

SOVIN BIRLA

sovinbirla.com; <https://github.com/sovinbirla>; Cell: 509-288-0988.

SUMMARY

Seasoned Backend Engineer with 4+ years of experience building production-grade services for high-scale SaaS and healthcare platforms. Skilled in designing and scaling APIs, modeling multi task systems, and integrating complex data stores like PostgreSQL, MongoDB, Redis, and Elasticsearch. Expertise in GCP and AWS infrastructure, CI/CD, Kubernetes, and PyTorch. Proven ability to lead distributed teams and deliver high-quality backend architecture and integrations.

EDUCATION

Masters in Applied Data Science, Remote

University of Chicago

Mar 2025 - Mar 2026

Visiting Graduate Student

Stanford University

Jan 2024 - July 2024

BS Statistics + Computer Science

University of Illinois Urbana-Champaign

Aug 2017 - May 2021

EXPERIENCE

Machine Learning Intern, Elevance Health, Atlanta, GA

May 2024 - Jul 2024

- Developed a real-time medical chatbot for cardiologists using AWS, Docker, Ollama, FastAPI, and React. Enabled text-to-3D lung reconstructions by integrating vision transformers and diffusion models.
- Leveraged Ray Train to parallelize training of a diffusion-based vision transformer for text-to-3D lung reconstruction.
- Expanded the training dataset by 21% using fine-tuned LLMs (Phi-3, Mistral 7B) and BiomedCLIP to improve automated CT scan analysis.
- Built a stroke prediction model with 81% accuracy using feature engineering and Random Forest/AdaBoost classifiers for early risk assessment.

Software Engineer, Maven Wave, Chicago, IL

Jun 2021- Dec 2023

- Architected a cloud agnostic AI platform to enable clients to fine tune their end-to-end platforms resulting in an improvement of 22% in client churn
- Deployed 5+ cloud-native (GCP, AWS), full-stack, solutions using Django, React, Next.js, Python, Java, C#, PostgreSQL, MySQL, ArcGIS, RabbitMQ, Firebase, and Kubernetes
- Led an 8-member team across 3 continents, employing Agile and DevOps methodologies, with ticket management on Trello/JIRA/ADO
- Worked alongside leaders from EY, McDonald's, Capital One, Disney, In-N-Out, Match Group, and United Health Group to drive customer engagement by accessing 200 million active users

KPMG Data Science Intern, NYC

Jun 2020 - Aug 2020

- Built a Python application to collect MLS data from Zillow Bridge API to perform rigorous statistical tests in any given zip code in the United States
- Used classification and regression techniques to determine areas with high appreciation rates amongst other various factors using Keras, Matplotlib, and Pytorch
- Built CRUD application in React to handle POST requests to Zillow Bridge API to filter, append, and delete MLS listings from MongoDB database

PROJECTS

Distributed Raspberry Pi SLURM Cluster

- Set up a mini HPC cluster with 4 Raspberry Pi nodes to run and schedule compute-intensive batch jobs.
- Implemented multithreaded task execution in C++ and configured shared storage across nodes

Quantitative Developer

- Built trading systems with Interactive Brokers API that can effectively survey, test, and beat SPY with 35% returns in intraday trading
- Designed a neural network to manage my portfolio using Deep Q learning

Simple Shots

- Developed a few-shot machine learning model for analyzing tennis performance, focusing on tasks such as ball tracking, player tracking, and court line detection
- Integrated pre-trained YOLO models for object detection through Roboflow, used OpenCV to analyze frames, and used Pytorch to fine-tune the YOLO models

SKILLS

- **Languages:** Python, C++, Java, C, R, JavaScript, TypeScript, PyTorch
- **Databases:** MySQL, PostgreSQL, MongoDB, Oracle SQL
- **Cloud & DevOps:** AWS, GCP, Azure, Docker, Kubernetes, Firebase, Helm, SLURM