CHIRAG PATIL

AI/ML Engineer, Data Scientist

Email: chirag251120@gmail.com
Phone: +917276307991

Results-driven AI/ML Engineer with 4+ years of hands-on experience in **Machine Learning**, **Deep Learning**, **NLP**, and **Generative AI**. Skilled in developing **end-to-end AI solutions** using **Python**, **SQL**, and advanced **neural network architectures**. Proficient in leveraging **Large Language Models (LLMs)**, **Retrieval-Augmented Generation (RAG)**, and **LangChain** for intelligent document understanding and **predictive analytics**. Knowledgeable in **cloud-based deployment (AWS, FastAPI)** and **MLOps practices** such as **model monitoring** and **CI/CD pipelines**. Passionate about building **scalable AI systems** that **automate workflows** and **drive measurable business impact**.

Professional Summary

- **AI/ML Engineer** with 4 years of experience in **Machine Learning**, **Natural Language Processing (NLP)**, and **Generative AI**, delivering **end-to-end data-driven solutions** across **Logistics**, **E-commerce**, and **Insurance** domains. Skilled in building **scalable predictive systems** and **LLM-powered applications**.
- Built AI systems such as a **Legal & Compliance Assistant Chatbot** using **Retrieval-Augmented Generation (RAG)**, integrating **LangChain**, **OpenAI GPT**, **Pinecone**, and **Tesseract OCR** for **legal document retrieval and summarization**, achieving a **30% reduction in review time**.
- Designed and deployed **predictive models** including **cart abandonment** and **customer churn prediction**, leveraging **XGBoost**, **Random Forest**, **scikit-learn**, and **SHAP** for **model interpretability** and **actionable insights**.
- Proficient in **Python**, **SQL**, and key **Machine Learning frameworks** such as **Pandas**, **NumPy**, **TensorFlow**, and **PyTorch**. Experienced in **feature engineering**, **hyperparameter tuning**, and **model optimization** for **production-grade pipelines**.
- Knowledgeable in **MLOps** and **cloud deployment** using **AWS (EC2, Lambda, S3)** and **Docker**, with knowledge of **MLflow**, **CI/CD pipelines**, for **reproducible model training and deployment**.
- Integrated **RESTAPIs (FastAPI)** and **dashboards** for **real-time monitoring**, **model performance tracking**, and **decision automation**.
- Adept at **cross-functional collaboration**, and **business translation** of AI solutions to improve **operational efficiency**, **customer retention**, and **regulatory compliance accuracy**.

Technical Skills

- **Programming & Data Handling:** Python, SQL, Pandas, NumPy, Data Cleaning, Feature Engineering
- **Machine Learning:** Scikit-learn, Supervised & Unsupervised Learning, Ensemble Methods, Bagging & Boosting, Clustering, Hyperparameter Tuning
- **Deep Learning:** TensorFlow, Keras, PyTorch, Neural Networks, Transformers, Transfer Learning, Model Evaluation
- **NLP & Generative AI:** Hugging Face, LangChain, LangGraph, OpenAI API, LangSmith, Prompt Engineering, Retrieval-Augmented Generation (RAG), Word2Vec, TF-IDF, NLTK, spaCy, AI Agents, Large Language Models (LLMs)
- **Vector Databases:** Pinecone, FAISS
- **Statistical Analysis:** Hypothesis Testing, A/B Testing
- **Data Visualization:** Matplotlib, Seaborn, Plotly
- **Web Frameworks & APIs:** FastAPI
- **MLOps & Cloud Platforms:** AWS (EC2, Lambda, S3), Docker, MLflow, CI/CD Pipelines, Version Control (Git, DVC)
- **Tools & Collaboration:** Jupyter Notebook, VS Code, JIRA, Agile/Scrum Practices
- **Soft Skills:** Communication, Collaboration, Problem-Solving, Critical Thinking, Leadership & Time Management

Work Experience

- **LANCE SOFT ENGINEERING PRIVATE LIMITED, Hyderabad, India**

  AI/ML Engineer (June 2021 - Present)

Education

Bachelor of Engineering (June 2021) - Ramrao Adik Institute of Technology (Mumbai University)

Projects

Legal & Compliance Assistant Chatbot

- Designed and implemented a **Legal & Compliance Assistant Chatbot** using **Retrieval-Augmented Generation (RAG)** to **automate document retrieval** and **summarization** for legal and compliance teams.
- Successfully **deployed** the system in production on **AWS**, enabling real-time access to customs, shipment, and **contractual regulations through a secure**, conversational interface.
- Integrated **LangChain**, **OpenAI GPT-4**, and **Pinecone** to deliver context-aware **semantic retrieval**, allowing users to query large document repositories conversationally with cited, explainable responses.
- Built **automated document ingestion pipelines** using **Python, Tesseract OCR, and PyMuPDF** to process **structured** and **unstructured** data (PDF, Word, Excel, scanned files), improving document **accessibility and searchability by 40%.**
- **Containerized and deployed** the application using **FastAPI, Docker,** and **AWS Lambda**, with **S3-based storage** for efficient **retrieval** and **horizontal scalability**.
- Achieved a **40% reduction** in compliance review turnaround time and a **30% improvement in retrieval accuracy**, directly improving decision speed and regulatory adherence.
- Implemented **performance monitoring** and **logging** for tracking **API usage**, retrieval latency, and user satisfaction, ensuring high availability and reliable performance in production.

  Tech Stack: Python, Tesseract OCR, LangChain, OpenAI GPT-4, CLIP, Pinecone, PyMuPDF, SpaCy, FastAPI, Docker, AWS (Lambda, S3), Streamlit

Cart Abandonment Prediction & Retargeting System

- Developed and deployed a **Cart Abandonment Prediction & Retargeting System** to identify users likely to abandon carts using **e-commerce clickstream and session data**, enabling targeted remarketing campaigns.
- Built and optimized predictive models using **XGBoost** and **Random Forest**, with engineered behavioral features such as **browsing depth**, **interaction frequency**, and **price sensitivity** to improve prediction accuracy.
- Addressed **class imbalance** through **stratified sampling** and evaluated models using **Precision**, **Recall**, **F1-Score**, and **ROC-AUC**, achieving **high recall for at-risk customers**.
- Implemented **model explainability** using **SHAP** to identify key behavioral drivers, ensuring transparency and actionable insights for marketing teams.
- **Containerized and deployed** the model as a **RESTAPI** and **Docker**, integrated with the company's marketing pipeline on **AWS EC2** for real-time cart recovery predictions.
- Achieved a **9% increase in cart recovery conversions** and **improved targeted campaign efficiency by 12%, enhancing customer retention and revenue growth.**
- Implemented **MLflow-based model tracking** and **performance monitoring dashboards** to ensure continuous model reliability and retraining efficiency in production.

  Tech Stack: Python, Pandas, NumPy, scikit-learn, XGBoost, SHAP, Flask, Docker, AWS EC2, MLflow

Customer Churn Prediction Model for Insurance

- Built and deployed a **Customer Churn Prediction Model** to identify high-risk policyholders using **insurance policy, claims, and demographic data**.
- Built and optimized a **XGBoost Regression model** in scikit-learn to accurately identify customers at risk of policy non-renewal.
- Engineered features using **SQL** and **Pandas**, and trained an **XGBoost model** in **scikit-learn** to predict churn probability with high recall.
- Improved **customer retention strategy**, achieving a **15% reduction in churn** through actionable insights and targeted interventions.