SUJAY PANDEY

Butwal, Nepal

sujaypandey345@gmail.com | +977 9746384190 | GitHub | LinkedIn | Portfolio

PROFESSIONAL SUMMARY

Final-year BSc. CSIT student with hands-on experience building end-to-end ML systems, including an embedding-based recommendation engine and an NLP sentiment-analysis pipeline, plus Django/MySQL web applications deployed with Docker. Comfortable across the full pipeline: data cleaning, model iteration, and deployment. Seeking an entry-level Machine Learning, Data Science, or Backend Engineering role.

EDUCATION

BSc. CSIT (Expected Graduation: December 2026) — Lumbini City College, Tribhuvan University, Drivertole, Nepal

Relevant coursework: Data Structures & Algorithms, Database Management Systems, Operating Systems, Computer Networks, Object-Oriented Programming

PROJECTS

RAG-Based Resume-to-Job-Description Matcher | *Python, Django, Sentence-Transformers, ChromaDB, Gemini API, Groq API*

- Built a tool that matches resumes against job descriptions and generates an evidence-backed narrative explaining strengths and gaps, so applicants can see exactly where they align with a role and where they don't.
- Built a retrieval pipeline to ground generation in source material: PDF/DOCX parsing, section-aware chunking, sentence-transformer embeddings, and ChromaDB vector storage.
- Designed a two-stage, multi-pass LLM pipeline (fast requirement extraction → evidence mapping → self-check → narrative summary) with Gemini-primary/Groq-fallback generation, served via a Django REST API.
- Added multi-user support with per-session vector-store isolation and a DB-backed daily usage cap that degrades gracefully instead of hard-blocking.

Movie Recommender System | *Python, Pandas, Scikit-learn, Sentence-Transformers*

- Built a content-based recommendation engine spanning over 49,000 movie titles, encoding metadata into sentence-transformer embeddings (all-MiniLM-L6-v2) and ranking via cosine similarity search.
- Filtered a raw TMDB dataset of 1M+ movie entries down to 49,504 high-vote-count titles, removing low-signal noise before similarity scoring.
- Root-caused genre-substring and title-overlap bias skewing results, then resolved it with fractional genre-overlap scoring and a logarithmic vote-count multiplier.

Sirupata Farm House (Academic Group Project) | *Django, MySQL, Docker, HTML/CSS*

- Designed and developed a full-stack Django/MySQL booking platform with a responsive front-end covering listings, availability, and booking management.
- Built a TF-IDF engine that surfaces guest-favorite rooms in the listings section, paired with a TextBlob sentiment-analysis pipeline run on guest reviews.
- Containerized the full Django/MySQL stack with Docker for streamlined deployment.

TECHNICAL SKILLS

Languages: Python, C, C++, Java, JavaScript, SQL, HTML/CSS

Frameworks & Tools: Django, Docker, Git/GitHub, REST APIs, Sentence-Transformers, TF-IDF, TextBlob, .NET

Data & ML: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Exploratory Data Analysis (EDA), Clustering, Classification

CERTIFICATIONS

- Kaggle Certification — Pandas
- CS50P Certificate — Harvard University / edX

LANGUAGES SPOKEN

Nepali, English, Hindi