Prakul

+17165076412 | prakul.95.pc@gmail.com| linkedin.com/in/prakul-95

Technical Skills/Technologies

Languages: Python, Java, Go, JavaScript, Ruby, R programming, MySQL, SQL, Bash scripting, C, C++, C#, HTML

Tools, Services, and Techniques: AWS (CloudFormation, Lambda, ECS, EKS, EC2, S3, Fargate, Kubernetes, CDK, IAM roles), GitHub, webhooks, MySQL, Data mining, Object-oriented programming, Object-oriented design, Python Flask, Django, Docker, Cradle, Data structures, Restful Api, MATLAB, Tableau, MongoDB, Machine learning, PySpark, Hadoop, HDFS, Big Data, Data Analytics, Big Query, Pandas, GCP, Distributed computing. Hugging face, LLM, gen-AI, version controlling, software design, TypeScript, Prometheus, Angular, Node JS, Documentation, JSON, YAML, TensorFlow, Minikube, kubectl, Multithreading, Redis, HelmCharts, Pytorch, Load balancing, Distributed systems, Spark, unit testing.

Work Experience

Software Development Engineer, Amazon.com (Jun 2022 -Feb 2024) - Full Stack Development

As part of the Alexa Conversational AI Dynamic Routing team, I designed, implemented, and maintained critical Tier 1 components for the HypRank service, ensuring accurate matching of customer intent to the best provider and hosting machine learning models using AWS distributed systems.

- Designed, implemented, and tested Alexa self-learning, a java-based service to optimize request routing between new LLM and classic stacks, reducing excessive GPU compute usage. Developed ETL pipeline for model data ingestion using AWS lambda. Created automated monitoring and alerting systems to perform large scale data analytics using TypeScript CDK framework to enhance cost efficiency.

- Created orchestration tool using Kubernetes EKS and Golang API to deploy model configurations, reducing science team man-hours, minimizing deployment errors, and catching faulty configurations before deployment for A/B experiments, and further developed webhooks using Helm in Golang on Kubernetes (AWS EKS), reducing the failure rate for the orchestration tool's API calls by 30%.

- Rendered new fields, validations and maintained the Orchestration tool webpage based on Node.js which allowed customers throughout Alexa to submit configurations for deployment through Kubernetes controller-based API calls.

- Accomplished ~1 million defect reduction per week as measured by Customer Perceived Defect Rate by leading the integration of customer interaction embeddings signal into HypRank's workflow using Java, git and CI/CD pipelines, improving accuracy by 85 bps. Achieved a seamless tier 1 system migration with 0 downtime, maintaining high availability.

- Managed capacity planning for Alexa peak traffic and 24/7 on-call rotations, promptly resolving tickets and addressing service issues within a stringent 15-minute timeframe. Followed software engineering practices including coding standards, code reviews, source control management, build processes, testing, and operations.

Data Science Intern, EITACIES INC Software Services, (Sep 2021 - Jan 2022)

Project: Racial and Ethnic Slur Detector

- Engineered a voice transcriber using Mozilla DeepSpeech for streaming transcription.

- Implemented Emotion Detection from Voice Audio Files using transfer learning from the ResNet34 architecture.

Software Development Senior Associate, NTT DATA, (Aug 2017 - Mar 2019)

- Developed a web app using Python Flask and SQL, exposing APIs for generating business reports.
- Optimized client report runtime from 1hr to under 2 mins by creating efficient bash scripts.
- Monitored daily application health checks for the Asia region, version controlling and implemented client -side login encryption through hashing.

Projects

Map-reduce using HDFS

- Developed optimized map-reduce programs for co-occurrences, shortest path, and page rank for large-scale web-scraped datasets utilizing the Hadoop file system. Enhanced performance tuning for efficient data processing. Integrated machine learning models for data analysis, leveraging deep learning techniques. Employed high-performance computing practices to handle large datasets, ensuring efficient data processing and analysis.

Improving COMPAS to Minimize Racial Discrimination

- Compared 15 ML models (SVM, Neural Networks, Naïve Bayes) using Keras and TensorFlow, focusing on fairness metrics like Demographic Parity and Equal Opportunity. Optimized models for performance and minimized racial discrimination, producing a report based on causal inference for model selection.

Education

Master of Science, Data Science

University at Buffalo, The State University of New York (Aug 2020 - Feb 2022), GPA: 3.778/4.0

Bachelor of Engineering, Electronics & Communication

Guru Gobind Singh Indraprastha University, New Delhi, India (Aug 2013 - Jul 2017)