Zachary Allen

zackallendata@gmail.com | linkedin.com/in/allenza | ZackAllen1.github.io

WORK EXPERIENCE

Data Scientist — March 2024 – Present
Abt Global — *Tampa, FL*

- Automated AWS Transcribe JSON-to-Word conversion using EventBridge and Lambda, saving $80,000 in transcription costs.
- Optimized statistical model by implementing a Julia-based GLMM with parametric bootstrapping, improving execution time by 55% and memory efficiency by 64% compared to R and Python equivalents.
- Led migration of SAS-based claims ETL to PySpark on Databricks for the Centers for Medicare and Medicaid Services, improving scalability and maintainability.
- Developed entity resolution clustering methodology with PySpark and scikit-learn to group Medicaid enrollees with poorly labeled identifiers, reducing processing time from weeks to minutes.
- Fine-tuned XGBoost model with Optuna to predict carbon emission report resubmission probability for the Environmental Protection Agency.

Data Scientist Co-op — Jan. – June 2023
Johnson & Johnson, Janssen BioResearch Quality & Compliance — *Tampa, FL*

- Developed a full-stack Dash Application to analyze regulatory articles and perform BERT topic modeling, reducing search time by 33%.
- Performed cost-effectiveness analysis of LangChain and LLM related solutions for NLP tasks on multiple unstructured PDFs that exceed most context windows.
- Designed a data pipeline from SharePoint to PostgreSQL with Power Automate and AWS Lambda Functions.
- Created 12 Tableau Dashboards to monitor QA Tool Usage and report on Real-time Transcription Tools.

Data Analyst Intern — June – Aug. 2021
TotalSDS — *Tampa, FL*

- Leveraged Power BI to develop reports for development productivity, improving Pull Request Lead Time by 40%.
- Completed multiple Data Cleaning/Analysis projects, resulting in a 30% reduction in data entry time and enabling the sales team to make data-driven decisions.

EDUCATION

Georgia Institute of Technology — Jan. 2025 – Present
Masters in Analytics, Computational Data Analytics Track

University of Florida — Dec. 2023
Bachelors of Science in Computer Science, Minor in Statistics — *GPA: 3.88/4.0*

PROJECTS

MAISTRO | *Python, Next.js, FastAPI, MongoDB* — Aug. – Dec. 2023

- Designed a chatbot to convert conversation into Spotify song recommendations using GPT-4 and a custom emotion classification model based on Google's 50k observation GoEmotion dataset.
- Implemented front-end with multiple themes and layouts using Next.js, TailwindCSS, and shadcn/ui components.

After the Bell Data Challenge | *R, Git* — Sept. – Dec. 2022

- Investigated impact of school choice and family engagement on K-12 student education to provide recommendations on factors that optimize academic excellence using multiple statistical models.
- Led a team of four to clean, explore, and evaluate over 25,000 student-outcome surveys spanning multiple years using Cumulative Logit and Multiple Logistic Regression Models.

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, Julia, TypeScript, C++, Java
Python Libraries: TensorFlow, Scikit-learn, NumPy, Pandas, Matplotlib, Plotly Dash, Streamlit, Optuna, HuggingFace Transformers, LangChain
Analytics: PostgreSQL, Databricks, Apache Spark, Snowflake, PowerBI, Tableau, Power Automate, Excel
Tools: React, Jupyter Notebook, Docker, GitHub/Git, LaTeX, PowerPoint