Priyadharshan Sengutuvan

sengutuvan.p@northeastern.edu | (857) 397-8847 | linkedin.com/in/priyadharshan-sengutuvan

EDUCATION

Northeastern University, Khoury College of Computer Sciences — Boston, MA
Master of Science in Data Science — May 2026
Courses: Machine Learning, NLP, Algorithms, MLOps

Sri Krishna College of Engineering and Technology — Coimbatore, India
Bachelor of Technology in Information Technology — May 2023

SKILLS

Programming: Python (NumPy, Pandas, Scikit-learn, PyTorch, TensorFlow), SQL, Java, R, C#
Data Engineering: ETL Pipelines, Apache Airflow, GCP BigQuery, Data Modeling, Data Quality, Docker, Git
Analytics: A/B Testing, Statistical Analysis, Data Visualization, Tableau, Business Intelligence, Product Analytics
Machine Learning: Supervised/Unsupervised Learning, XGBoost, Feature Engineering, Model Evaluation, MLOps
Frameworks: FastAPI, LangChain, Prometheus, Grafana, MLflow, CI/CD

PROFESSIONAL EXPERIENCE

AriesView — Boston, MA (Remote)
AI/ML Engineer Intern — Sept 2025 – Dec 2025

- Architected data models and real-time analytics dashboards enabling 1K+ investors to evaluate property performance through NOI, cap rate, and DSCR metrics, directly impacting investment decisions and user satisfaction.
- Built and maintained production data pipelines automating multi-scenario financial projections with sensitivity analysis, processing 500+ properties daily and accelerating analyst workflows by 60%.
- Owned data quality for financial records by implementing OCR validation rules and automated checks, improving extraction accuracy to 95% while eliminating manual data entry errors.
- Optimized RAG chatbot retrieval system through systematic A/B testing of embeddings and search strategies, improving user query accuracy from 68% to 87% and reducing support resolution time.

Psiog Digital Private Limited — Chennai, India
Software Engineer Intern — Feb 2023 – Dec 2023

- Partnered with business stakeholders to design and deploy React/C# dashboards displaying critical KPIs, replacing manual Excel workflows and enabling data-driven decisions across finance and operations teams.
- Performed data analysis on transaction patterns and optimized SQL Server through query rewriting and indexing, reducing analytics query load times by 75% and improving data accessibility for 50+ business users.
- Designed and implemented REST APIs connecting previously siloed backend services, streamlining data access and improving cross-functional team productivity by consolidating data sources.

PROJECTS

RAG-Based Operations Assistant for E-Commerce (GitHub) — *Python, LangChain, GCP, Airflow, Docker*

- Architected end-to-end data pipeline ingesting seller support data into BigQuery using Airflow ETL processes with incremental loading, enabling real-time analytics dashboards for operations team.
- Built RAG chatbot using LangChain and GPT-4 to automate seller support queries, improving response accuracy from 68% to 87% through systematic experimentation with embeddings, chunking strategies, and prompt engineering.
- Established data quality framework ensuring semantic search reliability and maintaining data freshness for 100K+ product records.

Customer Churn Analytics & Prediction System (GitHub) — *Python, FastAPI, Docker, Prometheus*

- Performed exploratory data analysis on 7K+ telecom records identifying three high-risk customer segments representing 35% of churn, enabling targeted retention strategies and informing product improvements.
- Deployed production ML system with A/B testing framework comparing Random Forest and Gradient Boosting models, optimizing latency by 84% (36ms to 5.7ms) while maintaining comparable accuracy (0.84 vs 0.83 AUC).
- Implemented monitoring infrastructure with Prometheus and Grafana tracking data pipeline health, prediction volume, and error rates across containerized services to ensure data quality and system reliability.

F1 Race Position Predictor (GitHub) — *PyTorch, FastAPI, MLflow, scikit-learn*

- Analyzed 26K+ historical racing events to engineer 47 temporal features including rolling statistics, momentum indicators, and circuit-specific metrics, achieving 3.0 MAE and 70% podium prediction accuracy.
- Built bidirectional LSTM model with attention mechanism (990K parameters) and tracked training experiments across 100+ epochs using MLflow, monitoring validation loss decay to optimize model performance.
- Deployed production API serving real-time race forecasts with 50ms latency using FastAPI with Pydantic validation, enabling data-driven insights for racing analytics applications.