Arun Sathiya Vaithiyam Ramasami

San Francisco, California, US | +1 (510) 973-9100 | arun@arun.blog
github.com/arunsathiya | linkedin.com/in/arunsathiya | arun.blog

PROFESSIONAL SUMMARY

Experienced Data Engineer with expertise in statistical analysis, machine learning, and predictive modeling. Skilled in Python, R, SQL, and data visualization. Proven ability to derive insights from large datasets and inform strategic decisions

EDUCATION

Master of Science - MS, Data Analytics, California State University, East Bay **Dec 2023**

- Machine Learning, Big Data, Data Mining, DBMS, Data Warehousing & BI
- EC2/EMR, Hadoop, Redshift, Tableau, SPSS, Jupyter notebooks, PyTorch, TensorFlow, scikit-learn
- Capstone project: Langchain integration with OpenAI GPT 3.5-Turbo and Spotify Oauth for recommendations

SKILLS

- Python, R programming, SQL, Machine Learning, Deep Learning, LLMs (Large Language Models), AI Fine-Tuning
- Hadoop, Spark, Tableau, PowerBI, MongoDB, SQL, Amazon Redshift, Hive, Pig, Hadoop HDFS, Kafka, Confluent
- Qlora, Lora, Langchain, TensorFlow, PyTorch, AWS Sagemaker, Pandas, Numpy, Seaborn, Matplotlib, scikit-learn

WORK EXPERIENCE

Support Engineer, Automattic **Apr 2018 – Aug 2022**

- Managed 50,000+ production queries, demonstrating strong data-driven problem-solving skills
- Translated customer needs into technical specs, highlighting requirements analysis for data projects
- Developed Slack and Google Sheets integration in Node.js for data research, automating tasks and saving 20 hours weekly

PROJECTS

Langchain with LLM and Spotify for recommendations (Link) **Dec 2023**

- Integrated Langchain with Spotify OAuth for dynamic music recommendations, leveraging swappable LLM models
- Implemented and compared different LLMs, including Mistral 7B and OpenAI GPT-3.5 Turbo, to optimize recommendation accuracy
- Utilized Gradio to create a chat interface, offering an interactive chatbot experience for personalized music suggestions

Household Energy Consumption Forecasting (Link) **Dec 2022**

- Utilized R and packages (forecast, zoo) for time series analysis on a 2-year household energy consumption dataset from Kaggle
- Conducted data preprocessing, removing anomalies, and EDA to identify consumption patterns and trends
- Implemented and compared models: Naive Forecast, Tailing MA, Two-level Forecasting, ARIMA, Holt's-Winter
- Evaluated model performance using MSE and MAE, successfully forecasting future household energy consumption

Housing Market Analysis and Prediction (Link) **May 2023**

- Developed a Python-based tool to scrape housing data from a website using BeautifulSoup and requests
- Utilized pandas, numpy for data manipulation and performed exploratory data analysis to identify key market trends
- Analyzed factors influencing house prices: city-wise price increase, bedroom count, bath count, and price per sq ft.
- Conducted data visualization using Seaborn and Matplotlib to depict housing price distributions and city-wise averages
- Applied OLS Multiple Linear Regression for predictive analysis of house prices and evaluated models using residual plots