

SAJAN THAPA

Pokhara, Nepal | +977 9761871988 | tsajan001@gmail.com | github.com/Coolsajan

PROFESSIONAL SUMMARY

Data Scientist with proven expertise in Machine Learning, NLP, Computer Vision, and Generative AI. Top 7% Kaggle competitor with 5+ deployed ML projects. Skilled in Python, SQL, deep learning frameworks (TensorFlow, PyTorch), cloud deployment, and delivering data-driven solutions. Seeking entry-level Data Scientist, ML Engineer, or Data Analyst roles.

TECHNICAL SKILLS

Programming: Python (Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, Keras), SQL, MongoDB, HTML, Flask, Streamlit

Machine Learning: Supervised/Unsupervised Learning, Deep Learning (ANN, CNN, LSTM), NLP, Computer Vision, Time Series, Feature Engineering, Hyperparameter Tuning, Statistical Modeling

Generative AI: LangChain, RAG, Prompt Engineering, OpenAI API, Hugging Face Transformers, Vector Databases

MLOps & Tools: Docker, CI/CD, REST APIs, Model Deployment, AWS, Power BI, Matplotlib, Seaborn, Plotly, Excel, Web Scraping

VIRTUAL EXPERIENCE

British Airways - Data Science Virtual Intern | Forage

May 2025

- Conducted sentiment analysis on 50,000+ customer reviews using NLP techniques to identify service improvement areas
- Built Random Forest model achieving 82% accuracy to predict customer buying behavior and inform marketing strategy

Deloitte Australia - Data Analytics Virtual Intern | Forage

2025

- Analyzed business datasets to identify trends supporting strategic decision-making through interactive dashboards
- Applied data cleaning and transformation techniques to ensure data quality and integrity for stakeholder presentations

PROJECTS

Flight Log Analyst - Multi-Agent AI System | Python, AutoGen, LLMs, OpenAI API

github.com/Coolsajan/flight-log-analyst

- Engineered 5-agent autonomous system using OpenAI AutoGen to automate flight log analysis, reducing manual time by 90%
- Integrated computer vision and NLP for automated risk assessment and report generation from flight log images

Great Food Buddie - RAG Chatbot | NLP, LangChain, Prompt Engineering, Hugging Face

github.com/Coolsajan/great-food-buddie

- Developed RAG-based AI chatbot processing 10,000+ restaurant reviews using vector embeddings and semantic search
- Deployed production application on Hugging Face Spaces with 95% user query satisfaction rate

Stock Price Prediction System | LSTM, Docker, Streamlit, Time Series

github.com/Coolsajan/stock-prediction

- Built LSTM deep learning model forecasting stock prices with 85%+ directional accuracy over 10-day horizon
- Engineered time series features (moving averages, RSI, volatility) and containerized with Docker for deployment

KAGGLE COMPETITIONS

Aeroclub RecSys Cup

Aug 2025 | Rank: 150/2,298 (Top 7%)

- Designed collaborative filtering recommendation system with matrix factorization and gradient boosting, improving baseline by 23%

Bird CLEF Audio Classification

May 2025 | Top 30%

- Applied CNN architectures on Mel-spectrogram features for bird species classification using transfer learning techniques

CERTIFICATIONS

Oracle Cloud Infrastructure Data Science Professional | Oracle | 2025

Google Data Analytics Professional Certificate | Google (Coursera) | Expected Oct 2025

AWS Cloud Practitioner | AWS Training | 2025

SQL (Advanced) | Python Programming | HackerRank | 2024

EDUCATION

Self-Taught Data Science & Machine Learning

2020 - Present

- Completed 500+ hours through Coursera, Krish Naik Academy, CampusX, and DataCamp covering Statistics, ML, Deep Learning, NLP, MLOps
- Built portfolio of 5+ production-grade projects demonstrating end-to-end ML lifecycle expertise