Derek Albosta

derek.albosta@gmail.com | linkedin.com/in/derek-albosta | https://github.com/dalbosta

Summary

Software and Data Engineer with prior experience as a Data Scientist, specializing in distributed systems, cloud-native data platforms, and production machine learning systems. Proven track record of building high-impact architectures, conducting rigorous experimentation (A/B testing), and translating statistical insights into scalable engineering solutions.

Technologies

Languages: Python, C#, SQL, Java
Cloud & Infrastructure: Azure (Functions, Service Bus, Blob Storage, ADF), Microsoft Fabric, AWS
Backend & Frameworks: .NET 8, Flask, Django, REST APIs
Databases: PostgreSQL, MongoDB, MySQL, SQLite
Data & Machine Learning: NumPy, Pandas, Scikit-Learn, TensorFlow, PyTorch, PySpark, A/B Testing
Tools: Git, Jenkins, Tableau, Datadog, LangChain

Experience

Software Engineer (Contract)
Invene, Remote | Nov 2025 - Present

- Designed and built a **Value-Based Care Enterprise Data Warehouse on Microsoft Fabric** using a **medallion architecture (Bronze/Silver/Gold)**, engineering **PySpark** pipelines to process **19 payer-specific healthcare file layouts**, validated **595M+ rows (~60 GB)** of synthetic production-scale load.
- Built a **production-grade testing framework** for the EDW by extracting notebook transforms into **reusable utility modules**, implementing **unit tests** per transform, and embedding **data quality gates** at each pipeline stage to ensure correctness ahead of go-live.
- Implemented a threshold-based anomaly detection system following a critical vendor-file failure that incorrectly marked **10,000+ data rows**, eliminating recurrence of a **six-figure financial loss risk**.
- Architected and built a distributed health plan data validation platform capable of processing **100GB+ file batches** using **.NET 8 Azure Functions, Blob Storage, Azure Service Bus, and ADF**.
- Designed a **database-driven configuration validation system** supporting **500+ health plan and subject-area rule combinations**, eliminating code changes when onboarding new file types.
- Reduced client manual validation effort by **80-90%** by automating complex cross-field, threshold, and file-level business rule enforcement across large-scale datasets.

Technical Lead
Nice.Industries, Remote | Dec 2024 - Nov 2025

- Led greenfield architecture and development of Nice Industries' full-stack web platform, **owning infrastructure design, database schema, and backend services from inception to deployment.**
- Engineered a dynamic page-generation engine that populated templates directly from SQL-backed content models, **reducing new page deployment time from hours to minutes**.
- **Established engineering processes** (ticketing, sprint cadence, standups) to improve delivery velocity and technical alignment across a distributed team.
- **Selected for Beta University pre-acceleration program**, a competitive startup accelerator within the Bay Area ecosystem.

Software Engineer and Data Scientist
Indeed.com, Austin, TX | May 2021 - May 2024

- Designed and productionized an employer pricing model and supporting REST APIs in Python for SimplyHired.com **(1.2M+ monthly users)**, improving pricing accuracy (MAE) by **15%** and increasing revenue through statistically validated model enhancements.
    - Reduced model prediction latency by **70%+** by precomputing and caching outputs in MongoDB, significantly improving response time and scalability.
    - Validated model performance through rigorous **A/B testing**, ensuring statistically significant impact before full rollout.
- Developed a **GPT-3 powered content generation pipeline** using LangChain and curated career advice datasets to produce scalable, high-quality career guide content.
- Optimized advertising spend by reallocating **~$5M+ (5% of total budget)** using K-means clustering to identify underperforming job markets.
- **Built real-time monitoring dashboards** in Tableau and Datadog to enable data-driven decision-making across three core product teams.
- **Promoted from Software Engineer to Data Scientist** based on impact in machine learning modeling, statistical analysis, and cross-functional leadership.

Machine Learning Intern
Juva Health, San Diego, CA | Dec 2020 - May 2021

- Designed the **video de-identification pipeline** using a deep learning auto-encoder architecture with Keras/TensorFlow, **safeguarding user privacy in alignment with HIPAA standards**.
- Built Tableau dashboards for analyzing user behavior, providing insights to improve mobile app engagement and retention.

Education

B.S. in Computer Science, Minor in Mathematics
University of Puget Sound, Tacoma, WA | 2016 – 2020

Projects

Resume Tailor: LLM based resume suggestion tool (https://github.com/dalbosta/ResumeTailor)
A GPT-powered app comparing job descriptions and resumes, providing compatibility scores, actionable recommendations and interview preparation for specific job roles.

ChirpGAN: Generative Modeling for Bird Call Synthesis (https://github.com/dalbosta/ChirpGAN)
A GAN-based tool for synthesizing artificial bird calls using a custom data pipeline, progressive GAN architecture, and wavelet scalograms.