Marcos Tidball

📱 +55 51 999269599 | ✉ marcostidball@gmail.com | 🏠 zysymu.github.io | 💻 github.com/zysymu | 💼 linkedin.com/in/zysymu

Work Experience

BTG Pactual
Sao Paulo, SP, Brazil (Remote)

Data Scientist
Aug 2022 - Now

- **Performed exploratory data analysis on time series data of user transactions across multiple stock market sectors**, using statistical methods and custom feature engineering to find meaningful and actionable correlations. Employed visualization tools to present findings to stakeholders and guide decision-making.
- **Designed and implemented an OCR API using Python and AWS**, leveraging machine learning models to extract personal information from Brazilian and foreign documents. The API was integrated with government databases and cross-checked with user entries to determine validity, **achieving an 86% reduction in manual analysis**. Implemented a human-in-the-loop system to handle low confidence cases.
- **Improved biometric network AWS API in Python** to enable employees to more efficiently analyze and correct problematic facial and document user data, **leading to a 70% reduction in correction time**.

BTG Pactual
Sao Paulo, SP, Brazil (Remote)

Data Science Intern
Sep 2021 - Jul 2022

- **Conducted a detailed analysis of computational resource usage on AWS EMR clusters, using unsupervised learning on time series data** to detect overloaded and underloaded clusters, **resulting in a 30% reduction in AWS EMR usage expenditures**. Visualized usage of clusters with an interactive Grafana dashboard that collected data with PostgreSQL using an AWS Lambda function.
- **Developed a multi-class document classifier using a combination of AWS' OCR and text-processing techniques** to differentiate between front and back for 8 different document types. Employed computer vision techniques and text-entropy calculations to infer image quality and improve classification accuracy. **Built an interactive Grafana dashboard with AWS Aurora and PostgreSQL to monitor classifier performance over time**, retrain it with new data when necessary, and inform decisions for future enhancements.
- **Designed a liveness detection system using facial landmarks and image features to differentiate between real faces and fake ones**. Utilized scikit-learn classifiers with custom features and trained the model with a diverse dataset of real and fake faces. Deployed the system using AWS Lambda functions and Docker containers, storing image data in S3 buckets and using DynamoDB (NoSQL) to manage the data. The system **achieved a recall score of 0.9986 for attacks, even with low-quality images and videos.**

University of California San Diego - Engineers for Exploration
San Diego, CA, USA (Remote)

Research Intern
Nov 2021 - Jan 2022

- **Implemented super resolution generative adversarial neural networks in PyTorch** to improve quality of low-cost satellite images for mangrove segmentation using a U-Net architecture, **significantly outperforming the IOU achieved when using low-resolution images**.

Google Summer of Code - Machine Learning for Science
Remote

Research Intern
May 2021 - Aug 2021

- **Developed a deep learning framework in PyTorch for unsupervised domain adaptation applied to gravitational lenses**, with a focus on adapting to new and unseen environments. Implemented 3 different domain adaptation algorithms, including convolutional and equivariant neural networks, and compared their results.
- **Published research findings in the Machine Learning and the Physical Sciences Workshop at NeurIPS 2022**, demonstrating the effectiveness of the proposed framework in simulation-based dark matter searches with strong gravitational lensing. Contributed to the open source Machine Learning for Science organization by making the code public.

Federal University of Rio Grande do Sul
Porto Alegre, RS, Brazil

Research Intern
Jan 2020 - Aug 2021

- **Created a pipeline for the detection of Low Surface Brightness Galaxies with an image-based DBSCAN algorithm and convolutional neural networks** applied to large astronomical images fetched from an online database. Published in Meetings of the Brazilian Astronomical Society 2021.

Education

Federal University of Rio Grande do Sul
Porto Alegre, RS, Brazil

Bachelor of Physics
Mar 2018 - Oct 2022

- Created a group to teach machine learning and computer science to other physics students.
- Undergraduate thesis achieved highest grade and was published in LatinX in AI Affinity Workshop at NeurIPS 2022.

Publications

- **Domain Adaptation for Simulation-Based Dark Matter Searches with Strong Gravitational Lensing**. Stephon Alexander and Michael W. Toomey, Sergei Gleyzer, Pranath Reddy, Marcos Tidball. Neural Information Processing Systems (NeurIPS) Conference: Machine Learning and the Physical Sciences Research Workshop, 2022.
- **Similarity Search of Low Surface Brightness Galaxies in Large Astronomical Catalogs**. Marcos Tidball, Cristina Furlanetto. Neural Information Processing Systems (NeurIPS) Conference: LatinX in AI (LXAI) Research Workshop, 2022. `https://doi.org/10.52591/lxai202211282`
- **Domain Adaptation for Simulation-Based Dark Matter Searches Using Strong Gravitational Lensing**. Stephon Alexander, Sergei Gleyzer, Pranath Reddy, Marcos Tidball, Michael W. Toomey. arXiv preprint, 2021. arXiv:2112.12121.
- **Deepfuse: Automatic Detection and Classification of Low Surface Brightness Galaxies with Convolutional Neural Networks**. Marcos Tidball, Cristina Furlanetto. Meetings of the Brazilian Astronomical Society (SAB), 2021.

Projects

- lsh-astro: **Implemented Locality-Sensitive Hashing (LSH) in PySpark for the approximate similarity search of Low Surface Brightness Galaxies** (LSBGs) in large astronomical catalogs. Utilized a dataset composed of the properties of 11,670,190 objects, developing a tool that is able to find new LSBG candidates using only one known LSBG. **Presented the findings at the LatinX in AI (LXAI) Research Workshop at NeurIPS 2022**, contributing to the advancement of AI research in astronomy.
- AdaMatch-pytorch: **First PyTorch implementation of AdaMatch**, an algorithm that unifies Domain Adaptation and Semi-Supervised Learning.
- ezML: Native Java implementation of different Machine Learning algorithms, CSV reader and a pre-processing class that performs train-test splits and normalizes the data.
- Unsupervised Adversarial Autoencoder on Gravitational Lenses: Implemented an adversarial autoencoder architecture in PyTorch to differentiate between gravitational lens images with and without dark matter substructure. Achieved AUC of 0.74.
- HR-Diagram Python: Used pandas, NumPy and Matplotlib to analyze star evolution data, separating each stage of the star life cycle.

Skills

Programming	Python (NumPy, pandas, Matplotlib, scikit-learn, PyTorch, Boto3, OpenCV, TensorFlow, SciPy, MMDetection, AstroPy), SQL (PostgreSQL), git, AWS (Lambda, S3, API Gateway, DynamoDB, EMR, RDS, EC2, Athena, Textract, Rekognition), Grafana, Docker, Apache Spark (PySpark, MLlib), C, Java, Latex, Excel.
Languages	Portuguese (fluent, native), English (fluent, 8.0 on IELTS Academic), Spanish (conversational level).
Soft Skills	Research background, strong collaboration skills, prototyping skills, problem-solver, communicative.

Honors

2022	**LatinX in AI travel grant**, LatinX in AI Affinity Workshop at Neural Information Processing Systems (NeurIPS)	*USA*