Divit Vasu

Bay Area, California | (+1) 925-418-9981 | divitvasu@gmail.com | LinkedIn | GitHub | Website | Open to Relocation

Summary

Driven Computer Science graduate with research experience in Machine Learning. Experienced in backend, infrastructure, and data engineering in a fast-paced startup environment, demonstrating strong skills in technical problem-solving, effective communication, and team collaboration. Seeking opportunities to apply expertise in Backend Engineering, Full-stack Development, Data Engineering, and Machine Learning to drive innovation and efficiency. Committed to continuous learning and keeping up with industry trends

Education

Master of Science in Computer Science — Jan 2021 – May 2023
Northeastern University, Khoury College of Computer Sciences, USA — **GPA:** 3.91/4.0

Bachelor of Technology in Computer Science Engineering — Jul 2016 – May 2020
Ganpat University, UVPCE, India, Minor in Big Data — **GPA:** 8.72/10.0

Relevant Courses: Algorithms, Programming Design, Data Science and Modelling, Machine Learning, Cognitive Computing, Compiler Design, Artificial Intelligence, Big Data Analytics, Cloud Computing, Data Mining, Natural Language Processing, Scalable Distributed Systems, Database Systems, Mobile Development, Operating Systems, Computer Networks, Software Engineering

Skills

Languages: Python, Java, Ruby, R, C++, C, JavaScript
Operating systems: Linux, Windows Server, Raspbian
Web Technologies: HTML, CSS, XML, XPATH
API Architectures: REST, RPC, Message Queuing, GraphQL
Testing Frameworks: Pytest, Rspec, JUnit, JMeter
Virtualization: Hyper-V, KVM-QEMU
Data Analysis: Tableau, WEKA, PowerBI
ML Libraries: Keras, Tensorflow, NLTK, Sklearn, Pandas, NumPy, Seaborn
DevOps: Git, Docker, Jenkins, Kubernetes, Nagios, Datadog, Rollbar
Frameworks: Rails, Django, Flask, Spring, Hadoop, React, RxJava, Retrofit
Tools: AWS, RabbitMQ, Sidekiq, Kafka, Databricks, Airflow
System Administration: Cacti, Spiceworks, Microsoft SCCM
Databases: MySQL, Postgres, SQLite, Firebase, MongoDB, ELK stack, Redis, Hive, HBase

Certifications

NPTEL: Data Mining - Indian Institute of Technology
AWS: Solutions Architect Associate (SAA-C03)
Red Hat: RHCSA (EX200), RHEV (EX318)
IBM: Big Data Mastery Award, AI Analyst Mastery Award

Work Experience

Software Engineer | **Miles,** ***Bay Area, USA*** — Jul 2023 – Jun 2024

- Architected and delivered critical product features for the Miles rewards app, ensuring a seamless **user experience for 4 million users** through scalable and cost-efficient database designs. Also performed on-call duties for AWS-hosted app infrastructure
- Developed and implemented a backend API for device version verification, prompting clients to forcefully upgrade if their version was below a specified threshold or optionally upgrade within a given range, strategically designed to deprecate older API versions and optimize costs while ensuring optimal code efficiency for widespread use across active user client devices. This resulted in 38% DAU upgrading, bringing about annual projected **cost-savings of $11,500**
- Spearheaded the App-Open Permissions feature, designing and deploying a system that displayed a notification prompt on the app every x days to users who had not granted notification permissions. Coded a client API to determine when to trigger the pop-up, owning the entire process from efficient table design, query planning, indexing, integration testing to post-rollout monitoring
- Implemented a feature to award bonus miles to users for completing activities during happy hours, utilizing a specified multiplication factor. Structured and implemented a push notification pipeline to efficiently **notify 2 million users within 30 minutes** by creating necessary queues and configuring worker nodes to drain the queues in a timed manner
- Independently managed the full scope of data engineering tasks, developing large-scale ETL pipelines in Apache Spark for B2B **data contracts valued at $1 million annually**, leveraging Airflow, Databricks, AWS S3, and Redshift to efficiently process over 200 million rows of geospatial data daily while maintaining effective partner communications
- Developed and managed ETL pipelines using Airflow to automate company-wide KPIs, ensuring metrics were accessible to the leadership and growth, marketing teams. Collaborated with various stakeholders and managed their expectations
- Single-handedly set up and scaled infrastructure to expand operations internationally, spinning up and seeding a network of 25 Postgres and 3 Redshift databases, 50 SQS queues, and 7 load balancers, along with VPC/Route53, SSM configs, S3 buckets, and Firehose data streams, with partial automation through Terraform. Set up Airflow, Tableau, and Databricks for data jobs
- Actively involved in weekly growth tasks, gathering and presenting company performance metrics during monthly business reviews, and collaborating with leadership to generate insights and develop POCs based on customer data

Graduate Research Assistant | **Khoury College of Computer Sciences,** ***Bay Area, USA*** — Feb 2023 – Aug 2023

- Researched on achieving accurate segmentations of MRI images, by leveraging CNN architectures such as 2D U-Net, 3D U-Net, and Computer Vision techniques, towards building personalized musculoskeletal, neuromechanical simulation models
- Assessed a range of annotation tools such as LabelMe, LabelStudio, SimpleITK, and DeepLabCut to streamline the preparation of training data. Additionally, collaborated on integrating the trained model files with a custom visualization utility

Graduate Teaching Assistant | **Khoury College of Computer Sciences,** ***Bay Area, USA*** — Jan 2023 – Apr 2023

- Managed TA duties for Cloud Computing, responsibilities: doubt-solving, mentoring, and supporting students academically

Software Engineer Intern | **Binary Works IO,** ***Bay Area, USA*** — Sep 2022 – Dec 2022

- Collaborated as a mobile engineer on a native Android app for an e-publication client on backend and frontend aspects, that allowed in-app purchases supported by WooCommerce, downloading, reading, and deleting purchased e-pub issues, editing user profile, alongside a user library with features such as filtering, bookmarking, highlighting, day/night mode

- Utilized Firebase Storage to securely store user-owned e-pub files, implemented encryption for enhanced security, and leveraged Retrofit and RxJava to enable asynchronous downloading of e-pub files
- Referenced UI design, themes based on Material Design guidelines for creating XML layout files of various app screens
- Stored application and user-profile settings in CloudFirestore database, allowing efficient management and retrieval of data
- Supported user authentication, authorization through Firebase, ensuring secure access to the app's features and functionalities

Software Engineer Intern | **Miles,** ***Bay Area, USA*** Jan 2022 – Apr 2022

- Built internal tools using Rails and Django (backend) and React toolbox components (frontend), streamlining admin tasks for marketing and customer support teams. This reduced daily efforts by 2-3 hours and decreased support tickets by 40% a day
- Created a data ingestion API that received JSON payloads from a third party at unpredictable intervals. Configured AWS Lambda as the entry point to receive the payloads and send them to an SQS queue. Implemented a secondary Lambda function triggered by the SQS queue, responsible for validating the payload's integrity, writing it into files categorized by day, in an S3 bucket. Additionally, integrated Datadog logging to track and monitor metrics whenever the entry-point Lambda was triggered. In case of corrupt or invalid payloads, a Dead Letter Queue (DLQ) was used to handle them appropriately
- Owned and implemented the app's welcome rewards feature end-to-end. Collaborated with the growth-marketing team to create a tool for managing reward details. Worked with the mobile team to establish API architecture and integrate a new API for reward claims at user sign-up. Additionally, coded an API to identify new users and deliver personalized user experiences

Graduate Teaching Assistant | **Khoury College of Computer Sciences,** ***Bay Area, USA*** May 2021 – Dec 2021

- Served as a TA for the graduate-level subject of Algorithms, key responsibilities included, grading homework assignments and exams, conducting weekly doubt-solving sessions, mentoring, and lending academic, moral support to the students

Software Engineer | **Capri Technosys,** ***Ahmedabad, India*** Apr 2020 – Jan 2021

- Designed and implemented a scalable document digitization project for Govt. of Gujarat using AWS, that enhanced accessibility, improved efficiency, and reduced the risk of loss or damage to important records and archives
- Contributed to developing a front-end web console on React and designed RESTful APIs to facilitate backend integrations
- Architected a pipeline with AWS S3, SQS, Rekognition, and Lambda for OCR processing, storing metadata in DynamoDB. Utilized AWS Comprehend to extract key phrases and inserted document data into Elasticsearch, enhancing data retrieval efficiency
- Tested successful handling of 400-500 documents during test runs and configured the system to send failed or interrupted documents to separate S3 buckets for manual review

Machine Learning Research Engineer Intern | **Dev Information Technology,** ***Ahmedabad, India*** Dec 2019 – Apr 2020

- Developed a chatbot from scratch using RASA stack for the financial domain, gained proficiency in working with TTS libraries, fine-tuning learning hyper-parameters, PowerBI, and researched upon building context-specific NLG systems
- Took ownership of creating effective content for the NLU file and focused on improving the training of RASA NLU and RASA Core. Employed various techniques to train the model, including the creation of diverse user stories to cover both successful and challenging conversational paths
- Engaged with the accounting team throughout the development process, seeking feedback and refining the chatbot's performance, ensuring that the chatbot delivered accurate, meaningful responses in the context of accounting conversations

Systems Engineer Intern | **Dev Information Technology,** ***Ahmedabad, India*** Mar 2018 – May 2018

- Coordinated procurement, assembly, installation of entire equipment for a custom NAS solution for a client, with RAID-5 arrays and monitoring on Nagios, coded custom shell scripts for monitoring, installed Webmin, and scheduled cron jobs
- Gained hands-on experience in deploying and managing VMs using Hyper-V and performed basic infrastructural monitoring and upkeep of server machines using Nagios, Spiceworks, and Microsoft SCCM to proactively identify and resolve issues
- Assisted in assembling, installing and configuring server machines, network patch panels, and other hardware onto server racks, ensuring seamless integration within the network infrastructure

Projects

Task Tracker App - *Java, Android SDKs, Firebase Authentication, Firebase Realtime Database*
Created an Android task tracker application using Java and Android SDKs, integrating Firebase Authentication for user registration and login, and utilizing Firebase Realtime Database to store and manage task data. Implemented features for creating, updating, and deleting tasks, along with user authentication and data synchronization across devices, providing an efficient and **synchronized task management and tracking** solution for roommates

Distributed Key-Value Store showcasing Consensus Algorithms - *JAVA, RPC (JAVA RMI), multi-threading, PAXOS*
Devised a concurrent, fault-tolerant, distributed client-server environment to achieve data consistency among server replicas by leveraging PAXOS and majority voting, successfully handling at least **10 server replicas and 3 client nodes**

Healthcare Assistant - *RASA, Python, Twilio, WhatsApp, Ngrok, Infermedica and Google Maps API, SQLite*
Engineered an NLP chatbot in RASA that gathered a set of user symptoms and leveraged the Infermedica primary care API to provide diagnoses and lists of relevant doctors sorted by ratings, nearby hospitals, and pharmacies. The bot operated with a **confidence level of 92-95%** in identifying user intents and performing entity-recognition

Prediction of Alzheimer's/Dementia - *Datasets from Kaggle and OASIS Research community, Colab, Python (nilearn, SimpleITK)*
Predicted patients at higher risk of dementia by applying Decision Tree modeling, XGBoost, Logistic regression, and SVM on datasets, relating the findings to actual MRI inferences, **exploring neuroimaging, and plotting libraries**

Publications

MRI Segmentation of Musculoskeletal Components Using U-Net: Preliminary Results, Divit Vasu, Seungmoon Song, Hans Kainz, Jeongkyu Lee, ICBBB '24, https://doi.org/10.1145/3640900.3640902