

Bijaya Kumar Pariyar

AI/ML Engineer | Production RAG & LLM Systems | Applied Machine Learning

Kathmandu, Nepal | +977-9767645335 | bijaybeezoe@gmail.com | github.com/bijay-odyssey | in/bijaya-pariyar | www.kaggle.com/bijaybeezoe

PROFESSIONAL SUMMARY

Data scientist and AI/ML engineer who builds production retrieval and decision systems. Built core components of a document-intelligence and automated-underwriting platform for a US commercial real estate lender - scoring and data-confidence models, the retrieval pipeline, and multi-provider LLM orchestration - architected so deterministic rules make the credit decision and the model only writes the narrative. Author and maintainer of an open-source preprocessing library on PyPI, on a foundation of 20+ applied ML projects across classification, regression, forecasting, clustering and anomaly detection.

EXPERIENCE

Data Scientist Intern

Dec 2025 - Jul 2026

CR Equity AI, Inc. · Remote

Production AI platform for commercial real estate lending, spanning document intelligence, underwriting, retrieval, scoring and LLM orchestration.

- Contributed to a production document-intelligence and AI underwriting platform across 7 repositories. Its core service was ~49k lines of Python at the July 2026 commit audit; I authored 150 of that service's commits, including 71 of its last 100 production-branch commits. The architecture evolved through five successive rebuilds, from initial prototype to modular production services.
- Built the deal-scoring model: a multi-factor weighted scorer producing an independent approve/conditional/decline call, with per-factor data provenance and confidence scoring, resolving each metric through verified, estimated, benchmark-derived and not-computable tiers that each carry their own weight downstream.
- Designed the decision architecture so a deterministic rule engine computes the recommendation and the LLM receives it as read-only context for narrative only - a hallucinating or prompt-injected model can change the wording but never flip an approval, which is a hard requirement in regulated lending.
- Implemented the retrieval pipeline end to end across 20+ financial document types: OCR routing for native-text vs scanned input, schema-driven chunking, embeddings with pgvector HNSW search, cross-encoder reranking, and content-hash deduplication.
- Built multi-provider LLM orchestration (AWS Bedrock, Google Gemini, OpenAI) with per-task fallback chains, a throttle-aware circuit breaker and graceful degradation; deployed as separate API and async-worker services on ECS Fargate with Redis-backed job state.

Skills: Python, pandas, NumPy, scikit-learn, FastAPI, PostgreSQL, pgvector, Redis, ARQ, AWS Bedrock, AWS Textract, AWS ECS Fargate, Docker, GitHub Actions, OpenAI API, Google Gemini

SKILLS

Machine Learning: Classification, Regression, Forecasting, Clustering & Anomaly Detection, scikit-learn, XGBoost, LightGBM, Prophet, SHAP & LIME, Optuna & Cross-Validation, Imbalanced-Data Methods

Data: pandas, NumPy, SciPy, SQL, Feature Engineering, ColumnTransformer Pipelines, Target & Categorical Encoding, Data Quality & Leakage Auditing

AI / LLM / RAG: Production RAG Pipelines, Multi-Provider LLM Orchestration, Vector Search (pgvector, FAISS, Qdrant), Cross-Encoder Reranking, Structure-Aware Chunking, Embeddings, OCR

Backend & Cloud: FastAPI, Flask, Pydantic, asyncio, ARQ, PostgreSQL, Redis, AWS (Bedrock, Textract, S3, ECS Fargate), Docker, GitHub Actions CI/CD

Engineering Practice: Evaluation Design, Reproducible Pipelines, Open-Source Maintenance, Audit Trails & PII Redaction

Languages: Python, SQL, JavaScript, Java, C#, PHP

PROJECTS

edaprep - open-source ML preprocessing library github.com/bijay-odyssey/edaprep

2026

Author & Maintainer

A published Python library for leakage-safe EDA and ML preprocessing with an explainable planner: it inspects a dataset, works out which preprocessing operations actually apply, states why in plain language, then fits every statistic on the training split alone. Actively maintained on PyPI under semantic versioning with a migration-noted changelog; fully typed, Python 3.9-3.13, with CI and benchmarks. Adopted beyond the author: 15 stars, 23 forks, and merged contributions from 10 outside contributors across 15 pull requests.

- Every decision is auditable: `explain()` reports the statistic that drove each choice - median imputation because skew is 3.22, target encoding with 5-fold cross-fitting because 163 levels exceeds the one-hot ceiling, an identifier column dropped because 100% of its values are distinct.
- Found, via a codebase audit, a target-leakage bug present in every release since 0.1.0 - the out-of-fold target encoder mixed in a smoothing prior computed over the whole training target rather than per fold, so a near-singleton category's encoding was measurably a function of its own label. Scoped the fix and root cause for a contributor, reviewed the implementation, and verified independently (flipping one row's label no longer changes its own encoding) before merging.

Stack: Python, scikit-learn, pandas, NumPy, SciPy, pytest, GitHub Actions

Qdrant - upstream open-source contribution

2026

Contributor

Fixed undefined behaviour in Qdrant's official Python client: formula evaluation accepted zero raised to a negative exponent instead of rejecting it. Reviewed and merged by a Qdrant maintainer (qdrant-client #1429).

Stack: Python, Qdrant

Applied ML Portfolio - 20+ end-to-end projects

2025-2026

Author

20+ end-to-end projects across classification, regression, forecasting, clustering and anomaly detection, each published with its evaluation design and a written analysis. 58 Kaggle notebook votes.

- Tuned XGBoost fraud detector: ROC-AUC 0.95 on ~1M heavily imbalanced transactions, with SHAP attribution behind every prediction.
- SKU-level demand forecasting over 1.7M+ rows (LightGBM + Prophet, TimeSeriesSplit CV, log-transformed target), reported on RMSE/MAPE/SMAPE rather than one flattering metric.

Stack: scikit-learn, XGBoost, LightGBM, Prophet, SHAP, Optuna

SHAP-Based Feature Selection for Churn Prediction

2025-2026

Author (research preprint)

Preprint on Preprints.org applying SHAP-based feature selection with Optuna hyperparameter tuning to telecom customer churn (7,043 records, 7 candidate models), supervised by Marshal Basnet.

- Audited the published pipeline and found its headline result was leakage: one field is present only for churned customers, another is a vendor-computed churn propensity. Re-ran it with both removed - test ROC-AUC 0.997 to 0.854, F1 0.980 to 0.604 - and corrected the repository record.

Stack: Python, scikit-learn, SHAP, Optuna

Smart Kar - automated tax document analysis

2026

Sole developer, 2-person college team

Final-year college project: an end-to-end applied AI system for Nepal's SMEs. It ingests tax documents, extracts and classifies them, flags anomalies, and produces IRD-ready VAT and TDS reports. Sole developer of the system.

- Bilingual document extraction: PyMuPDF for digital PDFs with a Tesseract OCR fallback for scanned input in English and Nepali, table reconstruction from OCR word geometry, and fuzzy vendor matching.
- LLM-assisted classification, anomaly detection and tax-rule extraction with a passthrough fallback when no provider key is configured, so the pipeline degrades to deterministic behaviour; Arq on Redis for async processing, ReportLab and XML report generation, and HMAC-signed eSewa payments.

Stack: FastAPI, PostgreSQL, pgvector, Arq, Groq, Gemini, Tesseract, ReportLab, React

EDUCATION

Bachelor in Computer Application (BCA)

Expected 2026

Divya Gyan College · Kathmandu, Nepal