THANAY GEESHPATHY NARAYANAMURTHY

San Jose, CA • thanay.murthy@gmail.com • (774) 365-7907 • LinkedIn • GitHub • Portfolio
(Open to Relocate)

"Diagnose with data, treat with design". Experienced **Data Engineer** and **Analytics Professional** with over **4 years** of industry experience and a Master's degree in **Data Science**, specializing in **Python, SQL, AWS, Machine Learning and NLP**. By developing and deploying intelligent data solutions, I help businesses uncover insights, build fantastic data products, and make data-driven decisions. Actively looking for new opportunities.

AWARDS, RECOGNIZATION, CERTIFICATIONS

- Generative AI with Large Language Models - (in progress)
- Google Advanced Data Analytics Professional Certificate - (in progress)

Associated with the University of Massachusetts

- 1st Place - Techstars Startup Weekend at New Bedford - Hackathon 2024 - link
- 1st Place - University of Massachusetts Dartmouth - Hackathon 2024 - link

Associated with Finflux

- Awarded Certificate of Excellence for my data migration and engineering work while onboarding clients onto the Finflux product platform

PROFESSIONAL EXPERIENCE

UNIVERSITY OF MASSACHUSETTS DARTMOUTH — **North Dartmouth, USA**
Graduate Research Internship - Artificial Intelligence — **Jun 2023 - Aug 2024**

Project 1: Deep Learning for Gravitational Waves Inference

The main goals of this project are to understand the math, develop appropriate analysis, and finally implement a robust and efficient guide for gravitational wave parameter estimation using Deep Inference of Gravitational Waves Observatory data.

- Generation and curation of the training datasets using **Python** and **YAML**, of pairs of intrinsic parameters like masses and spins, waveform polarizations, and noise power spectral densities for **feature-extraction (SVD)**
- Read and analyze waveform data, converting **HDF5** to **numpy** array and **pandas** dataframes, plot using **matplotlib**
- **Neural network** construction and training to model the Bayesian posterior distribution over the wave parameters
- Training involved the simulation of data sets that would mimic the noisy signals of gravitational waves detected by interferometers, and the **deep learning** model is optimized to infer the parameters from such signals precisely
- Used **PyTorch** for the neural network architecture design and **CUDA** with **Nvidia GPUs** for training
- Generalize the trained model to real or simulated observational data, using dingo-**APIs**, to enable fast and accurate inference of the parameters of GWs
- Investigate how well the model performs and scales when deployed at the production level on **AWS**
- Created analysis plots using **Tableau** dashboard, from data stored in **MySQL** database.
- Cut the inference time for real world GWs by over 90%

Project 2: Task Scheduling for Multi-Agent Systems

Develop and optimize autonomous agents for USAR missions in RoboCup Rescue Simulation (RCRS) testbed, enhancing multi-agent coordination and decision-making algorithms.

- Conducted research on **TAEMS** framework for multi-agent systems, a sophisticated task-scheduling framework to boost agents' collective intelligence, operational adaptability, and decision-making efficacy in dynamic, uncertain map environments
- Worked on different evaluation tests and **reinforcement learning** algorithms to make the framework robust
- Rewrote modules in **Python**, using clean and efficient code, improving the speed of execution by 10%

INDEGENE **Bengaluru, India**
Role: Sr. Data Engineer - Data & Analytics **Jan 2022 - Aug 2022**

Project 1: Oncology Product Research

Determine the sales and distribution of Pfizer medicines at various locations across the United States, to build targeted marketing campaigns for physicians. Conduct a study for the oncology product, tracking key metrics into effectiveness.

- Used **ML** models to predict physicians for new drugs, boosting prescription rates by 10% through targeted marketing strategies.
- Optimized inventory across various locations using **data visualizations**, leading to a more balanced stock distribution and reducing stock outs by 15%
- Analyzed EHR and pharmacy claims data using **Tableau**, for over 100 patients to assess drug effectiveness and side effects, providing insights that informed clinical decisions and improved patient outcomes
- Developed and optimized **MySQL** scripts in **Snowflake** to generate sales and distribution datasets, reducing data processing time by 13%
- Developed **Tableau** dashboards to aid the sales team in targeting physicians, resulting in a 6% boost in sales by providing actionable insights and data-driven recommendations

Project 2: Global Content Origination Analytics
Client: Pfizer, USA

Determine bottlenecks and optimize Pfizer's content origination workflow to improve the volume, velocity, and quality of published content.

- Streamlined existing **ETL** pipelines by optimizing workflows, reducing job runtimes by 20%
- Updated **data pipelines** based on change requests, ensuring they remained aligned with evolving operational changes
- Integrated data from multiple sources into a centralized **data warehouse**
- Implemented **data validation** to ensure data accuracy, using **DSS DataIku**.
- Optimized **Python & MySQL** code within ETL pipeline to ensure faster retrieval by 9%
- Developed analytics dashboards in **PowerBI** to help increase content generation by 16%

Project 3: Dry Eye and Punctal Plug Market Research
Client: Alcon, Switzerland

Conduct predictive analytics of the market share of eye care products in the United States, to determine the viability in selling niche eye care products and acquisition of assets.

- Collected and managed large volumes of market data from various sources using **AWS S3** for storage
- Automated data extraction tasks with **Apache Airflow**, reducing manual processing time by 20%
- Used **Athena** for querying data and preparing datasets
- Performed EDA using **SQL**, addressing inconsistencies and missing values, resulting in a 10% reduction in errors
- Developed **ML** predictive models with 90% accuracy in market share forecasts, guiding the successful acquisition of a new product manufacturer.
- Analyzed market trends and consumer behavior using **Quicksight**, identifying a 15% growth opportunity

FINFLUX **Bengaluru, India**
Role: Associate Product Manager **Jul 2021 - Dec 2021**

Crafted and executed a product enhancement strategy for a leading finance lending platform, including market analysis and user feedback integration, while leading proposals for new features and securing ongoing client partnerships

- Led a team of 10 professionals for managing all aspects of the **data analytics** lifecycle, using **Agile & sprint planning**
- Proposed analytics tools to enhance **data-driven** decision-making, effectively communicating these proposals to stakeholders, which led to the adoption of 2 new advanced analytics solutions within the company
- Used **Figma** and **wireframes** to effectively communicate product designs to stakeholders, ensuring alignment
- Utilized **customer analytics** to identify and address pain points, successfully cutting change requests by 15%, leading to more stable product releases

- Defined and tracked key customer metrics using **Metabase**, which improved customer satisfaction by 25% through data-driven product enhancements
- Streamlined internal communication and documentation processes using **Confluence**, reducing employee training times by 50%, and improving knowledge sharing across the team
- Implemented improvements on **Freshdesk**, which led to a 30% reduction in monthly support requests by automating responses and improving self-service options
- Gathered requirements from clients and authored **Business Requirement Documents** (BRDs), also monitored & reported **SLAs**, ensuring clear communication of project goals and expectations
- Onboarded **3 new clients** to the product platform, providing comprehensive training and support, which facilitated smooth transitions and higher client retention
- Delivered **4 product demos**, effectively showcasing product features and value to prospective clients
- Developed and integrated **new features** such as Bulk Disbursal, Bulk Approval, and Aadhar Vault, enhancing system functionality and compliance
- Worked with **cross-functional teams** to release new features from conception to deployment, ensuring timely delivery
- Prioritized bugs, features, and tasks for multiple clients using **JIRA**, which ensured critical issues were addressed promptly

Role: Data Engineer **Feb 2019 - June 2021**

Engineered and optimized data solutions for a financial lending software company through workflow automation and infrastructure management, leading to enhanced system performance and client engagement

- Optimized **MySQL** scripts to enhance backend performance, achieving a 25% increase in query efficiency
- Automated **Rest API** calls and data transformations with Pentaho, scaling loan service capabilities by 10x
- Managed **AWS** infrastructure, including **EC2**, **RDS**, and **CloudWatch**, which reduced downtimes by 15% and improved incident response times by 12%
- Managed client onboarding through multi-tenant configuration within **MySQL** databases, reducing setup costs by 25%
- Worked with **Postman APIs** testing to implemented 3rd party integrations using features like biometrics, geo-location, facial recognition, and payment gateways, expanding system functionality and ensuring compliance
- Automated deployment of web applications using shell scripts, **CI/CD** pipelines, **Jenkins**, and **Git**, and managed **AWS IAM** resources, which reduced server update downtimes by 30%
- Developed stored procedures and triggers in **MySQL** as well as data patches o fix issues caused by code bugs, thereby automating events and error corrections, reducing response times by 20%
- Created a data warehouse for scheduled reporting using **PostgreSQL** and **AWS Redshift**
- Analyzed Fineract application data to identify root causes of issues, using performance logging on **Apache Cassandra**, and developed correction plans, reducing issue recurrence by 12% and to improve system reliability
- Developed and tested Pentaho reports and created **Tableau dashboards**, increasing operational visibility
- Developed criteria checks in loan approval workflow stages using **Machine Learning** on **Amazon SageMaker** improving loan processing time by 30% and reducing delinquency rate by 5%
- Built and managed bulk repayment schedule changes using **RESTful APIs** and **Pentaho Data Integration Tool** (kettle), automating processes and reducing manual intervention by 40%
- Engineered Loan Management System workflows for **Loan Origination** and **Loan Approval**, cutting costs by 14%

PROJECTS

Gravitational Wave Parameter Estimation using Deep Learning - **link** *[Python, SVD, GPU Computing, Neural Networks]*

- Parameter estimation using neural posterior estimation to make fast inferences
- Used PyTorch to define and CUDA to train neural network architecture on GPUs

Spotify Playlist Enhancer - **link** *[Python, Machine Learning, Flask, HTML, CSS, Flask, Recommendation System]*

- Created a webpage and recommendation system that takes a Spotify playlist link from users and generates up to 30 songs.
- Used Python and Flask for the backend and HTML5 and CSS for frontend

Player Position Predictor - **link** *[R, Machine Learning, KNN-clustering, LDA]*

- Implemented ML models KNN-clustering and linear discriminant analysis (LDA) in R
- Validated model for predictions with 90% accuracy

World Energy Consumption Visualized - **link** *[Data Visualization, Git, HTML5, D3.js, CSS]*

- Project uses interactive visualizations to explore the dynamic connection between global energy consumption and economic prosperity. Offered insights through D3.js charts and Our World in Data's dataset

Face recognition using eigenfaces - **link** *[Python, Machine Learning, Numpy, Pandas, MATLAB, OpenCV, CuPy, PCA]*

- Built a facial recognition program using principal component analysis (PCA) dimensionality reduction, using OpenCV

E-commerce platform data analytics using AWS *[Amazon RDS, S3, Glue, DataBrew, Tableau]*

- Built an end-to-end data ETL pipeline on AWS
- Used AWS services RDS, S3, and DynamoDb for storage, Glue for batch processing, and DataBrew for transformations

AI 8 tile sliding puzzle solver - **link** *[Python, Artificial Intelligence]*

- Implemented an AI program that solves the 8-tile sliding puzzle game
- Uses three different search strategies; BFS, Greedy, and A-star to arrive at a solution

SKILLS

- **Programming Languages:** Python (NumPy, Pandas, Matplotlib, Scikit-learn, PyTorch, Flask), SQL, R, C++
- **Cloud & Databases:** MySQL, Snowflake, MongoDB, Redshift, DynamoDB, Neo4j, S3, EC2, Airflow, Pentaho (kettle), Kafka, Glue, EMR, Hadoop, Hive, Cassandra, Sagemaker, GraphQL
- **Tools & Libraries:** Excel, PowerPoint, JIRA, Confluence, Git, Kubernetes, Docker, Jenkins – CI/CD, Postman API, Flask, Django
- **Data Visualization:** Tableau, PowerBI, Pentaho reports, Metabase, Quicksight
- **AI and Machine Learning:** Transformers, LLMs - finetuning and integration, Pinecone, Langchain, Supervised ML (Linear and Logistic Regression, Decision Trees, RF, Xgboost, KNN, LDA), Unsupervised ML (K-means, PCA), NLP, Classification techniques (binary classification, multi-class classification), Statistical Analysis, Hypothesis Testing, A/B testing

EDUCATION

UNIVERSITY OF MASSACHUSETTS DARTMOUTH — **Dartmouth, US**
MS in Data Science (GPA: 4.0) — **Aug 2024**

DAYANANDA SAGAR COLLEGE OF ENGINEERING — **Bengaluru, India**
B.E in Electronics & Communication Engineering — **Jan 2019**