Sandeep Kumar Are LinkedIn
Mobile: +91 7989520772 Personal website
E-mail: sandeepkumar.are16@gmail.com AWS Certified AI Practitioner

CAREER OBJECTIVE

A proactive and fast learning individual seeking an opportunity to work as a dynamic data scientist/engineer, utilizing analytical & technical skills to help the company achieve business goals while driving innovation through AI, ML, and cloud solutions.

PROFESSIONAL SUMMARY

- 10+ years of IT software development experience, including 5+ years specializing in Python, Machine Learning, and Generative AI.
- Expertise in building end-to-end AI/ML solutions: data preprocessing, feature engineering, model training, deployment, and monitoring.
- Hands-on experience with Generative AI (LLMs, RAG, LangChain, Mistral, OLLAMA).
- Strong background in NLP, predictive modeling, and anomaly detection.
- Skilled in AWS cloud services (ECR, ECS, EKS, SageMaker, Lambda, SAM, S3, EC2) and CI/CD pipelines.
- Experience in data engineering pipelines, MlOps, and model lifecycle management with MLFlow.
- Strong collaboration skills with cross-functional teams to deliver scalable AI solutions.
- Conduct text-based analysis to derive insights and patterns from unstructured data sources.
- Perform data preprocessing tasks, including data cleaning, feature engineering, and transformation to prepare data for modeling.
- Implement machine learning algorithms for classification and regression tasks, optimizing model performance and accuracy.
- Utilize Qlik Sense at a beginner level to develop interactive data visualizations and dashboards for stakeholders.
- Implement MlOps practices, including version control, continuous integration/continuous deployment (CI/CD), and model monitoring to ensure efficient and scalable machine learning workflows.
- Develop and maintain documentation for data science projects, including model architectures, workflows, and deployment procedures.
- Stay updated on emerging technologies and best practices in data science, machine learning, and cloud computing to drive innovation and enhance capabilities.

Technical Skills:

- Generative AI & LLMs: Lang Chain, RAG, Mistral, OLLAMA, Transformers.
- Cloud & Deployment: AWS (ECR, ECS, EKS, SageMaker, SAM, S3, EC2, Lambda), CI/CD pipelines.
- Machine Learning & AI: Predictive Analytics, NLP, Regression, Anomaly Detection, ML/DL (CNN, RNN, LSTM).
- Data Engineering: Snowpark pipelines, SQL, PostgreSQL.
- Visualization: Qlik Sense, Matplotlib, Pandas
- MlOps, MLFlow, Model Monitoring, CI/CD
- Programming Language: Python.

WORK EXPERIENCE

Data Modeller - Manager-C2 - Capgemini **January -2025 - Till Date | Hyderabad, India**

ML Engineer - Capgemini Invent | Client: Entertainment domain (On-site)

- Contributed to the development of a GenAI-based automated billing generation system using AWS Lambda, Large Language Models, Azure Chat OpenAI, AWS SAM, and AWS ECS.
- Played a key role in building an Ad-Sales Chatbot— a complete RAG-based application with multiple integrated modules. Worked extensively on the RAG pipeline, including data ingestion, chunking, embedding, and retrieval workflows, leveraging LLMs, OpenAI APIs, embedding models, retrieval models, and prompt-engineering templates.
- Developed an AI-driven system to automatically generate Course Outcomes (CO), Program Outcomes (PO), and Program-Specific Outcomes (PSO) based on keywords and user queries (PCO-style generation).
- Contributed to the development of an Academic AI tool capable of generating detailed documentation using various input sources such as video links, website URLs, documents, and raw text. Enabled outcome editing and regeneration through specialized prompts using Lang Chain, embedding models, GPT-4-mini, and prompt-orchestration frameworks.
- Designed, developed, and deployed scheduled data-engineering pipelines on AWS ECS.
- Gained hands-on experience with Snowflake data warehouse, including building and optimizing data engineering pipelines within the Snowflake environment.
- Worked on AWS SageMaker model development and deployment using MlOps practices, integrating MLFlow for experiment tracking and model lifecycle management.

Senior Software Engineer – ATMECS Technologies

Jun 2022 – January 2025 | Hyderabad, India

Roles & Responsibilities:

- Delivered multiple client projects involving supervised, unsupervised, and multi-class classification models.
- Applied NLP techniques using NLTK, SpaCy, and Transformer-based models for text analytics.
- Developed GenAI-based applications focused on text generation.
- Built predictive models for payroll forecasting and anomaly detection.
- Deployed and monitored ML models, integrating performance dashboards in Qlik Sense.

Senior Software Engineer – Svobodha Infinity (Savart)

February 2020 – June 2022 | Hyderabad, India

Project: Qualitative Research (IRIS)

Iris is an AI-driven qualitative research system that analyzes textual, visual, and graphical data to uncover insights crucial for investment decisions. It identifies hidden patterns—such as compliance issues, governance risks, ethical concerns, or irregularities—across over 850 topics including ethics, corporate governance, brand strength, moat, and innovation.

The system processes multi-source data (subsidiary information, local jurisdiction compliance, company disclosures, and third-party data) to highlight non-compliances and generate actionable insights at scale.

Roles & Responsibilities:

- Scraped data from various websites, APIs, and external sources using Python and Selenium.
- Worked extensively with Python and MongoDB for data storage, processing, and integration tasks.
- Performed NLP-based topic modelling to categorize and extract insights from unstructured text data.
- Collaborated with the team on implementing and fine-tuning BERT-based models.
- Developed UI components using Flask (Python) for internal workflows and data visualization.

Project: Quantitative Analysis (QUANT)

Quant is a machine learning–based quantitative research engine designed to analyze billions of data points (financial statements, market data, statistical indicators) to generate unbiased, long-term investment strategies. Unlike traditional methods limited by human assumptions, QUANT builds patterns from scratch, performs deep stress-testing on historical and real-time data, and outputs optimized portfolios for further evaluation by IRIS.

Roles & Responsibilities:

- Evaluated and experimented with multiple modeling approaches to identify the most effective solutions.
- Performed data cleansing, preprocessing, and analytical transformations to prepare large-scale datasets.
- Implemented feature selection techniques to enhance model accuracy and reduce noise.
- Contributed to forecasting model development for generating predictions and investment signals.

Software Engineer – Netpeach Technologies

February 2015 – January 2020 | Hyderabad, India

Worked across multiple roles and technologies, beginning as a Junior Developer and progressing to Software Engineer. Contributed to enterprise application development, intranet portal solutions, and data analysis initiatives, including healthcare-focused research projects**.**

Project Name: Uncover – Autism Analysis Project

The objective of this project was to support healthcare professionals in identifying more reliable factors associated with autism in newborns. The work involved exploring existing scientific research, analyzing available datasets, and collecting insights through patient surveys to identify meaningful patterns. The goal was to provide the healthcare industry with analytical findings that could support early detection and risk mitigation.

Tools & Technologies:

- Environment: Anaconda, Jupyter Notebook
- NLP Tools: NLTK
- Data & Visualization Libraries: NumPy, Pandas, Matplotlib, Seaborn
- Machine Learning: Scikit-learn

Education

B.Tech in Electronics & Communication Engineering – JNTU University, India (2012).