KRITHIKA RAJENDRAN

Entry-level Software/Machine Learning Engineer

682 306-2302 | rkrithika1993@gmail.com | linkedin.com/in/krithika-rajendran
github.com/krithika93

EXPERIENCE

Artificial Intelligence Engineer Intern

RadicalX

08/2023 - Present

- Conceptualized impactful solution for podcast summarization, and implemented data pipelines using AWS Lambda, S3, and tuned Open AI GPT-4 LLM.
- Engineered an impactful AutoGen App using Vertex AI Embeddings and Langchain to simulate personalized Pair Programming user experiences where I orchestrated Task Manager and Junior Bot agents using GPT-4.
- Developed a Q&A Bot using Retrieval Augmented Generation (RAG), employed Chroma Vector database to boost query handling speed and retrieval accuracy by 80%

Data Engineer Intern

Ballotpedia

09/2022 - 01/2023

- Engineered efficient Python scripts using Selenium, Beautiful Soup, XPath for data scraping, Pandas for data cleansing, transformation, analysis
- Spearheaded an OCR tool leveraging state-of-the-art ML model development techniques with LayoutLMV3 models, detecting data in tables of complex ballot sheet images
- Propelled the delivery of product ahead of time by thriving in a fast-paced team environment, delivering a crucial 67% execution efficiency improvement in Selenium data extraction/ data scraping scripting causing bottleneck in the pipeline operations
- Directed the detailed documentation of technical procedures and identified potential areas for optimization; Ensured Code quality, reliability, and scalability through best practices and code reviews.

Big Data System Engineer

Tata Consultancy Services

12/2015 - 06/2018

- Migrated large volumes of Retail data in terabytes, utilizing Spark-SQL and Scala delivering result earlier than expected in mission-critical solutions thereby enhanced the deployment speed of solution
- Enhanced real-time data streaming pipeline for Scala Spark, Spark-MLlib credit card approval processing through debugging Kafka data ingestion flow and reduced time to deployment.
- Collaborated with stakeholders to address data quality issues after migration of Legacy tables from Mainframe to Hive using intelligent recursive algorithm and received appreciation from Team Lead for reducing the existing cost.
- Created a Word Cloud Tableau DashBoard for finding the second most frequently tweeted disease among each group of demography using data collected from Twitter API.

EDUCATION

Master of Science MS in Computer Science

The University of Texas at Arlington

08/2023

SKILLS

Machine Learning/Artificial Intelligence

Pytorch **AutoGen** **AutoML**
Scikit-learn **MLFlow** **LangChain**
Predictive Labeling/ Supervised Classification

Data Engineering

Artificial Intelligence **SQL** **Python**
Data Visualization **Data Pipelines**
Google cloud platform **Kafka**
MongoDB **Hadoop**
Natural Language Processing
Selenium **Tableau** **Statistics**
Data Mining **PySpark** **Neo4j**

Software Engineering

Algorithms **Github** **Java**
Scala

ACHIEVEMENTS

TCS Gems Award

During a retail project, I was awarded for delivering results earlier than expected in collaborative team work

TRAINING / COURSES

Software Engineering: Analysis, Design, and Testing

Neural Networks

Machine Learning

Advanced Database Systems

Database Implementation Techniques

PROJECTS

Machine Learning (ML) /Deep Learning Predictive Modeling Auto-Tagging Notes

10/2022 - Present

- Implemented Python and Scala Spark code for cleaning, pre-processing large volumes of unstructured data into a structured format as training data,harnessed Sklearn, NLP algorithms for getting features like word order to verify data augmentation, **enhanced processing speed by implementing Scala Regex code in Spark by 50%.**
- Achieved an **impressive 97% AUC metric accuracy in predictive new entries** on Deep Learning Bi-LSTM and Bi-GRU models after data filtering, data augmentation.
- Design an Auto-tagging ML solution using categorical target labels/tags prediction in note entries using NLTK, implemented data pipelines for **XG-Boost model** training using dataset of about 5000 individual entries.

Big Data – Scala Spark Graph substructure discovery

02/2022 - 03/2022

- Implemented Graph substructure discovery on large graph datasets**, performed a non-trivial task of translating 1000 lines of code, from Java to Scala MapReduce and Scala Spark**, focused on substructure duplicate removal.
- Optimized application performance by implementing best practices resulting in a **50% decrease in processing speed of the application as shown in Google Cloud Platform by utilizing parquet files in Spark to streamline data storage.**

Database Implementation Techniques

07/2023 - 07/2023

- Conducted simulation of sequences of read, write, commit, abort transactions using the implementation of Rigorous two-phase locking with deadlock detection protocol in Python. **Achieved a 50% reduction in response time**
- Developed and executed Python scripts to perform graph multi-hop analysis on local store Data using Neo4j graph database with cypher queries for the retail dataset.
- Created a Word Cloud in Tableau to show major Airport hubs for 2 major airlines ranked by centrality measure in Python NetworkX
- Developed more 30% optimal Local alignment and Global alignment string algorithms (Needleman-Wunsch & Smith-Waterman) using Dynamic Programming principles and duplicate data representation using Dictionary.

PUBLICATIONS

Learning to Grade Short Answers using Machine Learning Techniques

Women in Computing and Informatics,ACM Digital Library International Conference Proceeding

Krithika Rajendran, Jayasree Narayanan

2015 http://dl.acm.org/citation.cfm?doid=2791405.2791508

I am attempting to grade short answer automatically which can be efficient and helpful to both students and teachers

AWARDS

Innovation Jockeys -Cognitive Computing Jury's Choice Award by Accenture

CERTIFICATION

R for Data Science

Uplimit, May 2023

Information Systems Security Professionals, NSTISSI No. 4011, System Administrators, CNSSI No. 4013E

The Committee on National Security Systems and The National Security Agency, Dec 2022