

PRASHANT DARSHANWAR

+91 7083553011 | prashantdarshanwar70715@gmail.com | Nagpur, Maharashtra | linkedin.com/in/prashant-darshanwar-0310b222a | github.com/prashantdarshanwar

PROFESSIONAL SUMMARY

Data Scientist with 7 months of hands-on industry experience building end-to-end machine learning solutions using Python, SQL, Scikit-learn, XGBoost, FastAPI, and Power BI. Skilled in EDA, statistical analysis, hypothesis testing, feature engineering, cross-validation, hyperparameter optimization, and model evaluation across regression, classification, and clustering problems. Experienced in developing Generative AI and Retrieval-Augmented Generation (RAG) applications using LangChain, FAISS, and HuggingFace, alongside MLOps workflows with Docker, MLflow, and CI/CD. Adept at translating ambiguous business problems into scalable, data-driven, AI-powered solutions.

TECHNICAL SKILLS

Programming: Python, SQL

Machine Learning: Logistic Regression, Random Forest, XGBoost, Supervised & Unsupervised Learning, Regression, Classification, Clustering, Cross-Validation, Hyperparameter Optimization, Model Evaluation

Statistics: Probability, Statistical Inference, Hypothesis Testing, A/B Testing, Confidence Intervals, Sampling Techniques, Bias-Variance Trade-off

Data Science: Exploratory Data Analysis (EDA), Feature Engineering, Feature Selection, Data Cleaning, Missing Value & Outlier Treatment, Class Imbalance Handling

Generative AI: LLMs, RAG, Prompt Engineering, LangChain, FAISS,

Libraries: Pandas, NumPy, Scikit-learn, TensorFlow, Matplotlib, Seaborn,

Visualization : Power BI

Databases: MySQL, PostgreSQL

Cloud & MLOps: AWS S3 (basic), MLflow, Docker, GitHub Actions, CI/CD, Render

Developer Tools: FastAPI, REST APIs, Git, GitHub, Jupyter Notebook, VS Code, Google Colab

PROFESSIONAL EXPERIENCE

Data Science Intern — Hisan Labs Private Limited, Nagpur Nov 2025 – Jun 2026

- Built an end-to-end customer churn prediction model on 200K+ records using Logistic Regression, Random Forest, and XGBoost, optimizing performance with hyperparameter tuning and k-fold cross-validation to reach 88% accuracy and 0.91 ROC-AUC, then deployed it to the company's internal dashboard.
- Performed EDA, missing-value treatment, outlier detection, and feature engineering/selection, then applied statistical analysis and hypothesis testing to validate customer behavior patterns before presenting insights to stakeholders.
- Designed and automated ETL pipelines in Python, SQL, and Pandas, feeding MySQL-based reporting layers used daily by the analytics team.
- Built Power BI dashboards tracking weekly KPIs, revenue trends, and regional performance for the sales team, cutting manual reporting time by approximately 40%.

PROJECTS

Sanskrit RAG System — CPU-Optimized Hybrid RAG for Sanskrit Document QA

- Built a CPU-optimized hybrid Retrieval-Augmented Generation framework for Sanskrit document understanding and context-aware question answering, combining dense embedding search with sparse (keyword-based) retrieval.
- Implemented Sanskrit-specific text preprocessing, chunking, and embedding generation using lightweight HuggingFace models optimized for CPU inference, integrated with FAISS for vector search.
- Combined semantic and lexical retrieval signals to handle Sanskrit's complex morphology, improving answer relevance for context-aware question answering, and exposed the system via FastAPI for query-time inference.
- Impact:- Improved retrieval relevance by ~25% over dense-only retrieval and reduced response latency by ~40%, enabling deployment on low-resource machines without requiring GPUs.

Technologies: Python, LangChain, FAISS, HuggingFace, Hybrid Retrieval (Dense + Sparse), FastAPI, NLP, Prompt Engineering

Online Shopper Purchase Intention Prediction — End-to-End ML Deployment

- Built and deployed an end-to-end ML pipeline predicting purchase intent from browsing data, comparing Random Forest and XGBoost with hyperparameter optimization and cross-validation, tracking experiments with MLflow, and deploying the containerized FastAPI model on Render with CI/CD via GitHub Actions.
- Impact: Achieved 91% ROC-AUC and improved prediction accuracy by 8% over the baseline model while reducing deployment time from hours to under 15 minutes through automated CI/CD.

Technologies: Python, FastAPI, Docker, MLflow, GitHub Actions, CI/CD, Render, Scikit-learn, XGBoost, EDA, MLOps

AI Invoice Processing System — Automated Invoice Data Extraction & Validation

- Built an AI-powered invoice processing system that automates extraction of structured data (vendor, amount, line items, dates) from scanned and PDF invoices using OCR and NLP-based information extraction.
- Implemented a rules-based validation layer to flag anomalies, missing fields, and duplicate entries, and exposed the pipeline via FastAPI endpoints for integration with downstream accounting workflows.
- Impact: Automated 90%+ of manual invoice data entry, reduced processing time from 10 minutes to under 2 minutes per invoice, and minimized human validation effort.

Technologies: Python, OCR, NLP, FastAPI, Pandas, Regex-based Validation, REST API

EDUCATION

B.Tech, Artificial Intelligence and Data Science — Wainganga College of Engineering, Nagpur 2026

HSC — Science (PCM), Maharashtra State Board 2022

CERTIFICATIONS

Data Science Professional Certificate — Cloudblitz (2026)