

JAYESH SINGH

(213) 551-6993 | jayesh.singh@gmail.com | linkedin.com/in/jayeshsingh27 | Open to relocate

EDUCATION

University of Southern California

Master of Science in Applied Data Science, GPA: 3.83/4.00

Los Angeles, CA

Aug 2022 - May 2024

- Coursework: Machine Learning, Data Mining, Data Science at Scale, Fairness in AI, Database Systems, Data Management

NMIMS University

Bachelor of Technology in Data Science, GPA: 3.73/4.00

Mumbai, India

Jul 2018 - May 2022

- Coursework: Deep Learning, Statistical Models, Big Data, Computer Vision, Cloud Computing, Natural Language Processing, Business Analytics, Supervised & Unsupervised Learning, Programming for Problem Solving, Statistics, Mathematics

EXPERIENCE

Research Scientist

Biomedical Imaging Lab, USC

Aug 2024 - Present

Los Angeles, CA

- Automated data ingestion and pipelined macaque brain surface extraction using Apache Airflow, cutting weekly manual work by 8 hours and enabling automatic model retraining on 20GB+ of new MRI scans.
- Co-developed a U-Net segmentation pipeline in PyTorch with radiologists for macaque MRI brain scans, achieving 88% Dice accuracy and cutting processing time by 40%, enabling faster analysis for 500+ global researchers.
- Containerized inference using Docker/Kubernetes and orchestrated blue-green CI/CD via GitLab, shrinking model deployment period from 2 days to 4 hours and integrating real-time quality control checks, reducing quality assurance overhead considerably.

Data Scientist Intern

Veersa Technologies

Feb 2022 - Jun 2022

Noida, India

- Built a scalable time series forecasting pipeline (Python, YAML, Git), validated across 20+ datasets for robustness.
- Collaborated with a cross-functional team on data preprocessing, modeling, and analytics modules for a production ML system.
- Enhanced ML infrastructure in partnership with stakeholders, improving system efficiency by 12%, integrating flexible support for models including SVM, LSTM, Prophet, ARIMA, and RNN.

TECHNICAL SKILLS

Programming Languages: Python, SQL (PL-SQL, NoSQL), JavaScript, TypeScript, HTML, CSS, Java

Frameworks/Libraries: Node.js, Django, PyTorch, Flask, React, TensorFlow, Keras, scikit-learn, OpenCV, HG Transformers, Streamlit

Tools/DevOps: Git, Docker, Kubernetes, Apache Airflow, Spark, Firebase, Tableau, Power BI

Cloud/Infrastructures: Amazon Web Services (AWS), Azure, Kafka, HDFS, PostgreSQL, MongoDB

ACADEMIC PROJECTS

eTA - Virtual Teaching Assistant

- Innovated a teaching assistant AI for e-learning platforms, based on student feedback, delivering personalized academic support.
- Integrated technologies like CLIP, FAISS, and OpenAI LLM, deployed via Streamlit for an intuitive frontend user experience.
- Designed a semantic query system leveraging embeddings, cosine similarity and vector clustering, enabling course navigation with timestamped images, enhancing student retention while minimizing instructor intervention.

Call Prioritization based on Emotion Recognition

- Engineered a multi-modal emotion recognition platform, integrating BERT for text (78% accuracy) and Random Forrest for speech (89% accuracy), optimized with dimensionality reduction and hyperparameter tuning for real-time inference.
- Leveraged the emotion recognition output to develop a smart routing algorithm that assigns waiting callers to agents based on detected emotional states, deploying models with Node.js for asynchronous, scalable processing.

Advanced Yelp Recommendation System

- Established a hybrid recommendation engine to process over 2 million rows with more than 30 features on HDFS.
- Constructed item-based collaborative filtering with model-based ensemble techniques, employing PySpark and Scala on Ubuntu, attaining a 0.97 RMSE via hyperparameter tuning on a data warehouse containing 500GB+ of noisy and heterogeneous formats.

Media Preprocessor Library

- Led research and creation of an open-source Python library for high-throughput image preprocessing and manipulation.
- Architected a modular and extensible codebase compatible with various formats for image standardization, edge detection, and background removal, streamlining dataset preparation for CV pipelines, recognized with over 15,000 downloads on web.

Road Classification and Subsequent Pothole Detection

- Designed a full-stack system integrating a CNN-based road classifier (96% accuracy) and YOLO-based pothole detector (87% accuracy), deployed via a Django backend to serve model inferences and coordinate system services.
- Built a Flutter mobile app for real-time road condition visualization using Google Maps API, with live updates from the backend enabling drivers to avoid potholes dynamically.

AI-powered Security Assistant

- Built an AI assistant using RAG (LangChain, OpenAI, Qdrant) to retrieve and summarize threat intelligence data and internal logs.
- Integrated DeepLog for anomaly detection, flagging ~30% more suspicious activity and reducing incident triage time by 40%; deployed with FastAPI, Kafka, and Kubernetes for scalable, real-time monitoring.