Ruthvik Bacha

Data Scientist | AI & ML Engineer | Cloud & Big Data Enthusiast

Tampa, FL

-Email me on Indeed: http://www.indeed.com/r/Ruthvik-Bacha/b74b63520e85f897

Professional Summary

I am a Data Scientist and AI Engineer pursuing a Master's in Business Analytics and Information Systems at the University of South Florida. My expertise lies in machine learning, big data processing, and AI model optimization. At Moffitt Cancer Center, I worked on ML-driven survival models, improving oncology prognosis accuracy by 25%, and integrated clinical and pathology data for enhanced deep learning applications. Previously, at Iamneo.ai, I built real-time data pipelines using Apache Spark and Airflow, optimizing SQL analytics and query performance.

My academic projects focus on AI-powered automation and predictive analytics, including a multi-language invoice extractor using Google Gemini Pro Vision & OpenCV, an AI chatbot for medical queries with Retrieval-Augmented Generation (RAG), and a traffic crash severity predictor achieving 84% accuracy. Proficient in Python, TensorFlow, AWS, PostgreSQL, and LLMs, I am passionate about leveraging AI and data science to build scalable, impactful solutions in healthcare, automation, and predictive modeling. ✎

Willing to relocate: Anywhere

Work Experience

Data Science Intern

Moffitt Cancer Center-Tampa, FL

May 2024 to October 2024

• Analyzed clinical and pathology data from 1,000+ melanoma patients, optimizing ML-driven survival models.

• Standardized whole-slide images (WSIs) using VALIS, enhancing deep learning model performance.

• Merged clinical metadata with image data, enhancing dataset integrity for ML-based survival analysis.

• Engineered CNN-based feature extraction, boosting survival prediction accuracy by 25% in oncology prognosis models.

Data Scientist

Meraki Solutions-Coimbatore, Tamil Nadu

September 2022 to April 2023

• Engineered data ingestion pipelines using Spark Streaming to process real-time student interaction data.

• Architected SQL-based data warehouses (PostgreSQL, Snowflake), enhancing real-time user behavior analytics by 20%.

• Optimized AWS S3 data lake using Parquet/ORC, cutting storage costs improving query performance by 30%.

• Developed batch processing workflows in Apache Airflow, improving data pipeline efficiency.

Education

Master's degree in Computer System Analysis
University of South Florida-Main Campus - Tampa, FL
August 2023 to May 2025

BE in Electronics and Communication
RV COLLEGE OF ENGINEERING - Bengaluru, Karnataka
August 2017 to August 2022

Skills

- Python
- Java (Less than 1 year)
- ReactJs
- HTML/CSS
- Node.js
- Tableau (1 year)
- Power BI (1 year)
- Git
- MySQL
- GitHub
- JavaScript
- SQL
- AWS
- Communication skills

Languages

- English - Native
- Kannada - Intermediate
- Telugu - Fluent
- Hindi - Fluent

Certifications and Licenses

Generative AI with LLM's
February 2025 to February 2027

Publications

Website

https://github.com/BachaRuthvik/StyleProject

October 2022

Created and implemented components to ensure scalability, maintainability, and performance also used Routing, Craousal and SCSS.

DETECTION OF SIGN LANGUAGE USING MACHINE LEARNING

https://www.irjmets.com/uploadedfiles/paper/issue_7_july_2022/28456/final/fin_irjmets1658330455.pdf

July 2022

• Conducted data selection, extraction, cleaning, and exploratory data analysis.

• Created a system that detects sign language gestures with 92.5% accuracy.

• Enhanced this solution using Raspberry Pi 4 Model B with 4GB RAM to identify the given text and display the sign gesture for the text within 0.5 seconds.