Murilo **Gustineli**

murilogustineli@gmail.com • LinkedIn • GitHub • Website • 502-489-7705 • Raleigh, NC

Summary & Skills

Data Scientist with 3+ years of experience designing statistical machine learning solutions and implementing data engineering pipelines. Currently working full-time and pursuing a master's in computer science specializing in machine learning at Georgia Institute of Technology. Passionate about machine learning, deep learning, reinforcement learning, computer vision, NLP, and AI research

- **Programming Languages:** Python, SQL
- **Tech Stack:** Spark, Git, Databricks, Azure, GCP, AWS, MLOps, MLflow, Docker
- **Libraries:** Pandas, NumPy, Scikit-Learn, TensorFlow, PyTorch, Matplotlib, NetworkX, GraphFrames, OpenCV, Prophet, NLTK, SciPy
- **Research:** Google Scholar, Reinforcement Learning, Hyperparameter Tuning, Unsupervised Learning, Activation Functions for DL

Professional Experience

Insight Enterprises, Inc. ***Raleigh, NC***

Data Scientist II *September 2022 – Present*

- Developed a POC community discovery engine using network science, achieving 92% similarity between individuals' skill sets
- Discovered network clusters using GraphFrames and cosine similarity, enabling tailored project-based skill recommendations
- Leveraged the network to discover key opportunities to strengthen team skills and provide mentorship recommendations
- Initiated a patent application for cluster discovery using cosine similarity, based on individual skill level and project requirements

Data Scientist I *June 2021 – August 2022*

- Built an image classifier with Azure Cognitive Services, detecting anomalies with precision of 0.93/0.37 on training and test sets
- Developed a proposal for diagnostics on labeled images, improving labeling processes, and identifying data-quality issues
- Created a data pipeline using PySpark on Databricks for labeling 20 thousand images stored on AWS and Azure containers
- Implemented Feature Store and MLOps best practices, standardizing model training and testing procedures across various teams

Omdena ***New York, NY***

Jr. Machine Learning Engineer *January 2021 – May 2021*

- Built a computer vision model using PyTorch and torchvision, detecting pathologies in ultrasound images to help doctors in Africa
- Achieved 95% accuracy in distinguishing between benign and malignant tumors in pneumonia and breast cancer images
- Implemented Mask R-CNN fine-tuning for object-detection and image segmentation, detecting cancer lumps in breast sonograms

Amazon Web Services, Inc. ***Seattle, WA***

Business Intelligence Engineer (Internship) *June 2020 – August 2020*

- Developed an ETL pipeline identifying a 4% week-over-week customer growth for services generating $120 million in weekly revenue
- Created a QuickSight dashboard using SQL, handling billions of records for product launch analysis, helping engineers and PMs

Research and Projects

Scalable LLMs: Explored scalability of language models ranging from 3M to 169M parameters using the RWKV architecture and TinyStories dataset. Found that models with as few as 10M parameters can produce coherent and contextually relevant stories. Notably, models achieved BLEU scores between 0.04 to 0.049, and ROUGE-1 scores around 0.28, indicating a balance between original content generation and coherence. This demonstrates the potential for deploying language models in resource-constrained environments

BirdCLEF 2023: Developed a TensorFlow and PySpark-based method for classifying 264 African bird species in 192 hours of audio, achieving accuracies of 0.79 on training and 0.68 on test sets. Leveraged transfer learning with BirdNET and MixIT models for semi-supervised dataset annotation and extracting audio embeddings. Implemented multi-layer perceptron and gradient-boosted decision tree classifiers via scikit-learn and XGBoost. Performed hyperparameter search using sci-kit-optimize Bayesian optimization

Time Series Forecasting: Developed a univariate time series forecasting model using Prophet and PySpark, which accurately predicted seasonal patterns in apartment prices with an RMSE of 21.08 and MAE of 17.22, yielding annual rent savings of up to $1000

NLP Classification: Created a Random Forest classifier utilizing Python, Machine Learning, and Natural Language Processing techniques, which accurately predicted positive and negative sentiments in over 3,000 Amazon Alexa reviews with 94% precision

Education

Georgia Institute of Technology ***Atlanta, GA***

M.S. Computer Science; Machine Learning *August 2022 – Present*

University of Nevada, Las Vegas ***Las Vegas, NV***

M.S. Management Information Systems *December 2020*

B.S. Business Administration; International Business *December 2017*