Alex Lee

Senior Data Scientist | Applied AI & Machine Learning
Maryland, US | alexllee@umich.edu | (832) 795-6031
LinkedIn | Github | Website

PROFESSIONAL SUMMARY

Senior Data Scientist with 3+ years of experience building and deploying production ML and AI systems in healthcare and enterprise. Expert in end-to-end ML development, deep learning, MLOps, causal inference, and data analytics, with a record of delivering measurable revenue, operational, and clinical impact.

SKILLS

Programming: Python (Scikit-Learn, PyTorch, PySpark, Transformers, LightGBM), R (Tidyverse, Tidymodels, Shiny).
ML & Statistics: End-to-End ML, Time Series, NLP, Causal Inference, Deep Learning.
MLOps & Cloud: MLflow, CI/CD, Databricks, Azure Synapse Analytics, Git.
Data & Visualization: SQL, Hadoop, Tableau, PowerBI, Alteryx.
Generative AI: Prompt Engineering, Multi-Agent System, LTM/STM, VectorDB, RAG.
Other: Hugging Face, Ollama, REST APIs, Django.

EXPERIENCE

Humana

Senior Data Scientist ***Sep 2024 – Present***

- Designed and deployed an **ensemble forecasting framework** (LightGBM, ARIMA, Ridge) across millions of medical and pharmacy claims to predict healthcare quality-based bonus metrics, contributing to **~$100M** in annual revenue.
- Developed a **Monte Carlo-based Elastic Net regression** framework to quantify predictive uncertainty in clinician interaction outcomes, driving ~$4M in targeted interventions.
- Deployed a **daily** gradient-boosted classifier identifying members with gaps in follow-up care after ED visits, enabling prioritized outreach and improved outcomes.
- Designed and operationalized **TF-IDF** based condition modeling with regression analysis to measure and prioritize condition-specific impact on ED utilization.
- Implemented **CI/CD** pipelines for **26 production sub-models**, automating retraining, monitoring, and governance.

Senior Business Intelligence Engineer ***Jun 2023 – Sep 2024***

- Developed and deployed a **LightGBM** fall-risk classifier with **Bayesian hyperparameter optimization (Hyperopt)** for 5M Medicare members, doubling precision for high-risk identification and driving ~$2M in annual value.
- Applied **propensity score matching** and **difference-in-differences** for causal inference and intervention evaluation, yielding $100K+ in annual cost savings and $5M+ incremental revenue.
- Led 7+ cross-functional **A/B and quasi-experimental studies** from concept to deployment.

The Hershey Company

Lead Business Intelligence Analyst ***Dec 2022 – Jun 2023***

- Built a Random Forest classifier to predict employee turnover, enabling retention strategies for high value EEs.
- Developed a real-time REST API integrating **survival analysis** and **PCA** to monitor Talent Acquisition KPIs.

Senior Data Analyst ***Mar 2022 – Dec 2022***

- Conducted adverse impact analyses using **(Chi-square, t-tests, ANOVA)** to inform fair hiring decisions.
- Developed ROI models for tuition reimbursement programs, supporting $6M+ in annual funding.

Data Analyst ***Nov 2018 – Mar 2022***

- Automated People Analytics **dashboards**, replacing third-party tools and saving $200K annually.
- Optimized **ETL pipelines and data models**, reducing data prep time by 40 hours per month through SQL, and Alteryx.

EDUCATION

Master of Applied Data Science - University of Michigan, Ann Arbor ***Graduation Expected: Dec 2026***

Bachelor of Business Administration - University of North Texas ***Aug 2015***

PROJECT HIGHLIGHTS

- Multi-Agent RAG Education System for Children (Personal project; code not public): Designed and implemented a multi-agent LLM system with planner and tutor agents, retrieval-augmented generation, short and long-term memory, and safety-aware orchestration.
- Algorithmic Stock Trading (PyTorch): Built a Stock Market Prediction Platform combining technical analysis and deep learning (LSTM, Temporal Fusion Transformer) for multi-task forecasting of price changes and direction. Developed a modular pipeline covering data ingestion, preprocessing, model training, evaluation, and backtesting, with interactive notebooks for experimentation and detailed performance analysis.