

Contact

Briar Forest Dr, Houston, Texas
8606204718 (Mobile)
sailikhithcse@gmail.com

www.linkedin.com/in/sailikhithk
(LinkedIn)
sailikhith.me/ (Portfolio)
github.com/sailikhithk (Blog)

Top Skills

Apache Airflow
Technical Project Leadership
Spring Boot

Certifications

Microsoft Certified: Azure Data
Scientist Associate
Google Cloud Certified Professional
Data Engineer
Machine Learning Engineer
Nanodegree
Oracle Certified Associate, Java SE
8 Programmer

Honors-Awards

Best Evaluator, Best Speaker, Best
Role Player
Best Chess Player
Best Poster Presentation
Best Academic Project
Third prize in National Science
Talent Test

Sai Likhith K

Sr. Python Developer | Data Scientist | Data Engineer | MLOps | Java
| AWS 2X Certified | Azure 1X Certified | GCP 1X Certified
Houston, Texas, United States

Summary

Full stack Developer with experience at creating products to deliver insights and implement action-oriented solutions to complex business problems. Strong quantitative background in engineering, business application development, and Agile/Scrum methods. Project and work experience designing interactive web applications, leading development teams building custom enterprise applications, and using modern cloud computing technologies like AWS, and Google Cloud.

►► Looking for contract Backend Engineer, Data Engineer, Full Stack Software Engineer, Data Science, or similar position.

Can offer full-stack engineering skills using Agile methods for building scalable web and mobile applications using front end and backend development technologies.

Experience designing applications using Python, Java, CI-CD pipelines, AWS, Flask

HIGHLIGHTS

- ✓ Upgraded open-source chat messenger in 15 days. Earned 80% positive feedback from forum users.
- ✓ Improved deliverables productivity by 20%. Completed deliverables 2 months faster than expected.
- ✓ Automated processes for 13 business units, which saved 600+ monthly work hours.
- ✓ Developed a real-time analytics dashboard, which boosted quarterly client business activity by 20%.

TECHNICAL SKILLS

★ Python, Java, CI-CD

- ★ Distributed, scalable, backend systems , kafka
- ★ HTML, CSS, JavaScript, Typescript
- ★ Oracle SQL, MySQL, SQL Server, MongoDB, DynamoDB.
- ★ Web frameworks: React.js,Vite.
- ★ Cloud computing using AWS, Google Cloud, Firebase, Oracle Cloud Infrastructure
- ★ Oracle Certified Associate, Java SE 8 Programmer.
- ★ Machine Learning: Linear Regression, Logistic Regression, Decision trees, Random Forest, Association Rule Mining (Market Basket Analysis), Clustering (K-Means, Hierarchal), Gradient descent, SVM (Support Vector Machines), using TensorFlow (Keras).

I'm happy to receive new connections for opportunities in software engineering. Please connect with me on LinkedIn, or by email:

✉ ksailikhith007@gmail.com

📞 281 584 3017

🖱 <https://www.github.com/sailikhithk>

Experience

Airbnb

Sr. Data Scientist/ Sr. Python Developer - AI Development Experience

September 2024 - Present (1 year 9 months)

San Francisco Bay Area

Building ML infrastructure and GenAI tooling for Airbnb's AI data prep and labeling ecosystem, spanning LLM orchestration, PII protection, Labelbox workflows, Streamlit apps, Flask/Celery services, and Airflow-based data pipelines.

- Built and shipped internal GenAI workflow tools used by ~55 analysts, integrating 30+ LLM models behind a unified interface for structured analysis, summarization, and insights.

- Scaled analysis workflows from 600 to 10,000 rows per run and increased supported upload size from 10 MB to 40 MB through batching, chunked file processing, and runtime tuning.

- Built a PII de-identification pipeline using Microsoft Presidio with targeted entity detection, improving processing speed by ~30%.
- Contributed to backend and frontend data-labeling platforms supporting Labelbox uploads, model-assisted labeling predictions, benchmarks, and label exports.
- Improved reliability across long-running LLM workflows with token-lifetime checks, retry handling, observability, and traceable logging.
- Worked across Python, Streamlit, Flask, Celery, Redis, SQLAlchemy, Airflow, Hive/Trino, OpenTelemetry, Datadog, Docker, and AWS-backed infrastructure.

Eli Lilly and Company

Sr. Data Scientist

February 2024 - August 2024 (7 months)

Philadelphia, PA

- Architected and implemented a comprehensive dose order management system for radiopharmaceutical drugs using KUBED (Kubernetes-Based Enterprise Development) platform, integrating with AWS services and microservices architecture.
- Developed core microservices for dose ordering, manufacturing scheduling, and capacity management using Spring Boot and React, handling complex radioactive material decay calculations and time-sensitive delivery scheduling.
- Implemented secure authentication and authorization using Entra ID and Istio service mesh for service-to-service communication, ensuring compliance with pharmaceutical industry regulations.
- Created sophisticated algorithms for dose calculations and delivery scheduling, incorporating radioactive decay rates and transportation time constraints using Spring Boot backend services.
- Architected observability solution using Grafana, Loki, and CloudWatch, providing comprehensive monitoring and alerting for the entire system including manufacturing schedules and dose delivery tracking.
- Built React-based frontend applications with TypeScript for dose ordering and management, implementing complex form validations and real-time scheduling interfaces.
- Developed GitOps deployment pipeline using ArgoCD and GitHub Actions, ensuring consistent deployments across development, testing, and production environments.
- Regulatory compliance and audit systems for pharmaceutical manufacturing
- Real-time workflow orchestration using Temporal
- Infrastructure as code using Crossplane and Kubernetes

- Implemented audit logging and compliance tracking system for RAM licenses and regulatory requirements, ensuring adherence to pharmaceutical industry standards.
- Achieved 99.9% system availability through robust error handling, failover mechanisms, and comprehensive monitoring using AWS CloudWatch and custom alerting systems.

Southwest Airlines

Data Scientist / Senior Application Engineer

January 2023 - August 2024 (1 year 8 months)

Dallas, TX

- Architected algorithms to select the right activity & improving deployment and testing automation pipeline using GitLab.
- Created a local s3 stack with Docker and deployed it in multiple pods using Kubernetes to replicate DEV environments.
- Developed a microservices architecture using RabbitMQ as a message broker to manage communication between different components. Implemented publish-subscribe and point-to-point messaging patterns for real-time data processing.
- Designed and implemented a Kafka-based data pipeline for real-time stream processing, processing daily events. Utilized Kafka Connect and Kafka Streams APIs to develop data integration and transformation workflows.
- Implemented data governance policies to ensure data security and compliance.
- Developed data pipeline involving flight tracking files receiving the Elastic Flat Files (EFFs) from RabbitMQ, sending the file to s3 and then sending it to Kafka topic producer for a data transformation.
- Deliver oral or written presentations of mathematical modelling and data analysis results to management or other end users.
- Intense experience in Python backend API development experience using frameworks such as Flask or Django.
- Gradle builds with Java code conversion from Python and unit test cases with 95% test coverage.

Shell

1 year 11 months

Python Developer/ Data Scientist

June 2021 - December 2022 (1 year 7 months)

Houston, Texas, United States

- Automated the historical data extraction from various sources like Databricks, and Microsoft Azure Storage Explorer for Pi Data from 1995-2022 extracts using ETL transformation data pipeline-based operations.
- Worked on text-based categorization to identify various types of drilling losses that occurred for oil wellbores located in the Gulf of Mexico using the AI and NLP algorithms.
- Designed and developed scalable applications using python, agile methodologies, NLP, and data engineering and deployed them in the AWS cloud.
- Built a python-based application for an autoencoder algorithm using CNN (Conventional Neural Networks) to generate lossless Acoustic Impedance graphs for internal business users.
- Collaborated and helped onboard interns while working on a POC (Proof of Concepts) for a research project regarding building a Responsible-AI framework in python based on LIME and SHAP modules for identifying bias in AI models and data.
- Built well-structured and documented RESTful APIs so our apps can easily share data.
- Delivered oral presentations and written technical papers to business stakeholders and leadership to drive insights and facilitate business decisions.
- Created Pipelines in ADF using Linked Services/ Datasets/Pipeline/ to ETL from different sources like Azure SQL, blob storage, Azure SQL Data Warehouse, write back tool and backwards.
- Developed Spark applications using PySpark and Spark-SQL for data extraction, transformation and aggregation from multiple file formats for analyzing and transforming the data to uncover insights into the customer usage patterns in the oil keyword API tool.
 - Prediction of wellbore-based drilling event losses by collaborating with clients to label the weighted events.
 - Key Achievement: Complex wildcard regex insertion into the NLP data pipeline, thus improving the scalability of the algorithm.

Python Developer

February 2021 - May 2021 (4 months)

Houston, Texas, United States

- Developed custom supervised NLP based text-classification algorithms for the existing unsupervised prediction of well digging event losses by collaboration with the clients for labelling the weighted events. Deployed the model into AWS data pipelines.

- Key Achievement: Optimized the AI Risk modelling pipeline, thus improved performance increase by 25% and F1-score improved by 0.2. Training time reduced by an hour.

Oracle

Python Developer/ Data Scientist

August 2017 - July 2019 (2 years)

Bangalore, India

- ◆ Performed several roles in app development lifecycle serving enterprise clients in tech and fintech. Tech stack included Java, Python, HTML, CSS, SQL, NoSQL, and Oracle Cloud applications.
- ◆ Interacted with clients' executives throughout entire project lifecycle from requirements gathering to maintenance.
- ◆ Used Agile processes during implementation phase. Improved deliverables productivity by 20%. Completed deliverables 2 months faster.
- ◆ Developed test scenarios and test cases, which achieved 95% success rate.
- ◆ Built, enhanced, debugged, and troubleshot scalable enterprise software programs with 50,000+ lines of code.
- ◆ Automated Performance Improvement Plan process for 13 business units. Saved 600+ monthly work hours.
- ◆ Developed ERP analytics dashboard delivering real-time overview of client CEO's 13 business units, and see quarterly predictions using analytics and machine learning. Boosted quarterly client business activity by 20%.
- ◆ Integrated external modules into Oracle Cloud HCM. Developed custom reports with 2X speed.
- ◆ Developed test scenarios and test cases, which achieved 95% success rate.
- ◆ Performed an in-depth study about Elastic Search, Kibana, Logstash, Kafka, and Spark to establish two data pipelines with several small demos. The first data pipeline is ELK stack for processing and visualizing Apache Spark logs in a scalable way. Built the second AWS Kinesis Data Stream based credit card fraud detection data pipeline to process credit card transactions in real time using Kafka message broker to classify if fraudulent or authentic (based on user's buying pattern).

JetBlue

Python Developer

June 2015 - July 2017 (2 years 2 months)

Karnataka, India

- Improved data mining processes, resulting in a 20% decrease in time needed to infer insights from customer data used to develop marketing strategies.

- Automation testing using JMeter and load runner frameworks for unit and integration testing as part of the Test Driven Development (TDD), Load testing done in JIRA and Jenkins as well.
- Used predictive analytics such as machine learning and data mining techniques to forecast company sales of new products with a 95% accuracy rate.
- Improved data mining processes, resulting in a 20% decrease in time needed to infer insights from customer data used to develop marketing strategies.
- Implemented AWS Solutions using EC2, S3, RDS and EBS for scalability of the systems on top of the Application Load Balancer.
- Increased data security by updating companywide encryption, steganography, IP security, and secure wireless transmission practices.
- Experience in Text Analytics, developing different Statistical Machine Learning, Data Mining solutions to various business problems and generating data visualizations using R, Python.
- Developed predictive model of the fuel consumption using Tree based regression analysis using Decision Tree, XGBoost and Random Forest for a comparative analysis on predictive models in AWS SageMaker and Jupyter Notebooks.
- Performed Time series forecasting of the price prediction of the plane using Auto ARIMA models for next Christmas season of the year and new year eve.
- Developed an Flask API on top of the models mentioned above and deployed it into AWS using AWS Code Pipelines in integration with GitHub actions for CICD pipelines for automated code deployments.
- Enhanced existing system architecture for monitoring the optical fiber communication losses by partnering with hardware engineers. Slashed company's monthly maintenance budget by 2%.
- Developed applications and supported communication team analyzing and optimizing losses during optical fiber communication.

Education

New York University

Master's degree, Computer Science · (August 2019 - May 2021)

Jawaharlal Nehru Technological University, Anantapur

Bachelor of Technology - BTech, Electronics and Communications Engineering · (August 2012 - May 2016)

Narayana Junior College

Intermediate, M.P.C · (2009 - 2011)

Sai Saraswati English Medium High School

8th to 10th Class, Educational, Instructional, and Curriculum
Supervision · (2006 - 2008)