SANJAY VISHWAKARMA

Ahmedabad, Gujrat 382424

+919664538358 | sanjayskumar4010@gmail.com | linkedIn

SUMMARY

Data Scientist passionate about transforming data into actionable insights and driving business impact. Currently developing AI-based solutions, including real-time YOLO-based systems and LLM-driven natural language models. Leading efforts to build AI agents for demand forecasting and enhancing data governance and ETL pipelines to ensure data consistency and operational efficiency.

EDUCATION

B.Tech in Information Technology

2020 - 2024| Aditya Silver Oak University   CGPA: 9.59

Vidhya GPT – Multilingual Educational Chatbot (Final Year Project)

• Developed an AI-powered multilingual educational chatbot using open-source LLMs (LLaMA 2, Falcon) and Groq API for real-time, personalized learning assistance, enhancing user engagement by 30%.

• Designed a context-aware processing pipeline using Named Entity Recognition (NER) and intent classification to extract user context and fetch relevant learning resources from a structured knowledge base and third-party APIs.

• Applied transformer-based models and adaptive learning algorithms to generate dynamic, personalized responses, improving learning efficiency by 25% and boosting user retention by 20%.

• Enabled instant access to educational resources (articles, quizzes, videos) through an interactive interface, providing accurate, tailored recommendations and enhancing the overall learning experience.

Higher Secondary Education

2018 - 2020 | I.P. Vidyalaya, Ahmedabad   PERCENT: 64%

WORK EXPERIENCE

Associate Data Analyst | Azilen Technologies Pvt. Ltd. (Rootle.AI – Sister Company), Ahmedabad, India

Worked with Rootle.AI (sister concern of Azilen) for **1.7 years** before officially transferring into Azilen Technologies. Continued in the Software engineer (**GenAI**) role, focusing on **Generative AI, LLMs, and AI/ML solutions**.

JAN 2024 - Present

Data Analyst Trainee | GTCSYS Private Limited, Ahmedabad, India

SEP 2022 - NOV 2023

Data Science Intern | Maxgen Technologies, Ahmedabad, India

JUNE - JULY 2022

DOMAINS WORKED ON

• Natural Language Processing (NLP) • Large Language Models (LLMs) • Retrieval-Augmented Generation (RAG) • Prompt Engineering • Computer Vision • Image Processing • Deep Learning • Machine Learning • Speech Processing • Data Engineering • Data Governance • ETL Pipelines • Multi-Agent Systems (MAS) Generative AI (Lang Graph, Autogen, AI Avatar)

LIBRARIES AND FRAMEWORKS

- **PROGRAMMING LANGUAGES:** Python
- **DATA MANIPULATION AND ANALYSIS:** Pandas, NumPy, SciPy, plotly, Seaborn, Scikit-learn
- **NLP:** NLTK, Spacy, Gensim, Transformers, Langchain, Prompt Engineering
- **DATA ENGINEERING AND GOVERNANCE:** DataHub, Pyspark, Airflow, Superset, ClickHouse, PostgreSQL, Guardrails
- **COMPUTER VISION:** OpenCV, YOLO, Detectron
- **DEEP LEARNING FRAMEWORKS:** TensorFlow, PyTorch
- **WEB FRAMEWORKS:** Django, Flask, FastAPI, Swagger
- **AUDIO PROCESSING:** Pydub, PyAudio
- **MULTI-AGENT & AI-AVTAR:** Multi-Agent Systems, AutoGen, LangGraph, AI Avatar

PROJECTS

Automated Data Governance And ETL Workflow

• Designed and implemented a secure ETL pipeline to extract data from PostgreSQL, transform using PySpark, and load into ClickHouse for real-time analysis, ensuring 98% data accuracy and consistency with Debezium for change data capture (CDC).

• Strengthened data security and compliance by enforcing access control and visibility rules through DataHub, ensuring only authorized users can access specific data based on governance policies. For example, HR and Finance teams can only access employee data, while Sales leads can only access sales data — improving data security by 40% and reducing unauthorized access incidents by 30%.

• Automated the workflow with Apache Airflow, incorporating SMTP alerts for failure detection and quick recovery, reducing pipeline downtime by 25% and improving operational efficiency by 20%.

• Enhanced decision-making with interactive visualizations using Superset and improved security monitoring with Wazuh, leading to a 15% reduction in data discrepancies and faster response to anomalies.

Supply Chain Analytics & Inventory Management

• Built a Python and XGBoost-based supply chain solution for predictive demand forecasting and real-time inventory optimization, reducing stockouts by 15% and improving order fulfillment by 20% with a Chart.js dashboard.

• Engineered 35+ predictive features for multi-horizon forecasting (30, 45, 60, 90 days), achieving 75–80% accuracy (MAPE ~20, MAE ~10) by analyzing seasonal patterns and large historical datasets.

• Applied advanced feature engineering (lagging, rolling windows, EMA) and optimized XGBoost using grid/random search, outperforming SARIMAX and Prophet, improving inventory turnover by 12%.

• Developed a FastAPI-based GenAI chatbot for real-time insights, reducing manual analysis time by 30%, and automated inventory management with dynamic reorder points and safety stock levels.

Advanced Edge Computing Vision System with Multi-Modal Detection

• Developed an edge computing vision system using RKNN-3566 and YOLOv11n for real-time multi-class detection (gender, age, expression, brand), achieving 94% accuracy and reducing CPU load by 40% by offloading to the NPU.

• Designed a multi-threaded pipeline with frame-skipping and queue management, reducing inference time by 25% and enabling stable 30+ FPS live feed.

• Optimized deep learning models with RKNN-Lite, handling model loading and inference on the NPU, reducing overall latency by 20%.

• Finalized YOLOv11n over YOLOv8 and YOLOv5 for superior accuracy and low latency, achieving 92% mAP and improving detection consistency by 15% with Bayesian updating, Kalman filtering, and adaptive thresholding.

AI-Driven Test Case Generation for Agile Software Development

• Developed an AI-driven test case generation system using LLaMA2-70B-8192 with Chain of Thought prompting, outperforming other LLMs (GPT-4, LLaMA, Gemini) across 30 experiments, achieving higher accuracy and consistency in generating detailed test cases.

• Evaluated prompting techniques (Zero-shot, Few-shot, and Chain of Thought) and integrated Tree of Thoughts (ToT) framework, achieving superior performance with LLaMA2-70B-8192, improving ROUGE, BERTScore, and BLEU metrics by 18%.

• Enhanced interpretability and reliability using Explainable AI (XAI) techniques, including LIME and SHAP, ensuring improved traceability and alignment with ISTQB industry standards.

• Developed Story2TestGPT, an AI-based tool enabling Quality Assurance professionals to automate test case generation and analysis, increasing test case coverage by 25% and reducing manual effort by 30%.

Financial Analytics Dashboad & Query System

• Developed an AI-powered financial analytics platform using OpenAI API and LangChain, achieving high accuracy in processing complex financial queries across large-scale multi-company records with response times under 2 seconds.

• Built a context-aware query processing system using LangChain and AgentExecutor, enabling multi-turn conversations and intelligent data retrieval, increasing query resolution efficiency by 80% and improving user satisfaction by 85%.

• Designed a real-time data pipeline with Pandas and NumPy, integrating Highcharts and Streamlit for dynamic financial trend visualization, reducing data interpretation time by 75% and improving forecasting accuracy.

• Implemented system monitoring and logging using Python-based solutions, maintaining 90% uptime and ensuring consistent query resolution under 3 seconds, enhancing system reliability and scalability.

Movie Recommendation Platform

- Developed an AI-driven recommendation system using NLTK Porter Stemmer and Count Vectorizer for genre-based suggestions, improving accuracy by 25%.
- Integrated a real-time data pipeline to source and update movie data from APIs and databases, ensuring fresh content recommendations.
- Engineered a hybrid engine combining content-based and collaborative filtering for personalized suggestions, enhancing precision by 30%.
- Designed a user-specific dashboard with Stream lit, featuring dynamic updates, visual enhancements, and API-driven content for improved engagement.

CERTIFICATIONS & COURSES

- Generative AI with LLM Models – Deeplearning.ai
- Google Advanced Data Analytics – Coursera
- Diploma in Multilingual Computer Programming (DMCP) – C-DAC (Govt. of India)
- Multiple Badges – Google Cloud Skills Boost (Ongoing)