# Vijay Takbhate

Email — GitHub — LinkedIn — Kaggle — +91-8767363681

## Professional Summary

Risk Analyst with hands-on experience in applied machine learning, MLOps, and GenAI systems. Skilled in building, evaluating, and deploying end-to-end ML solutions using Python, MLflow, Docker, and cloud platforms, with strong exposure to risk and decision-focused use cases.

## Technical Skills

**Programming & Backend:** Python, R, Flask, SQL (MySQL)

**Machine Learning & Data Science:** Regression, Classification, Feature Engineering, Model Evaluation (MAE, RMSE, $R^2$), Cross-Validation

**Generative AI & LLMs:** LangChain, LLM Orchestration, Retrieval-Augmented Generation (RAG), Prompt Engineering, GPT Models (GPT-3.5, GPT-4)

**Embeddings & Vector Databases:** all-MiniLM-L6-v2, ChromaDB, Semantic Search, Chunking Strategies

**MLOps & DevOps:** MLflow, Docker, GitHub Actions (CI/CD), Kubeflow

**Data & Analytics:** Databricks, Metabase

**Cloud & Deployment:** AWS EC2, AWS Lambda, Model Deployment

## Experience

### InCred Financial Services
Dec 2024 – Present

*Risk Analyst (ML Systems, BRE Development & Experimentation)* — *Mumbai, Maharashtra*

- Develop and maintain credit risk decision policies inside the **Experian PowerCurve Strategy Management (BRE)** production system for large-scale automated decision workflows.
- Design and execute **A/B tests** to evaluate policy changes and measure business impact before production rollout.
- Built an internal **Python-based policy simulator** to validate implementations by comparing historical vs updated policy outcomes using aggregations and flow-level counts, enabling precise impact analysis.
- The simulator highlights affected customer segments and summarizes behavioral shifts, helping business stakeholders understand expected decision impact prior to deployment.
- Developed a centralized validation framework allowing any developer to test policy updates reproducibly, reducing verification effort by **30%**.
- Implemented **CI/CD pipelines (GitHub Actions)** for packaging, testing, and versioning the simulator repository to ensure reliable and standardized validation workflows.
- Analyze outcome distributions using **Databricks and SQL** to monitor post-deployment behavior and support data-driven policy iteration.

### Fox Solutions Pvt. Ltd.
Feb 2024 – Oct 2024

*Automation Engineer (Intern → Full-time)* — *Pune & Nashik, Maharashtra*

- Developed and deployed automated pipelines with a focus on **reliability, monitoring, and reproducibility**, reducing manual intervention in production systems.
- Applied best practices in **version control, system monitoring, and fault handling** to build scalable and maintainable automation workflows.

## PROJECTS

### AI-Powered Portfolio & Resume Assistant for HR (RAG-based) *Oct 2025 - Present*

- *End-to-end GenAI system enabling semantic search and interactive HR queries over resume and project data.*
- Built an **AI-enabled portfolio platform** allowing recruiters to conduct **virtual HR interviews** via an interactive chatbot powered by **LangChain** and **GPT-5-mini**.
- Enhanced the chatbot backend using **Retrieval-Augmented Generation (RAG)** to answer recruiter queries grounded in **resume content and GitHub project documentation**.
- Converted all GitHub project **README files into PDFs**, generated embeddings using **all-MiniLM-L6-v2**, and stored them in **ChromaDB** for semantic retrieval.
- Implemented a retrieval pipeline that fetches the **top-5 relevant context chunks**, enabling the LLM to generate **accurate, hallucination-reduced responses** about skills and projects.
- Deployed the application using **Flask, Docker, and AWS EC2**, integrated **MLflow** for experiment tracking, and automated workflows via **GitHub Actions (CI/CD)**.
- **Live Demo:** Portfolio Demo — **GitHub:** Repository

### Medical Insurance Cost Prediction (SVR Model) *Sep 2025*

- *Supervised machine learning project focused on predicting insurance costs to support pricing and risk assessment decisions.*
- Designed an end-to-end **regression pipeline** to predict medical insurance costs, achieving **$R^2$: 0.86** and **MAE: 0.034**, demonstrating applicability for premium estimation and risk assessment.
- Performed **data preprocessing** including categorical encoding, feature scaling, and outlier analysis to improve model stability and generalization.
- Evaluated baseline models (Linear Regression, tree-based regressors) and selected **SVR** to capture non-linear relationships among cost-driving features.
- Applied **cross-validation and hyperparameter tuning** to reduce overfitting and ensure robust performance across unseen data.
- Deployed the model via a **Flask** API, containerized using **Docker**, and hosted on **AWS EC2** with **MLflow** for experiment tracking and **CI/CD** automation; integrated **Kubeflow** for scalable ML workflows.
- **GitHub:** Repository

## EDUCATION

### SVERI's College of Engineering Pandharpur, Maharashtra

*B.Tech. & Diploma — Electronics and Telecommunication* *May 2024 & May 2021*

## LANGUAGES

**English** **Marathi** **Hindi**