gesture Recognition Using Virtual Mouse*, International Journal of Innovative Research in Engineering and Multidisciplinary Physical Sciences, Aug 2024. [ijirmps.org/research-paper.php?id=230945](https://www.ijirmps.org/research-paper.php?id=230945)