SUDHIR MAHAARAJA

+91 7397403525 | sudhirmahaaraja@gmail.com | Chennai |
https://www.linkedin.com/in/sudhirmahaaraja/ |
https://github.com/SudhirMahaaraja

EDUCATION

BACHELOR OF TECHNOLOGY IN ARTIFICIAL INTELLIGENCE AND DATA SCIENCE , *RMK COLLEGE OF ENGINEERING AND TECHNOLOGY* *2020 - 2024*

SKILLS

PROGRAMMING LANGUAGES: Python, SQL.

MACHINE LEARNING: PyTorch, TensorFlow, Keras, Transformer-based modeling, Large Language Models, DecisionTreeClassifier, NLP, OCR, Computer Vision (OpenCV, YOLO), RAG, AutoML (Optuna).

TOOLS AND LIBRARIES: Flask, Streamlit, LangChain, LlamaIndex, Hugging Face Transformers, Pandas, NumPy, N8N, Excel, Power BI, Matplotlib, Selenium, BeautifulSoup, Caffe.

CLOUD SERVICES & DATABASES: AWS (S3, Rekognition), Azure Vision Service, MongoDB, FAISS

INTERNSHIP EXPERIENCE

MACHINE LEARNING INTERN – *CODSOFT* Feb 2024 – Mar 2024, Remote

- Developed a text classification model for movie genres using logistic regression, achieving 85% accuracy.
- Created a Random Forest model for detecting fraudulent transactions, effectively handling imbalanced data.
- Built an SMS spam classifier using a Multilayer Perceptron with real-time classification features.

MACHINE LEARNING TRAINEE – *MERCURYMINDS* Dec 2024 – Jun 2025, On-Site

- Developed a system using YOLO, Caffe, Amazon Rekognition, and Azure Vision Service to detect visitor behavior with 92% accuracy metrics like foot traffic, dwell time, engagement.
- Built a dashboard providing insights on visitor demographics, peak times, staff interactions, and engagement metrics.
- Delivered data to optimize booth operations, enhance visitor experience, and improve staff performance.

PROJECTS

RAG-BASED PDF Q&A

- Built RAG PDF chatbot to extract text, chunk, and index vectors for semantic search.
- Implemented accurate chunk-to-page mapping, overlap-aware chunking, OCR fallback.
- Integrated Groq LLM (Llama 3.3 70B) with Streamlit UI for grounded, page-cited answers.
- Tech stack used: Python, PyPDF2, sentence-transformers, FAISS, Groq, Flask, Tesseract

VIRTUAL HEALTH ASSISTANT

- Developed a machine learning model for disease prediction with an 85% accuracy rate.
- Designed and implemented the user interface and backend integration.
- Conducted extensive testing and validation to ensure model robustness.
- Tech stack used: Python, TensorFlow, DecisionTreeClassifier, NLP, Flask.

TOURIST BEHAVIOR ANALYSIS

- Analyzed tourist data to provide pivotal insights for the travel industry.
- Visualized data using PowerBI to identify patterns and trends.
- Performed data cleaning and preprocessing to ensure data quality.
- Tech stack used: Python, SQL, Selenium, PowerBI.

CERTIFICATE

- Databricks Lakehouse Fundamentals - *Databricks*
- Data Analytics Essentials - *Cisco*
- Advanced Google Analytics - *Google*
- SQL for Data Science - *Great Learning*

PUBLICATIONS

- "Tourist Behaviour Analysis Using Data Analytics" in J. C. Koy, B.V. Parker, & D.A. Lopez (Eds.), Data-Driven Decision Making for Long-Term Business Success (pp. 8-12).

ACHIEVEMENTS & PARTICIPATIONS

- Won the CloudlyML Data Analysis Hackathon 2024
- Participated in the national conference on "Recent Advancements in Artificial Intelligence and Data Science"