CHRISTOPHER CABALLERO

(305) 733-3568 | chrismcaballero@gmail.com | chrismcaballero.com
github.com/chris-caballero | linkedin.com/in/christopher-caballero-696b6b219

EDUCATION

MS in Computer Science
Florida State University | *4.0 GPA*

BS in Mathematics
Florida State University | *Magna Cum Laude*

Machine Learning Engineering for Production (MLOps) Specialization
Coursera | *June 2023 - Sep 2023*

EXPERIENCE

Volunteer Data Scientist
Florida State University | *Sep 2022 - Nov 2022*

- Collaborated with an interdisciplinary team to enhance a COVID-19 research knowledge graph, actively contributing to standup meetings for project advancement.
- Worked extensively with a 900GB MongoDB sharded cluster of COVID-19 research papers.
- Engineered models for classification of semi-structured tables, employing efficient distributed training on PyTorch. This led to two publications, available for review on my LinkedIn profile.
- Improved topic classification F1-score by 5% over previous methods, and achieved a +6% gain over SOTA in binary metadata classification on large-scale heterogeneous data.

PROJECTS

Support Ticket Classification | April 2023 - Present

- Developed and deployed a simple full-stack Flask app to serve model inference requests.
- Created a Dockerfile to support reproducible local deployment.
- Architected a compact yet powerful Encoder Transformer model, yielding a 1% accuracy gain over fine-tuned BERT and a significant 15% improvement over CNN in 10-fold Cross-Validation. Supporting research is available on GitHub.

Fraudulent Transaction Detection | June 2023 - Present

- Conducted EDA and handled imbalanced data with under-sampling and outlier removal.
- Performed a comparative analysis on three simple models, logistic regression, random forest, and SVC, on the task of flagging fraudulent transactions.
- Utilized Google Cloud Storage for data management, Google AI Platform for XGBoost model training, and Google Cloud Functions for model deployment. Improved test set AUC by +2.3% over the baseline.
- Developed a Jupyter notebook to demonstrate model predictions using Google Cloud Function, showcasing proficiency in cloud services integration and API utilization.

Carbon Emission Analysis | July 2023 - Present

- Conducted comprehensive analysis of carbon emission data across England, focusing on Lower Layer Super Outer Areas (LSOAs), to extract critical environmental insights.
- Employed Python and GeoPandas for geospatial data visualization.
- Leveraged Docker to create a consistent and replicable analysis environment.

SKILLS

Programming Languages: Python, C, C++, SQL, JavaScript, HTML, CSS
Tools and Frameworks: Git, Docker, Google Cloud (AI Platform, Storage, Functions), Flask
Database: MongoDB, SQLite