Sandeep Sah

Data Scientist | Machine Learning | Artificial intelligence | NLP | GEN AI

Contact

Phone
+91 6362322548

Email
sandeepsah50714@gmail.com

Github
https://github.com/SandeepSah12

Linkdin
https://www.linkedin.com/in/sandeep-sah-bb9000273/

Address
Hyderabad, India

Skills.

- Programming Language : | Python | | R Programing |
- Databases : | SQL |
- Cloud : | AWS | AZURE ML | Azure Ai
- MLOps : | Git | GitHub | MLflow | Airflow |
- Data Visualization: | Tableau | | Power BI | Automated EDA |
- Big Data : | pyspark | Hadoop |
- Key skills : | Data Visualization | Predictive Analysis | Statistical Modeling | Data Preprocessing | Clustering And Classification | Time Series Analysis And Forecasting | Machine Learning | Deep Learning |
- Packages : | Numpy | Pandas | Matplotlib | Seaborn | Scikit-learn | Scipy | NLTK | Spacy | Mediapipe | TensorFlow | Keras | OpenCV | Flask |
- Statistics : | Inferential Statistics & Descriptive Statistics |
- Machine Learning :| Linear and logistic regression | Gradient descent | Decision Tree | Random-Forest | SVM | Naive-Bayes | KNN | Xgboost | Adaboost | Polynomial Regression | Lasso & Ridge Regression | Gini | Entropy | AUC & ROC | Clustering | Cross Validation |
- Deep Learning : | ANN | CNN | RNN | LSTM | Optimization Techniques | Transfer Learning | Encoders & Decoders | Attention Model |

Summary

Data Scientist with 3.2 years of hands-on experience in Python, data analysis, machine learning, and Artificial intelligence developer, Computer vision, NLP, Generative Ai, and Agentic Ai, LLMs, GPT-4 statistical modeling. Proficient in Python, R, AWS Sagemaker, MLOps, SQL, Time series, Forecasting and data visualization tools Power BI, with a track record of delivering data-driven solutions and actionable insights. Strong problem-solving skills and the ability to communicate complex findings to both technical and non-technical stakeholders. Passionate about leveraging data to drive business decisions and optimize.

Work Experience

Data Scientist
Aria Aerotech Pvt Ltd
May 2022 - Jun 2025 **(Bengaluru, India**

- Data Cleaning, Preprocessing & exploration: Prepare data for analysis, ensuring quality, consistency, and completeness by handling missing values, outliers, and transforming data. Explore and analyze large and complex datasets to identify patterns, trends, and anomalies
- Machine Learning Model Development
- Design Machine Learning, Natural Language and Decision Optimisation applications architecture.
- Design ML, NLP and DO models, algorithms for predictive & prescriptive analytics.
- Translate business requirements to reporting dashboard and analytics.
- Perform advanced analytics and statistical modelling on structured and unstructured data.
- Work with software engineers to integrate and deploy AI applications.
- Design, develop, and deploy machine learning models and data science solutions.
- MLOps/LLMOps platform development & automated pipelines focusing on deploying
- Work with large datasets, preprocess and analyze data for insights.
- Implement and optimize deep learning and statistical models.
- Design and implement machine learning models and Artificial intelligence algorithms for various use cases (e.g., prediction, classification, NLP, computer vision).
- Design and implement generative AI models (e.g., LLMs, diffusion models) for text, image, audio, or multimodal content generation.
- Develop agentic AI systems capable of autonomous decision-making, planning, and tool use in complex environments.
- Integrate AI agents with APIs, databases, and external tools to enable real-world task execution.
- Fine-tune foundation models for domain-specific applications using techniques like RLHF, prompt engineering, and retrieval-augmented generation (RAG).
- Familiarity with Ai and ML frameworks such as Tensorflow, PyTroch, Keras,
- Familiarity with Generative ai and Agentic ai frameworks such as Langchain, Langgraph, AutoGPT, OpenAgents, CrewAi, Hagging Face.
- Deploy models to production using tools such as Docker, Kubernetes, or cloud services (AWS, GCP, Azure)

Projects

- Natural Language Processing : | Word Embedding | tf-idf | word2Vec | Bag of Words | Tokenization | Stemming | Lemmatization | Text Summarization | Web Scrapping | Speech Recognition
- Generative Ai | GAN | LLMs | GPT | ML | BERT | T5 | LANGCHAIN | Gemini-pro| | Google gemini-pro-vision | Hagging Face
- Agentic Ai | LangGraph | Crew Ai | Grok | Deepseek | Vertex Ai | Embeding | ChromaDB |
- Open Ai | Bloom | LIama2 | Falcon | DALL-E2 | ChatGPT | Whisper | GAMMA2 | Meta Ai | Gemini Ai | Llama3 |
- Tools : | Jupyter notebook | Google Collab | Spyder | Visual Studio |
- Software Skills : | MS Office | MS Word | Excel |

Education

2018 - 2022
B.TECH (Computer Science) 8.37CGPA
Rajiv Gandhi Proudyogiki Vishwavidyalaya (Bhopal)

CERTIFICATION

- SPSS Modeler Professional V3 (IBM)
- Data science Certificate by Microsoft
- Artificial intelligence certificate by Microsoft
- Machine learning certificate by IBM.
- image processing Certificate by Cisco.

Operating systems

- Microsoft Windows.
- Linux.

Language.

- English.
- Hindi.

Solving complex problems

- Python,
- Machine Learning,
- Deep Learning,
- Data science

TripChain. Ai Using Agentic ai
Role: Gentic Ai Researcher

Developed an AI-powered assistant capable of interpreting natural language instructions and executing complex, multi-step tasks—such as booking flights, reserving hotels, and scheduling calendar events—by coordinating external APIs through LLM-based planning

Technologies Used: Python, Git, GitHub, Docker, LLMs, Open Ai, Gemini, GPT-4, Grok, RAG, Langchain, LangGraph, AutoGen, Hagging Face, Vector-Embeding, GCP, Vector DB, LLMOps, Llama, AWS-bedrock, faiss-GPU, Fine-tuning, RHLF.

Product Recommendation Using LLMs
Role: Generative Ai Researcher/Scientist

An end-to-end generative AI project is presented, focusing on building a product recommendation chatbot for e-commerce using Flipkart data. The process involves data ingestion, creating a chatbot using a large language model, and deploying the application using Docker and AWS EC2, Speech - text - Speech

Technologies Used: Python, Git, GitHub, Docker, LLMs, Open Ai, Gemini, GPT-4, Grok, RAG, Langchain, Hagging Face, Vector-Embeding, GCP, ChromaDB, Llama, LLMOps, AWS-bedrock, AWS EC2, faiss-GPU, Fine-tuning, RHLF.

Vehicle Detection & Speed Estimation using YOLO V8
Role: Computer Vision Engineer / AI Developer

This project is an advanced Vehicle Speed Detection and Tracking System using YOLO, OpenCV, and PaddleOCR, integrated with a Django-powered dashboard for real-time monitoring. It detects vehicles from a video stream, estimates their speed, extracts number plates using OCR, and sends the data to a Django API. The data is then displayed on an interactive dashboard with analytics & Features

Technologies Used: Python, numpy, Pandas, YOLO V8, Tracker, CV2 Open CV, CNN, Speech recognition, AWS, GCP, SQL, Flask, git github

Credit Finance Using MLops
Role: Machine Learning Engineer

The objective of this project is to develop a machine learning model that predicts whether a customer's loan application will be approved or not based on their financial and demographic details. By leveraging Python and machine learning techniques, we aim to create an efficient, data-driven approach to assist financial institutions in making informed lending decisions

Technologies Used: Python, Numpy, Pandas, Matplotlib, Sebron, Skit-learn, Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting. Git, GitHub, MLflow, SQL, AWS, Fast API, faiss-GPU.

Sentiment Analysis Using NLP
Role: NLP Engineer

The food industry is highly competitive, and customer reviews play a crucial role in shaping restaurant reputations. Analyzing these reviews and restaurant data can provide valuable insights for both customers and business owners

This project aims to analyze customer reviews to extract meaningful insights using Natural Language Processing (NLP)

Technologies Used: Python, Numpy, Pandas, Matplotlib, Sebron, EDA, NLP, NLTK, Pytorch, Skit-learn, Wordcloud, Clustering, K-Means, TF-IDF, SVM, Word2Vec, stopwords, lemmatization, streaming, tokenization, and GridSearchCV, PCA, Git, , GitHub,, SQL, AWS, Fast API, faiss-GPU.