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 - Machine Learning | Florida State University
BS in Mathematics | Florida State University
Introduction to Machine Learning in Production | Coursera | June 2023
Machine Learning Data Lifecycle in Production | Coursera | June 2023
Machine Learning Modeling Pipelines in Production | Coursera | August 2023

PROJECTS

Support Ticket Classification | April 2023 - August 2023
- Engineered a user-friendly Flask application with seamless model serving, offering an intuitive interface for instant classification of user-generated ticket data.
- Created an optimized Encoder Transformer language model to enhance sentence embeddings and classification accuracy.
- Utilized scikit-learn pipelines for modular data transformations and developed efficient data input mechanisms using the TicketDataset class and DataLoader.
- Achieved significant performance gains, including a 0.925% accuracy improvement over fine-tuned BERT and a remarkable 15% boost over CNN in 10-Fold Cross-Validation, supported by a research paper available on GitHub.

Fraudulent Transaction Detection | June 2023
- Trained three models, logistic regression, random forest and SVC, to flag fraudulent transactions. Additionally trained XGBoost for optimized performance on this task.
- Conducted exploratory data analysis to visualize class distribution and inform data preprocessing decisions.
- Implemented data preprocessing techniques to enhance model performance and handle class imbalance effectively.
- Conducted rigorous model evaluation using cross-validation and ROC analysis.

Fake News Detection | July 2023 - Present
- Developed and deployed a binary classification model using 'distilbert-base-cased' and 'distilbert-base-uncased' language models for fake news detection on the LIAR Dataset.
- Implemented text pre-processing techniques considering potential inconsistencies in writing style, capitalization, and punctuation in fake news data.
- Developed a binary classification model for fake news detection with a notable accuracy of 67%, showcasing significant improvement over the state-of-the-art 6-class classification results.
- Finetuned DistilBERT models were added to the HuggingFace Model database, accessible through the HuggingFace inference API.

EXPERIENCE

Machine Learning Research Assistant
Florida State University | Sep 2022 - Nov 2022
- Collaborated effectively with an interdisciplinary team to enhance a coronavirus research knowledge graph, actively participating in standup meetings and contributing to the project's advancement.

- Managed a MongoDB sharded cluster housing extensive research papers, overseeing a distributed data storage exceeding 900GB.
- Led the creation and implementation of innovative machine learning architectures and embeddings, driving forward the team's efforts with strategic contributions.
- Engineered an automated model for improved classification of semi-structured tables, ensuring efficiency and enhancing retrieval and tuple classification.

Graduate Research Assistant
Florida State University | May 2022 - Aug 2022
- Conducted literature reviews and analyzed linguistic and computational resources to understand word polysemy.
- Developed a novel idea to study polysemy by exploring the relationship between polysemy encoding by Graph Neural Networks (GNNs) and simple natural language models.
- Utilized statistical analysis and dictionary learning to process and interpret data for identifying patterns and separating valid semantics from words.

SKILLS
Machine Learning:
- Natural Language Processing (NLP)
- Classification, Regression, Clustering, Anomaly Detection

Libraries:
- PyTorch, Keras, TensorFlow, scikit-learn
- numpy, pandas, matplotlib, seaborn

Programming Languages:
- Python, C, C++, SQL, MongoDB Query Language (MQL)

Methodologies:
- Agile, Object-Oriented Programming (OOP)

Database:
- MongoDB, SQLite

Frameworks:
- Docker

PUBLICATIONS
Kandibedala, B., Pyayt, A., Piraino, N., Caballero, C., and Gubanov, M., 2023. COVIDKG.ORG-a Web-scale COVID-19 Interactive, Trustworthy Knowledge Graph, Constructed and Interrogated for Bias using Deep-Learning.