Dan Billmann

dan.billmann.13@gmail.com LinkedIn GitHub

Relevant Skills

Languages: Python (Pandas, NumPy, Matplotlib, SciPy, Scikit-Learn, PySpark, MLFlow, PyTest, XGBoost, TensorFlow, PyTorch, SQLAlchemy, OpenCV, Streamlit, shap, AIOHTTP), Bash, Rust, Go, C++
Machine Learning: Regression, Classification, Clustering, Deep Learning, Computer Vision, NLP, CNNs
Technologies: Azure (DevOps, SQL Server, Databricks, Data Factory, Maps), Docker, Hadoop, Spark
Other Technical Skills: ETL, SWE, OOP, Data Visualization, Data Analysis, Image Segmentation, Forecasting
Soft Skills: Collaborative, communicative, problem-solver, technical leadership, mentoring

Professional Experience

Senior Machine Learning Engineer Oct. 2023 – Present
Oak Street Health | CVS Health Remote (Philadelphia, PA - based)

- Drove $12M in new revenue within a month by scaling a GenAI RAG model from 1K to 30K generations per hour using PySpark, asynchronous API calls, and thread-safe shared queues.
- Engineered model serving pipeline with Databricks, MLFlow, and NumPy to improve model explainability, reduce deployment time by 83%, and include 3x more ML Models, resulting in 2X faster inference.
- Deployed a geocoding ETL pipeline to production with Azure Maps SDK, Bash scripts and SQLAlchemy, for the Population Health team to study patient demographics, resulting in 90% faster runtime on 330K additional patients
- Authored a in-house Python package using the standard Python library and the Databricks SDK to facilitate faster R&D for several cross-functional teams, amassing 13K+ downloads.
- Established software testing practices with PyTest by setting up 173 unit and integration tests at 100% code coverage.
- Introduced PEP8 standards, reusable code practices, and software packaging that improved the quality of all team members' software, dataset generation, and documentation contributions.
- Constructed CI/CD pipelines using custom bash scripts in Azure DevOps to publish PR completions to Databricks, create Docker images for ML training, and run PEP8 checks to ensure code quality and consistency.
- Created automated documentation deployment system using Sphinx, Diagrams, and Azure DevOps Pipelines to build two internal websites, allowing for easier collaboration and onboarding across all data science teams.

Machine Learning Intern May 2022 - Aug. 2022
VideaHealth Boston, MA

• Implemented a Gaussian-smoothed heatmap cost function on segmentation landmark points with NumPy and OpenCV to align training convergence and the visual representation of the predictions.
• Conducted A/B testing to study the runtime performance of 3 lightweight CNN encoder backbones to confirm alternative encoders, though faster, were less performant.

Data Engineer Jun. 2017 – Jun. 2021
Bloomberg LP Princeton, NJ

• Designed a custom algorithm using NumPy and Pandas to capture billions of missing data points.
• Redesigned a batch processing ETL pipeline with Python multithreading to reduce runtime by 10 hours and improve scalability and maintenance.
• Built an end-to-end document classification system using Python microservices, Hadoop, proprietary cloud storage and Splunk to save $250K / yr on headcount.
• Implemented a TF-IDF model on 50K municipal documents on a Hadoop cluster using PySpark, achieving 95% precision and 90% recall across 5 different document types.

Education

M.S. in Computer Science - Thesis SUNY at Stony Brook May 2023
B.B.A. in Management Information Systems University of Cincinnati April 2017