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

Machine Learning Research Assistant
Florida State University | *Sep 2022 - Nov 2022*

- Collaborated with an interdisciplinary team to enhance a COVID-19 research knowledge graph, actively contributing in 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.

Graduate Research Assistant
Florida State University | *May 2022 - Aug 2022*

- Conceptualized an innovative approach to investigate polysemy by exploring the relationship between Graph Neural Network and simple language embeddings.
- Utilized statistical analysis and dictionary learning to process and interpret data for identifying patterns and separating valid semantics from words.

PROJECTS

Support Ticket Classification | April 2023 - August 2023

- Designed a user-friendly Flask application for seamless model serving, offering instant classification of user-generated support tickets with over 93% validation accuracy.
- Created an optimized Encoder Transformer model, yielding a 0.925% accuracy gain over fine-tuned BERT and a remarkable 15% improvement over CNN in 10-Fold Cross-Validation, supported by a research paper available on GitHub.

Fraudulent Transaction Detection | June 2023 - Present

- Conducted EDA and handled imbalanced data with under-sampling and outlier removal.
- Utilized Google AI Platform on GCP to deploy a fine-tuned XGBoost model, leveraging Google Cloud Functions for predictions. Improved test set AUC by +2.3% over the baseline.
- Performed a comparative analysis on three models, logistic regression, random forest and SVC, on the task of flagging fraudulent transactions.

SKILLS

Machine Learning: Topic Clustering, Text Classification, Classification, Regression, Clustering, Anomaly Detection, Cloud Computing
Tools and Frameworks: Docker, PyTorch, TensorFlow, Google Cloud (AI Platform, Storage, Functions), Flask, scikit-learn, numpy, pandas, matplotlib, seaborn
Programming Languages: Python, C, C++, SQL, MQL
Database: MongoDB, SQLite