Kaushik Tummalapalli

+1 (917) 994-4435
github.com/kaushik-42/ kaushiktummala55@gmail.com kaushik-tummalapalli Portfolio

Education

New York University **Sep'22 – May'24**
Masters of Science in Computer Engineering, 3.9/4.0 GPA *Brooklyn, New York*
Courses: Natural Language Processing, Data Science, Big Data, Machine Learning, Deep Learning, AI

BV Raju Institute of Technology, India **Aug'18 – May'22**
Bachelors of Technology in Computer Science and Engineering, 8.38/10 GPA *Hyderabad, India*
Courses: Data Structures and Algorithms, Optimization, Object Oriented Design, Database Systems.

Skills

Languages/Domains: Python, R, SQL, Java, Applied Machine Learning, Generative AI, Time Series Forecasting, NLP
Developer Tools: Azure, Jenkins, Git, MLFlow, Docker, CI/CD, Rest API, Snowflake, Airflow, AWS, Bash
Libraries/Frameworks: PySpark, Flask, PyTorch, SciKit Learn, MLLib, Jupyter, Langchain, Pandas, Streamlit, RAG

Experiences

CVS Health **May'24 – Present**
Machine Learning Engineer - LLM's, Python, Microsoft Azure, OCR, Prompt Engineering, NLP, Testing

- Integrated GPT's multimodal capabilities into the Data Extractor, addressing untrained variations to boost **extraction coverage from 65% to 95%**, saving over thousands of hours annually by automating processing for 2M+ documents.
- Designed an evaluation strategy for GPT-based entity extraction to **refine acceptance criteria for LLM outputs**, leveraging custom metrics and advanced parsing to ensure accurate field-level extraction and workflow integration.
- Automated Cloudability reporting with **Azure SDKs**, reducing manual effort by over **30 hours monthly**.
- **Enhanced scalable AI solutions** by introducing asynchronous programming in Azure Function Apps, **reducing latency by 30%**, and enabling seamless high-volume request handling.
- Developed automation scripts to identify and redirect failed documents for re-training, improving model performance by 20% and enhancing adaptability across diverse document types.

Machine Learning Engineer Intern - Time Series Forecasting, Machine learning modeling *May'23 – August'23*

- Addressed cloud spend granularity and budgeting challenges at CVS Retail by deploying a time series forecasting model (Prophet) with Snowpark, **achieving 20% MAPE and 80% accuracy**, and enabling informed resource allocation.
- Conducted comprehensive experimentation with a range of time series models, including Prophet, ARIMA, SARIMA, and XGBoost, to forecast budgets for cloud compute and storage usage across various applications.
- Built a front-end application utilizing Streamlit with forecasting models integration, enhancing financial control.

New York University **Jan'23 – May'24**
Research Assistant - Data Scientist / TA

- Led a movie ratings analysis project under Professor Pascal, analyzing 10,000+ ratings to explore correlation structures.
- Built predictive models (linear regression, ridge regression, and logistic regression) with advanced feature engineering and hyperparameter tuning, **achieving a 15% improvement in accuracy, a 20% increase in R^2 for regression tasks**, and an **AUC score of 0.85 for classification tasks**.

Zemoso Technologies **Dec'21 – May'22**
Software Developer - React, JS, Java, MongoDb, GIT, Html, CSS, Code Coverage tools, CI/CD

- Designed and built a Blinkist-style web application, achieving an exceptional **code coverage of over 85%**, significantly reduced system vulnerabilities, and increased reliability using **React and Javascript**.
- Fostered a cross-functional team environment, applying agile methodologies to **reduce time-to-market by 20%** and **increase team productivity by 35%** with project debugging.
- Enhanced user engagement by implementing a responsive and intuitive interface using React and JavaScript, leading to a **50% increase** in average session duration within the first three months post-launch

Projects

Starbucks Recommendation System (Recommendation Systems) **Project Link**

- Implemented the FUNKSVD algorithm to recommend best-selling offers to current users and new customers, achieving a **Mean Squared Error** of **0.003823** for **15** latent features, and providing insights into customer purchasing behaviors.

Cab Cancellation Prediction Using Machine Learning **Project Link**

- Estimates the cancellation rate of a cab at the current time of booking based on historical data by EDA (Exploratory Data Analysis) with **85% accuracy** by using various ML Algorithms like Decision Trees, Logistic Regression, and SVM.