Mughilan Muthupari

mughilan.muthupari@gmail.com
240-779-6852
www.linkedin.com/in/mughil-pari/
www.github.com/MughilM

PROFILE

I am a **data scientist** with more than 6 years of experience in machine learning and MLOps in the fintech, biotech, and climate domains, with specialized experience in natural language processing and image recognition. I am looking to be a senior data scientist as the next step of my career. I can bring my AI/ML expertise to your company and implement new AI workflows in your company's products and further your goals.

SUMMARY

- Expert in many machine learning models and concepts, including NLP, image recognition, and time series
- Built products following full SDLC and agile methodologies, including Jira for issue tracking
- Extensive experience in designing and implementing in cloud platforms, including AWS and Google Cloud Platform (GCP)
- 4+ years of work experience in DevOps and MLOps to monitor model training and post deployment metrics

SKILLS

- **ML technologies:** PyTorch (incl. PyTorch Lightning), Tensorflow 2.x, numpy, pandas, scikit-learn, Jupyter
- **Specializations and Interests:** NLP, LLMs, image recognition and segmentation (UNet), reinforcement learning, model interpretability
- **MLOps experience:** MLFlow, Weights and Biases, AWS Sagemaker, Apache Spark, Jenkins, Github Actions
- **Programming Languages:** **Python**, R, SQL, Java
- **Cloud Platforms:** AWS, Google Cloud Platform

EDUCATION

M.S / Data Science — Columbia University / New York, NY

Sep 2019 – Dec 2020

B.S / Computer Science & Statistics — University of Maryland – Gemstone Honors / College Park, MD

Aug 2015 – Dec 2018

WORK EXPERIENCE

Machine Learning Engineer

Meta / New York, NY / Apr 2024 – Present

- Currently on the Social Impact Ecosystems (SIE) team working on reducing prevalence of violating content towards teens across Facebook, Instagram, and Threads

Data Scientist II

Battelle / Columbus, OH / Mar 2022 – Apr 2024

- Designed a CV model to examine biological pathways in corn using large images, and evaluated SOTA CV models with MLFlow
- Led analysis of employee data for workplace shocks as Model Lead
- Modeled evolving bioplumes using anomaly detection (PyG) with graph neural nets and MLPs to detect changes in smoke concentration
- **Technologies:** Python, Java, PyTorch, MLFlow, convolutional neural network (CNN), ViT, graph neural network (GNN), scikit-learn

Machine Learning Engineer/Developer

FINRA / Rockville, MD / May 2019 – Mar 2022

- Presented and visualized results of explainability techniques for risk predictions to business stakeholders as team lead
- Developed and maintained a risk advert ML system in EC2 and Sagemaker using CNNs and transformers
- Integrated data pipelines and storage in S3 using Apache Spark to support model training data versioning
- Performed abstractive document summarization of long PDFs using BERT LLM from Huggingface and Amazon Textract
- **Technologies:** Python, Java, PyTorch, AWS EC2, Sagemaker, BERT, LongFormer, CNN, LIME, Shap, Spark, Agile, Git, scikit-learn

INTERNSHIPS

Moody's Investors Service / New York, NY / Jun 2020 – Aug 2020

- Optimized labels in a weak supervision NER task by generating refined annotation labels, reducing labor costs by 70%
- Devised heuristic labeling functions using Snorkel to reduce burden on human labelers
- **Technologies:** Python, NLP, NER, Snorkel

FINRA / Rockville, MD / May 2018 – August 2018

- Predicted complaint types of disclosures, and classified enforcement documents using CNNs and LSTMs.
- Models were built using both TensorFlow and PyTorch

NASA Center for Climate Simulations / Greenbelt, MD / Jun 2017 – Aug 2017

- Assessed the Advanced Data Analytics Platform (ADAPT) to examine historic daily temperature cycle and other statistical calculations
- Compared calculations across different systems to detect discrepancies and presented findings
- **Technologies:** Python, Jupyter, pandas, numpy

FINRA / Rockville, MD / May 2016 – August 2016

- Modeled time series trading data from the BATS exchange using Java with the goal of analyzing possible patterns in actual trading data

PUBLICATIONS AND OTHER PROJECTS

Where's the Learning in Representation Learning for Compositional Semantics and the Case of Thematic Fit – Columbia University

- Improved sentence word structure through thematic fit by utilizing roles and words using **TensorFlow.**
- Measured effectiveness of various word embeddings (XLNet, FastText, Glove) and network structural changes on performance and evaluation using Google Cloud Platform (GCP) systems
- Enhanced legacy codebase to improve preprocessing performance and pipeline efficiency
- **Continued research after graduation resulted in paper published at the Conference on Empirical Methods in Natural Language Processing (EMNLP) Blackbox and ready on arXiv: https://arxiv.org/abs/2208.04749**
- **Repository is at https://github.com/MughilM/RW-Eng-v3-src**

Wildfire Incidence using Spatial Logistic Regression – Columbia University

- Forecasted wildfire incidences across the United States using historic wildfire incidence, weather, and land use data.
- Logistic regression was used to determine the probability of a wildfire occurring in a specific 5-minute resolution cell (~6 miles) using past climate data, and nearby wildfire incidences and land use data up to 30 miles away.
- Performance was shown to be ~15% better than null accuracy.
- **Notebooks and final report is at my AcaDS repo under INAF6506/Final Project/**

An Immersive Experience: Visualizing Large-scale Climate Data using Virtual Reality and Infrared Hand-tracking Technology – University of Maryland

- Created a novel climate visualization tool constructed with virtual reality in a team of multi-disciplinary students using **Python and Unreal Engine.**
- Awarded the James M. Wallace Outstanding Gemstone Thesis Award. Available for reading at the Digital Repository at the University Maryland (DRUM): **https://doi.org/10.13016/wgrz-tatt**

CERTIFICATIONS AND CREDENTIALS

Python for Financial Analysis and Algorithmic Trading – November 2017 – Credential