LinkedIn
GitHub
Google Scholar

Sibi V. Thirukonda

Open to Relocation; Start immediately

thirukonda.s@northeastern.edu
Website
+1-6092332195

Experience

- **Squark AI - Automated AI ML Solutions** — Boston, MA
  Machine Learning Engineer; Python, CUDA, SQL — *Mar 2025 - Present*
  - **Research on Advanced Time-Series Modeling**: Investigated and developed ***state-of-the-art*** time-series forecasting techniques, including ***ARIMA***, ***GARCH***, ***N-BEATS***, ***Temporal Fusion Transformers (TFT)***, and ***self-supervised learning***, improving predictive accuracy on financial datasets by ***30%***.
  - **Exploration of New ML Architectures**: Researched and benchmarked novel ML models, including ***graph neural networks (GNNs)***, ***normalizing flows***, and ***diffusion models***, for capturing complex dependencies in structured and unstructured financial data.
  - **Explainable AI (XAI)**: Integrated ***SHAP*** and ***LIME*** for model interpretability, enhancing transparency in financial predictions and improving trust in AI-driven asset allocation by ***25%***.
  - **Feature Engineering and Representation Learning**: Developed automated feature engineering pipelines leveraging ***self-supervised learning*** for financial time-series data, improving generalization and robustness against market volatility.
  - **Scalable ML Research Pipeline**: Designed and optimized ***scalable ML research pipelines***, enabling rapid experimentation and evaluation of cutting-edge ML methodologies for financial applications.

- **John Hancock - Manulife Investment Management Division** — Boston, MA
  Machine Learning Engineer; Python, CUDA, SQL — *Jan 2024 - Jan 2025*
  - **Research on Transformer-Based Risk Modeling**: Investigated ***transformer architectures*** for financial risk assessment, leveraging ***self-attention*** and ***sequence modeling*** to improve forecasting accuracy, reducing valuation errors by ***20%***.
  - **Efficient Monte Carlo Simulations with GPU Acceleration**: Researched and implemented ***parallelized Monte Carlo methods*** using ***CUDA*** and ***TensorFlow XLA***, reducing risk model simulation runtime from 30 minutes to ***15 seconds***.
  - **Bayesian Inference for Portfolio Risk Modeling**: Developed ***Bayesian deep learning models*** for probabilistic risk estimation, incorporating ***stochastic volatility processes*** to optimize capital allocation and reduce financial exposure by ***15%***.
  - **Survival Analysis Using Deep Learning**: Designed and evaluated ***neural survival models*** using ***recurrent*** and ***attention-based architectures*** to predict policyholder lapse rates, improving risk estimation by ***25%***.

- **Morgan Stanley - Capstone Project, Parametric Portfolio** — New York, NY
  Quantitative Researcher; Python, R — *Aug 2024 - Dec 2024*
  - **Identified Key Investment Opportunities**: Pinpointed ***Singapore*** and ***China*** as top investment destinations, enabling sector-specific insights that drove ***25%*** better portfolio allocation strategies.
  - **Improved Forecasting Accuracy**: Enhanced ***GDP*** and sector trend predictions by ***30%*** across nine countries using ***model stacking*** and ***advanced imputation techniques***.
  - **Quantified Market Growth Opportunities**: Developed a ***Bayesian predictive framework*** to analyze economic indicators, uncovering ***$10B+*** potential investment opportunities in ***emerging markets***.

- **Northeastern University - Research Computing** — Boston, MA
  Research Analyst; Python, R, CUDA — *Sep 2023 - Dec 2024*
  - **Research on LLM-Based Answer Recommendation**: Investigated and developed ***LLM-based retrieval-augmented generation*** techniques, optimizing response contextuality and accuracy for domain-specific queries. Enhanced knowledge retrieval by ***40%***.
  - **Preference Alignment via RLHF**: Researched and implemented ***Reinforcement Learning from Human Feedback (RLHF)*** for fine-tuning LLMs, improving preference alignment and response coherence by ***35%*** using scalable feedback mining frameworks.
  - **Multi-Step Reasoning with Autonomous LLM Agents**: Designed ***LLM-based autonomous agents*** with ***multi-hop reasoning*** capabilities, integrating ***self-reflection*** and ***tool use*** to improve task planning and execution. Increased automation efficiency by ***50%***.
  - **Efficient Fine-Tuning of Large Language Models**: Explored and applied ***parameter-efficient fine-tuning (PEFT)*** techniques such as ***LoRA***, ***QLoRA***, and ***AdapterFusion***, reducing computational cost while maintaining model adaptability.

- **Lennox International - Samsung America** — Chennai, India
  Software Engineer; C++, Java — *Aug 2022 - Dec 2022*
  - **Integrated Smart Assistants for HVAC Control**: Developed ***middleware*** for seamless integration of ***Apple HomeKit*** and ***Amazon Alexa*** APIs with HVAC systems, improving interoperability by ***30%***.
  - **Enhanced Distributed System Optimization**: Optimized ***multi-threaded communication protocols***, reducing latency by ***40%*** and increasing system responsiveness.
  - **Low-Latency Edge-Based Multimodal Processing**: Implemented efficient ***multimodal models*** for real-time speech and vision data processing on ***edge devices***, reducing inference times by ***20%***.

- **Madurai Smart City - Industry Institute Partnership Cell & Capstone** — Madurai, India
  Senior Computer Vision Researcher; Python, C++, Vulkan, CUDA — *May 2020 - May 2022*
  - **High-Performance Inference for Vision Models**: Deployed ***computer vision models*** on ***NVIDIA Xavier***, leveraging ***TensorRT optimization*** to achieve ***3x faster inference*** for real-time face mask detection.
  - **Scalable Distributed Computing for Video Analytics**: Designed ***distributed computing architectures*** to process ***500+ concurrent video feeds***, improving real-time surveillance through ***adaptive bitrate streaming***.
  - **Efficient Object Detection Model Compression**: Implemented ***quantization-aware training (QAT)*** for ***YOLO-based object detection***, achieving ***30% faster inference*** on edge devices.

Education

- **Northeastern University** — Boston, MA
  Master of Science in Data Science; *GPA: 3.73; Courses: Machine Learning, NLP, Geo-Spatial Analytics* — *2024*

Leadership and Achievements

- **Best Project Winner 2018-2022 Batch**: Developed a **COVID-19 monitoring system** that enhanced city-wide response by **50%**.
- **IoT Traffic Management Leader**: Created an traffic system that reduced emergency vehicle delays by **50%** using routing algorithms.
- **Blockchain-Based Auction Innovations**: Published Ethereum-based auction mechanisms, improving transaction efficiency by **20%**.
- **IoT Patent Contributor**: Contributed to a **patent for sand moisture IoT sensors**, enhancing accuracy by **15%** with ML integration.