SANTHOSH KUMAR S

2/57 North Street Jayankondan, 608601
+91 7812869940
sksanthosh4246@gmail.com
linkedin.com/in/santhoshkumar-s-9404aa298

PROFESSIONAL SUMMARY

AI/ML Engineer with expertise in OCR, Generative AI, Computer Vision, and Machine Learning. Experienced in building OCR pipelines (PaddleOCR, Regex), GenAI chat assistants (LangChain, LLMs), fraud detection systems (pHash, ORB, Qdrant), and sales forecasting models (ARIMA, regression, ensemble ML). Skilled in Python, FastAPI, LangChain, Scikit-learn, SQL/NoSQL, and delivering scalable AI solutions for document automation, claims processing, and analytics

EDUCATION

M.Sc Information Technology [2019 - 2024]
Annamalai University
CGPA 8.5

PROFESSIONAL EXPERIENCE

AI/ML Engineer - ASSIST360 Singapore **Mar 2025 - Present**

Working Projects Details:

OCR Data Extraction System (Python, PaddleOCR, Regex)

- Built a robust OCR pipeline using PaddleOCR integrated with advanced Regex-based parsing to extract structured fields (amounts, dates, invoice numbers, payment methods, and taxes) from receipts and invoices.
- Enhanced extraction accuracy with regex normalization, fuzzy matching, and domain-specific heuristics for global currency and tax formats.
- Improved document parsing performance for large-scale processing with support for multi-format (PDF, JPEG, PNG) inputs.

Generative AI Chat Assistant (LangChain Framework)

- Developed an intelligent chatbot using LangChain, integrating LLMs for dynamic query handling, claims-related Q&A, and contextual document retrieval.
- Combined vector-based document search (FAISS/Qdrant) with FAQ retrieval to provide accurate and policy-compliant responses.
- Designed modular APIs for external system integration, enabling deployment in both cloud and on-premise environments.

Fraud & Duplicate Receipt Detection (Computer Vision + Vector Search)

- Implemented a receipt fraud detection system leveraging Perceptual Hashing (pHash) and ORB feature matching for visual similarity detection.
- Used Qdrant Vector DB with embedding + cosine similarity search to detect duplicate or fraudulent receipt uploads across large datasets.
- Integrated with FastAPI backend and MongoDB/PostgreSQL storage for real-time duplicate detection and analytics.

Sales Forecasting (Machine Learning)

- Built and deployed a time-series sales forecasting model using regression, ARIMA, and ensemble methods to predict future sales trends.
- Improved forecast accuracy with feature engineering, seasonal trend analysis, and hyperparameter tuning.
- Delivered dashboards for decision-making using Python (Pandas, Scikit-learn, Matplotlib).

Network Intrusion Detection Analysis using Machine Learning **Feb 2024 - Mar 2024**

- This project focuses on analysis of various machine learning algorithm performance in classification and detection of network intrusion detection.
- The project aims to provide insights into the strengths and weaknesses of each machine learning algorithm, offering a comparative analysis that aids in selecting the most suitable approach for network intrusion detection.

SKILLS

- Python Fullstack Developer
- Web Development - HTML, CSS, JavaScript.
- API - Fast API,REST API
- Database - MySQL, SQL,Mongo DB,PostgreSQL
- AI / ML - Pandas, NumPy, Matplotlib, OpenCV, SciKit, Tensorflow, Keras, SCIPy, Theano, Seaborn, XG BOOST, Light BGM,NLTK
- Machine Learning - supervised, Un-supervised Learning Algorithm
- Cloud Computing - S3 ,VPC, EC2, RDS, IAM, Cloud Watch,ELB, ROUTE 53,SNS + 20 Services

CERTIFICATIONS

- Artificial Intelligence with Python | Great Learning
- Software Development Methodology in Agile | IBM -SkillBuild
- Learning Python -Infosys | Springboard
- Implementing DevOps with AWS - Infosys | Springboard
- Data Analytics in AWS | Coursera
- Excel for Data & Analytics Completion Of Certificate | CHEGG