

Anil Kudumula

Data Scientist

+1 (740) 298-4590/ kudumula.anil74@gmail.com

Professional Summary

- Over 4 + years of experience in **data science, machine learning**, and **artificial intelligence**, delivering impactful insights and driving data-driven decision-making.
- Expertise in **statistical modelling, data mining, predictive analytics**, and **deep learning** to solve complex business challenges.
- Proficient in **Python, R, SQL, TensorFlow, PyTorch**, and cloud-based analytics platforms, including **AWS, GCP, and Azure**.
- Extensive experience in building end-to-end machine learning pipelines, optimizing models for performance, and deploying scalable solutions.
- Strong background in big data technologies such as **Hadoop, Spark**, and **Kafka** for real-time data processing.
- Skilled in exploratory data analysis, data visualization, and reporting using tools like **Tableau, Power BI**, and **Matplotlib**.
- Experience in **A/B testing**, hypothesis testing, and experimental design to validate data-driven strategies.
- Proven ability to work with cross-functional teams, bridging the gap between technical and non-technical stakeholders.
- Expertise in **Natural Language Processing (NLP)** and computer vision for text and image analysis applications.
- Adept at developing recommendation systems, customer segmentation models, and predictive forecasting solutions.
- Strong experience in **data governance, data security**, and compliance with regulatory requirements.
- Hands-on experience in optimizing data pipelines, **ETL processes**, and feature engineering for model enhancement.
- Excellent problem-solving and analytical skills with a keen eye for identifying trends and business opportunities.
- Experience leading and mentoring data science teams, conducting technical training, and establishing best practices.
- Passionate about continuous learning, research, and applying cutting-edge techniques to improve business outcomes.

Technical Skills:

Programming Languages	Python, R, SQL, Java, Scala
Machine Learning	TensorFlow, PyTorch, Scikit-learn, Keras
Data Engineering	Hadoop, Spark, Kafka, Airflow, ETL
Cloud & DevOps	AWS (S3, Lambda, Sage Maker), Azure, GCP
Visualization & BI	Tableau, Power BI, Matplotlib, Seaborn
Databases	PostgreSQL, MySQL, MongoDB, Big Query
NLP & Computer Vision	NLTK, spaCy, OpenCV, Transformers
Big Data & Analytics	Hive, Presto, Redshift, Snowflake
MLOps & Deployment	Docker, Kubernetes, CI/CD, Flask, FastAPI

Professional Experience:

Client: Ally Bank Detroit, MI

July 2024 – Nov 2024

Data Scientist

- Designed and deployed **machine learning models** for customer behaviour prediction, increasing customer retention rates by 15%.
- Optimized **ETL** pipelines and automated data workflows, reducing data processing time by 40%.
- Implemented A/B testing frameworks, improving conversion rates and optimizing marketing strategies.
- Developed **NLP** models for sentiment analysis, enhancing brand perception analysis and customer feedback processing.
- Created real-time fraud detection models using anomaly detection techniques, reducing fraudulent transactions by 25%.
- Fine-tuned open-source **GenAI models** (LLaMA 2, Mistral) on custom datasets using LoRA/PEFT to improve domain-specific language understanding.
- Built and deployed deep learning models for image classification and object detection applications.
- Explore and experiment with cutting-edge tools and frameworks such as Hugging Face Transformers, Lang Chain, LlamaIndex, and vector databases like FAISS and Weaviate.
- Deployed ML and **GenAI models** as REST APIs using FastAPI and Docker on AWS Lambda/Sage Maker, enabling scalable inference.
- Implemented Retrieval-Augmented Generation (RAG) architecture using FAISS and Lang Chain to provide real-time, context-aware document answers.
- Regularly read and summarize state-of-the-art papers, including breakthroughs in transformers, large language models (LLMs), and retrieval-augmented generation (RAG).
- Developed scalable data pipelines using Apache **Spark**, processing terabytes of data efficiently.
- Spearheaded initiatives to implement explainable AI techniques, increasing transparency in model decision-making.
- Integrated cloud-based solutions for large-scale data processing, enhancing computational efficiency.
- Worked closely with engineering teams to deploy models into production using MLOps best practices.
- Integrated multimodal **GenAI** models (e.g., BLIP, DALL·E) to extract insights from image-text datasets for marketing analytics.
- Conducted regular model retraining and performance monitoring to ensure continuous improvement.
- Participated in roadmap planning and solution design to integrate ML/NLP models into production systems and customer-facing products.
- Developed an LLMOps pipeline to manage lifecycle of **GenAI models**, including evaluation, versioning, and prompt testing.
- Established data governance policies to improve data quality, security, and compliance with industry standards.
- Ensured data pre processing, tokenization, and evaluation followed NLP best practices with metrics like F1-score, ROUGE, and coherence score.
- Developed internal GenAI tools using Streamlit and FastAPI, enabling non-technical users to interact with LLM-powered features.
- Provided mentorship and technical guidance to junior data scientists, fostering a culture of innovation.

Environment: Python, R, SQL, TensorFlow, PyTorch, Scikit-learn, AWS (Sage Maker, Lambda, Redshift), Spark, Hadoop, Tableau, Power BI, Docker, Kubernetes, Airflow, Git, Snowflake.

Client: Nivis Info Pvt Ltd
Data Scientist

Sep 2022 – Nov 2023

- Designed and implemented predictive analytics models, reducing churn rates by 18% through targeted interventions.
- Developed and deployed deep learning models for text classification and entity recognition in large datasets.
- Built machine learning pipelines to automate data preprocessing and model training, reducing manual efforts.
- Analysed customer behaviour patterns to improve personalized marketing campaigns, increasing revenue.
- Implemented data visualization dashboards in **Tableau**, providing executives with real-time insights.
- Collaborated with cross-functional teams—including engineering, product, and business stakeholders—to identify and prioritize high-impact data science opportunities.
- Developed recommendation algorithms, increasing product adoption and user engagement.
- Worked on reinforcement learning models for optimizing dynamic pricing strategies.
- Improved data engineering workflows, enhancing data accessibility and reducing latency.
- Applied advanced statistical techniques for demand forecasting and inventory optimization.
- Conducted deep dives into unstructured data sources to extract meaningful insights.
- Enhanced model interpretability using SHAP and LIME, increasing stakeholder trust in AI-driven solutions.
- Collaborated with software engineers to deploy machine learning models in microservices architecture.
- Conducted workshops on best practices in data science and machine learning for cross-functional teams.
- Researched and experimented with novel AI techniques to maintain technological competitiveness.

Environment: Python, R, SQL, TensorFlow, Keras, Scikit-learn, GCP (Big Query, Dataflow, AI Platform), Spark, Kafka, Tableau, Power BI, Docker, Kubernetes, Airflow, Git, Snowflake.

Client: Neptune information solutions limited
Junior Data Scientist

Sep 2020 – Aug 2022

- Developed machine learning models to optimize pricing strategies, leading to a 12% increase in revenue.
- Implemented a fraud detection system using anomaly detection techniques, reducing fraudulent transactions by 30%.
- Spearheaded a project to automate data processing pipelines, decreasing **ETL** time by 50%.
- Designed and deployed a chatbot powered by **NLP** to improve customer support efficiency.
- Built deep learning models for sentiment analysis, providing valuable insights into customer feedback.
- Conducted exploratory data analysis to uncover new market opportunities and enhance strategic planning.
- Improved model interpretability using SHAP, ensuring transparent AI-driven decision-making.
- Optimized cloud-based data storage solutions, reducing infrastructure costs by 20%.
- Provided training and mentorship to junior data scientists, fostering a culture of continuous learning.
- Conducted statistical analysis to refine product positioning and market segmentation.
- Designed **scalable** machine learning pipelines, ensuring efficient deployment in production environments.
- Developed reinforcement learning models to optimize customer engagement strategies.

- Researched cutting-edge AI methodologies and applied them to solve business challenges.

Environment: Python, R, TensorFlow, Scikit-learn, AWS (S3, Lambda, Redshift), Spark, Tableau, SQL, Docker, Kubernetes.

Education details:

- Masters In Campbellsville University 2024, USA