Sai Nikhil Kunapareddy

Boston, MA | +1 (617)-818-2367 | kunapareddy.s@northeastern.edu | LinkedIn | Github

Available from January 2025 through August 2025

EDUCATION

Northeastern University **Boston, USA**
Master's of Science in Data Science; **GPA: 4.0** January 2024 – December 2025

Koneru Lakshmaiah University **India**
Bachelor's in Mechanical Engineering; **GPA: 3.9** July 2016 – May 2020

Coursework: Algorithms and Data Structures, Supervised Machine learning, Data Management and Processing, Deep Learning, Large Language Models, Artificial Intelligence, Machine Learning, Mathematics, and Statistics.

SKILLS SUMMARY

Programming Languages: Python, R, SQL, and C.
Libraries and Frameworks: PyTorch, TensorFlow, Hugging Face, Keras, Scikit-learn, XGBoost, NumPy, Pandas, Matplotlib, Seaborn, tidyverse, NLTK, spaCy, RegEx, dplyr, glmnet, ggplot2, Requests and BeautifulSoup.
Miscellaneous Tools: Spark, Kubernetes, Docker, Hadoop, MLFlow, MySQL, Tableau, Power BI, AWS and GCS.
Developer Tools: Jupyter, Git, Microsoft Excel and Visual Studio.
Data and Software Engineering: Agile methodology, Object oriented programming, CI/CD, Algorithm development, A/B testing, ETL, Data science pipeline (cleansing, wrangling, visualization, modeling, interpretation), Hypothesis testing.

PROFESSIONAL EXPERIENCE

Goodwill Computing Laboratory Boston, USA
Graduate Research Assistant July 2024 - Present

- Calibrated a large language model to identify spatial and temporal patterns in computational workflows as DAG executions, leading to improved wait times, enhanced energy efficiency, and reduced carbon emissions.
- Assembled a comprehensive data corpus from various sources using web scraping and natural language processing and employed statistical learning methods to accurately estimate and predict operational, embodied, and end-of-life carbon emissions.

ZF Group India
Senior Engineer January 2023 - July 2023

- Deployed an end-to-end material recommendation system using K-Nearest Neighbors, Bayesian optimization, Principal Component Analysis, and AWS SageMaker, enabling selections from over 150,000 variants.
- Trained and deployed a topology optimization deep neural network on-premise using PyTorch, reducing high-performance computing node usage to 7%.
- Fine-tuned a Hugging Face large language model on ISO and DIN standards datasets to build a custom chatbot capable of accurately answering industry-specific standards queries.
- Designed and implemented ETL pipelines to process one gigabyte of material data using Spark, SQL, and Power BI for real-time inventory control and material selection.

Engineer September 2020 - December 2022

- Conducted data wrangling, feature engineering, and exploratory data analysis on test data to predict assembly screw failures, implementing logistic regression and random forest models.
- Performed time-series forecasting using recurrent neural networks to predict system frequency behavior, reducing development costs by up to €13,000.
- Developed a robust inventory management system schema using MySQL, optimizing data storage, real-time tracking, and querying of over 2 million records.
- Created detailed and compelling data visualizations for documentation, reports, and dashboards using PowerPoint and Power BI, effectively communicating complex findings.

PROJECTS

TL;DR: Implemented and optimized a state-of-the-art text summarization tool for group chats, leveraging Hugging Face large language models such as BART and T5 to distill lengthy conversations into impactful, concise summaries. Employed TextRank and semantic filtering techniques to enhance summarization quality, ensuring clarity and relevance while improving named entity recognition to highlight frequently mentioned product attributes and issues.
Market Analysis of S&P 500: Leveraged advanced time-series forecasting techniques to predict the closing price of the S&P 500 index. Developed and compared models, including ARIMA, SARIMAX, and Long Short-Term Memory (LSTM) networks, employing rigorous statistical analysis and hyperparameter optimization to enhance prediction accuracy. Implemented ensemble methods to integrate model outputs, driving deeper insights and actionable investment strategies.