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

- Developed podcast summarizaton tool using ML techniques , and implemented data pipelines using AWS Lambda, S3, and optimized Open AI GPT-4 LLM.
- Participated in the adaptation of transformer models like Whisper for analyzing audio data, aiming to improve the accuracy of content recommendations.
- Developed AutoGen with Vertex AI Embeddings and Langchain-LLM technology to simulate personalized Pair Programming, orchestrating Task Manager and Junior Bot agents with GPT-4.
- Created a Q&A Bot using RAG, utilized Chroma Vector database for 80% faster query handling and retrieval accuracy.
- Contributed to the creation of a journaling tool that employs generative AI models to facilitate users in reflecting on their daily experiences using ranking and recommendation models , leading to an improvement in user engagement metrics.
- Develop an adaptive learning algorithm using iterative feature engineering, deep reinforcement learning and twin tower architecture, significantly enhancing content personalization and user engagement in educational platforms.
- Learned about utilizing Explainable AI (XAI) techniques to ensure transparency in AI decision-making processes, enhancing stakeholder and user trust in AI solutions.

Data Engineer Intern

Ballotpedia

09/2022 - 01/2023

- Engineered 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 , 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.
- **Managed technical documentation by simplifying complex concepts in an engaging way** and identified potential areas for optimization.
- Ensured Code quality, reliability, and scalability through best practices and code reviews.
- Applied Python unit tests in Python-Selenium data scraping scripts resulting in a 40% reduction in data validation and data integrity errors.
- Developed an automated OCR tool for Ballotpedia, utilizing AWS Textract, SageMaker, Lambda, and S3, along with API Gateway and DynamoDB/RDS for efficient ballot sheet processing and data management.

SKILLS

Machine Learning/Artificial Intelligence

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

Data Engineering

R **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

Database Management Systems

Machine Learning

Advanced Database Systems

Database Implementation Techniques

EXPERIENCE

Big Data System Engineer

Tata Consultancy Services

12/2015 - 12/2020

- Managed migration of terabytes of data of major marketing research firm, 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 by developing Kafka data ingestion flow and reduced time to deployment.
- Played in key role in ensuring seamless integration of Cloudera Data Platform infrastructure, Apache Kafka and HDFS by managing the libraries in Linux.
- Collaborated with stakeholders to address data quality issues after migration of tables from Legacy environment On-Prem Mainframe to Hive using custom "Divide and conquer" recursive algorithm.
- Developed Proof of Concept using SparkSQL to show Change Data Capture between source and target tables and optimized scalability of data processing workflows.
- 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.
- Executed a comprehensive Proof of Concept by applying KNN neighbors to cluster large-scale documents in the Software Engineering realm using NLP methodologies, optimizing Nearest Neighbor Selection parameters and features to attain a 0.67 accuracy metric in offline development.
- Developed and implemented K-Means and bisecting K-Means clustering algorithms in Keras for a dataset of 10,000 handwritten digits, achieving high V-measure; ranked among top 3 for accuracy in TCS Internal Competition
- Developed complex Excel models to optimize stochastic, linear, and integer functions using Frontline Solver Platform; achieved a reduction in operational costs.
- Implemented advanced procedures like natural language processing techniques like Locality Sensitive Hashing, using the optimized memory configurations in Spark.
- Used Hive to analyze the Partitioned and Bucketed data, develop UDFs, optimized and implemented Update queries in Hive using Spark SQL and Delta Lake.
- Developed and maintained custom MapReduce jobs for data processing, analysis, and transformation including developing a MapReduce job for processing Compressed files as input.
- Worked on ingestion, parsing, and loading the data from CSV, Parquet, Avro and JSON files using Spark.
- Ingested data to Hadoop using Sqoop and performed validations and consolidations for the imported data.

EDUCATION

Master of Science MS in Computer Science

The University of Texas at Arlington

08/2023

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

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