Alip Kamate

Data Scientist | Machine Learning Engineer

+91 73877 05851 • kamatealip1@gmail.com • linkedin.com/in/kamatealip • github.com/kamatealip • kamatealip.github.io

Jath, Sangli, Maharashtra

SUMMARY

Data Scientist with hands-on experience building end-to-end machine learning systems including recommendation engines and NLP search platforms. Skilled in Python, SQL, and ML libraries such as scikit-learn and PyTorch. Experienced in data processing, feature engineering, model training, and deployment using FastAPI, Docker, and AWS.

PROJECTS

Book Recommendation Engine | https://github.com/kamatealip/shelf-sage

Python • Pandas • NumPy • Scikit-learn • Scrapy • TF-IDF • SvelteKit

- Engineered a personalized recommendation engine using content-based filtering on a dataset of 50,000+ scraped book records
- Reduced data preprocessing time by ~30% through optimized Pandas pipelines and vectorized NumPy operations
- Implemented TF-IDF similarity matching for relevant recommendations with clean SvelteKit browsing interface
- Published full source code with reproducible notebooks and dataset setup instructions on GitHub

Local Document Search Engine | https://github.com/kamatealip/scout

Python • Flask • SQLite • SQL • TF-IDF • NLP

- Built a full-stack document retrieval system capable of indexing and querying 1,000+ local documents with sub-second query latency (~300ms average)
- Designed a **SQLite schema** to store term frequencies and document metadata, enabling efficient SQL-based retrieval and reducing search computation time by ~35%
- Implemented TF-IDF vectorization and cosine similarity ranking, achieving ~90% top-5 retrieval relevance on evaluation queries
- Developed REST APIs using Flask to handle document indexing, search queries, filtering, and pagination for scalable document exploration

Fraud Detection System | ML Classification | https://github.com/kamatealip/transaction-fraud-detection

Python • Pandas • Scikit-learn • Feature Engineering • EDA

- Built a machine learning pipeline to detect fraudulent financial transactions using the **PaySim dataset (6.3M transactions, 10 features)**
- Performed **data cleaning, outlier detection (IQR), correlation analysis, and multicollinearity checks (VIF)** to prepare features for modeling
- Engineered features including **log-scaled transaction amounts and encoded transaction types** to improve model performance
- Addressed severe **class imbalance (~0.13% fraud rate)** using class-weighted models and optimized training strategies
- Trained and compared multiple classifiers including **Logistic Regression, Decision Tree, Random Forest, and XGBoost**
- Evaluated models using **precision, recall, F1-score, ROC-AUC and confusion matrix**, identifying the most effective fraud detection model
- Analyzed **feature importance and transaction behavior patterns** to identify high-risk transaction types and fraud signals

TECHNICAL SKILLS

Languages: Python (primary), SQL, R
ML / Data: Scikit-learn, Pandas, NumPy, PyTorch (learning), TF-IDF, NLP, Machine Learning, Regression, Classification, EDA, Statistics
Backend: FastAPI, Flask, REST APIs, SQLite
Cloud: AWS (EC2, S3), Docker, Linux
Frontend: HTML, CSS, JavaScript, Svelte, SvelteKit
Tools: Git, GitHub, Jupyter, VS Code

MACHINE LEARNING EXPERTIES

Supervised Learning: Linear Regression, Logistic Regression, classification, gradient decent
Recommendation Systems: Content-based filtering
NLP: TF-IDF, vector space models
Model Evaluation: Cross validation, precision, recall, F1-score
Feature Engineering

EDUCATION

Bachelor of Computer Applications (BCA)2021 – Present (Expected 2025)
Shivaji University, Kolhapur, Maharashtra
CGPA: 8.7 / 10 • 82%

HSC – Science (PCMB)2019 – 2020
Raje Ramrao MahaVidyalaya, Maharashtra
CGPA: 8.7 • 80.69%

ACHIEVEMENTS

- 1st Place, Web Development Hackathon – designed and deployed a scalable full-stack web application under a 24-hour time constraint
- Participant, UIDAI Data Hackathon 2026 – built a data-driven solution addressing real-world government identity management challenges

CURRENTLY LEARNING

PyTorch (deep learning) • Advanced AWS services • Kubernetes • LLM APIs & prompt engineering • Kaggle competitions