Amit Prajapati

aprajapati@wpi.edu | +1 (508)410-9274 | LinkedIn | GitHub | United States

EDUCATION

Master of Science in Data Science, Worcester Polytechnic Institute, MA, USA   **Aug 2023 - May 2025**
Relevant courses: Data Structures and Algorithms, Big Data Management, Generative AI, Business Intelligence, Machine Learning

Bachelor of Technology in Data Science*, NMIMS, India   **Aug 2019 - May 2023**
Relevant courses: Artificial Intelligence, Cloud Computing, Statistics, Data Ethics, Finance, SAS, Advance Database System, Deep Learning

EXPERIENCE

Offshore Construction Associates, Boston, MA – *Data Science Intern*   **May 2024 – Present**
- Automated data collection pipelines by integrating Selenium, Beautiful Soup, and cron scheduling, reducing manual extraction efforts by 70% and ensuring accurate, real-time tracking of offshore construction metrics.
- Designed interactive Power BI dashboards to visualize wave sensor data and assess sea conditions, enhancing decision-making for offshore wind project safety and informing future development strategies.
- Fine-tuned LLaMA 3 on scraped scientific documents using Hugging Face Transformers; built an agentic RAG-based pipeline with semantic search to autonomously generate and summarize domain-specific knowledge, simulating open-ended scientific inquiry.

True Light Energy, Boston, MA – *Machine Learning Intern*   **Jan 2025 – Present**
- Engineered scalable ML data pipelines with automated ingestion and transformation using REST APIs and AWS EC2; deployed models in production environments, aligning with MLOps best practices for model serving and versioning.
- Forecasted energy consumption using LSTM variants (Vanilla, Bi-LSTM) to perform predictive analysis, enabling data-driven business decisions and improving demand planning as well as improving model prediction accuracy by 60%.
- Crafted a normalized PostgreSQL schema for time-series energy data, accelerating query performance by 90% and reducing dashboard load times across analytics platforms which led to faster decision making and development.

Munich RE, Mumbai, India – *Data Analyst*   **Dec 2022 – May 2023**
- Analyzed large-scale insurance records using PySpark on Azure Databricks by connecting to Azure Data Lake; applied Apriori and FP-Growth algorithms to detect fraud and risk patterns, raising red flags for the client to act.
- Architected an R-CNN-based model to classify diverse insurance claims documents, enhancing triage accuracy and accelerating downstream ML workflows by reducing manual processing time and improving claim authenticity validation by 36%.
- Constructed interactive dashboards in Power BI to visualize fraud risk patterns identified; enabled stakeholders to monitor high-risk claims, prioritize investigations, and support data-driven decision-making in fraud management.

SKILLS

- **Programming & Data Engineering**: Python, R, JavaScript, TypeScript, PL/SQL, C, C++, Swift, MySQL, MongoDB, Apache Spark
- **Data Analytics & BI** :Power BI, Tableau, Excel
- **Machine Learning & AI**: TensorFlow, PyTorch, Scikit-Learn, XGBoost, Langchain, OpenCV, NLTK, NumPy, Pandas, Dask, SciPy, Matplotlib
- **MLOps & Cloud**: MLflow, Docker, Git, AWS EC2, Azure, GCP, Snowflake, CI/CD (cron-based), Kubernetes (basic)

RESEARCH PROJECT

Anomaly Detection in Side-Channel Traces for Data Leak Prevention **| Python, CNN, MLP, PCA.**
- Engineered a side-channel anomaly detection pipeline to identify Points of Interest (POIs) using PCA, forward-difference, and Z-score methods, significantly reducing the number of traces needed to detect encryption key leakage.
- Integrated a high-throughput data acquisition and transfer system using NFS and FTP protocols, enabling parallelized trace collection and anomaly detection over 60,000+ encrypted samples with 99.9% accuracy.

PROJECTS

WPIBot | **AWS, LLMs, Python** | **Link**
- Finetuned a chatbot using LLaMA 3 (via Groq API) and FAISS to semantically retrieve and answer campus-related queries, enhancing student access to WPI-specific information.
- Deployed a Q&A platform with a custom web scraper and Sentence-BERT embedding pipeline to index 2,000+ WPI pages, then launched a full-stack Streamlit app on AWS EC2 and optimized low-latency LLM performance via FAISS.

Object Detection with Satellite Images | **Python, Computer Vision, GitHub, Yolo** | **Link**
- Evaluated the trade-off between object detection accuracy and pixel resolution in satellite imagery, identifying the optimal data resolution that maximizes detection performance while minimizing computational resources and processing time.
- Structured a pipeline using the xView dataset and YOLO, generating resolution-performance curves to identify optimal "knee" points, reducing data usage by 70% while retaining 90% detection accuracy, significantly improving satellite image analysis.

Real-Time Traffic Sign Classification| **Python, Deep Learning**| **Link**
- Built a real-time traffic sign classifier with low-latency inference pipeline; containerized system using Docker and deployed on edge-compatible architecture simulating production ML deployment.
- Implemented a robust preprocessing and augmentation pipeline (direction-aware vs. non-directional), improving model generalization and enabling smooth integration with live video input for real-time inference.