

SATWIK NALLAMILLI

☎ 669-224-5382 ✉ ssn9@illinois.edu 🔗 [linkedin.com/in/satwik-reddy](https://www.linkedin.com/in/satwik-reddy) 🐙 github.com/SaiSatwikReddy07

Education

University of Illinois Urbana-Champaign

Aug 2025 – May 2027 (Expected)

Master of Science in Statistics

Relevant Coursework: Advanced Data Analysis, Statistical Modeling, Statistical Inference

Technical Skills

Frameworks/Libraries: NumPy, Pandas, Apache Spark (PySpark), Scikit-learn, PyTorch, Hugging Face, TensorFlow, LangChain, CrewAI, Chroma

Specializations: Machine Learning, Data Engineering, Data Analysis, Natural Language Processing, Generative AI, Big Data, Agentic AI, Retrieval-Augmented Generation (RAG)

Programming Languages: Python (Advanced), SQL (Advanced), Java (Intermediate)

Tools & Cloud: Docker, Git, Jira, Databricks, Tableau, AWS (S3, Kinesis, Glue, Athena, Lambda, EC2, QuickSight)

Experience

Volkswagen Group

Sep 2024 – Feb 2025

Data Science Intern | Python, SQL, PySpark, AWS, REST APIs, Postman

Pune, India

- Developed an end-to-end **HR analytics solution** on AWS to support workforce decision-making for **70,000+ employees**, integrating OAuth-secured data from **SAP SuccessFactors** and internal enterprise sources.
- Engineered a robust **ELT pipeline** using API-based data ingestion, **AWS Glue with PySpark scripts** for scalable and efficient data transformations, and **Amazon S3** for storage of raw and processed datasets.
- Accelerated data querying by **30%** with optimized **Amazon Athena** queries and partitioning strategies, and built a **near real-time**, interactive dashboard in **Amazon QuickSight** for HR analytics.

Amrita Hospital

Nov 2023 – Feb 2024

AI Intern | Python, PyTorch, NumPy, Pandas, Matplotlib, GANs

Kochi, India

- Built and integrated a **PyTorch-based image classifier** into a healthcare web app, enabling automated diagnosis support using **25,000+ clinical images**.
- Designed and implemented a complete **data preprocessing pipeline** including cleaning, normalization, resizing, and artifact removal, to prepare high-quality training data.
- Fine-tuned **ResNet**, **EfficientNet**, and **Vision Transformer (ViT)** on **GAN-augmented datasets**, boosting classification accuracy by **20%**.

Projects

Depression Detection using NTK and KANs | *Python, PyTorch, Scikit-ntk, Pykan, NumPy, Matplotlib*

Apr 2025

- Built a depression detection model using **10,000+ Reddit posts**, applying TF-IDF feature extraction and training with **Neural Tangent Kernels (NTK)** and **Kolmogorov-Arnold Networks (KANs)**.
- Outperformed SVM and transformer-based baselines by **7.2% in F1-score**, demonstrating the potential of emerging architectures in NLP classification tasks.

AI-Powered Trading Intelligence Platform | *Python, CrewAI, ChromaDB, Streamlit, LSTM, TensorFlow*

Jun 2025

- Developed a **multi-agent AI trading system** using CrewAI orchestration with 4 specialized agents, integrating **5+ data sources** (Yahoo Finance, news APIs, Reddit, Fed data) and implementing RAG-based knowledge retention with vector embeddings.
- Architected modular system following **SOLID principles** with LSTM price prediction and sentiment analysis, enabling real-time market intelligence and automated signal generation through a Streamlit dashboard interface.

Selected Publications

Fast Iterative Filtering-Based DBN for Accurate Short-Term Electric Load Forecasting

ICICC 2024, Springer, doi: 10.1007/978-981-97-4149-6_35

Sep 2024

Transfer Learning-Based Emotion Recognition Using Augmented Speech Data

15th ICCCNT, IEEE, doi: 10.1109/ICCCNT61001.2024.10725674

Jun 2024

Enhancing Product Categorization in E-commerce using NLP and Machine Learning

8th ICICT, IEEE, doi: 10.1109/ICICT60155.2024.10544665

Apr 2024

Transfer Learning Approach for Differentiating Parkinson's Syndromes Using Audio

IACC 2023, Springer, doi: 10.1007/978-3-031-56703-2_18

Jun 2024

Accurate Estimation of Cargo Power Using Machine Learning Algorithms

Book Chapter, Springer, doi: 10.1007/978-3-031-72636-1_11

Jan 2025