

KALYAN TOPPO

Email: kalyan.toppo.2021@gmail.com | New Delhi

LinkedIn: linkedin.com/in/kalyantoppo | GitHub: github.com/KToppo

PROFESSIONAL SUMMARY

Associate Data Scientist skilled in statistical modeling, exploratory data analysis, feature engineering, machine learning, and deploying end-to-end analytical solutions. Experienced with real-world datasets across healthcare, finance, and food-tech domains. Strong foundation in Python, ML algorithms, visualization, and data pipelines. Seeking an entry-level Data Scientist role to apply analytical and ML expertise.

CORE PYTHON & DEVELOPMENT SKILLS

- Python (OOP, Functions, Modules, File Handling)
- API Integration (TMDB API, KaggleHub, NSElib)
- Data Processing Pipelines (Pandas, NumPy)
- ML Model Integration & Deployment
- Streamlit App Development
- Logging, Error Handling, Clean Code Practices
- Version Control: Git & GitHub

MACHINE LEARNING SKILLS

- Model Development: Classification, Regression, Recommendation Systems
- Feature Engineering, Vectorization, TF-IDF, CountVectorizer
- Ensemble Models: Random Forest, XGBoost, Gradient Boosting, Stacking
- Balancing Techniques: SMOTE, SMOTENNN
- Model Serialization with Pickle
- Deployment with Streamlit & Python Scripts

TECH STACK

- Python, Pandas, NumPy, Scikit-learn
- Streamlit, Jupyter Notebook
- SQL (Joins, Queries), Excel
- Git, GitHub, APIs
- Basic: NLP, PCA Pipelines

TOOLS

Streamlit, GitHub, Jupyter Notebook, Power BI (Basic), SQL Server, KaggleHub

PYTHON + ML PROJECTS

Emotion Detection System – Ensemble ML + Deployment

Built an entire ML pipeline using Python: PCA, SMOTEENN, boosted models, voting systems, and real-time webcam app.

Movie Recommendation System – API + ML + Streamlit

Developed a full Python application using CountVectorizer, cosine similarity, and TMDB API integration. Built clean modular scripts for preprocessing, similarity computation, and UI layers.

Book Recommendation System – Collaborative Filtering

Implemented user-item matrix creation, similarity engine, pipeline optimization, and deployed via Streamlit.

Spam Classifier – NLP + TF-IDF

Wrote Python preprocessing pipeline, TF-IDF vectorizer, and integrated a Naive Bayes model inside a Streamlit app.

Zomato Data Cleaning Script (51,717 rows)

Created Python automation scripts for data cleaning, type standardization, null handling, feature extraction, and review parsing.

PCOS/PCOD Analysis – Python Data Engineering

Handled complex categorical transformations, normalization, and preprocessing for healthcare dataset.

EDUCATION

B.Com (Hons.), College of Vocational Studies, Delhi University (2021–2024)

Minor Degree in Management (2021–2023)

Class XII (CBSE – 2021), Class X (CBSE – 2019)

CERTIFICATION

Master in Data Science & Analytics with Artificial Intelligence – IT Vedant (2025–Present)

STRENGTHS

Clean Coding, Debugging, Problem Solving, Fast Learning, Documentation, Analytical Thinking