

PRANITA LOKHANDE

DATA SCIENTIST

📍 Pune | +91- 7020554995 | pranitalokhande2607@gmail.com

PROFESSIONAL SUMMARY

Experienced Data Scientist with over 4 years of hands-on experience in developing and deploying Machine learning, Deep learning models across various domains. Strong background in Natural Language Processing (NLP), including tasks like text classification, entity recognition, and text analysis. Currently working on building Generative AI solutions, such as RAG (Retrieval-Augmented Generation) systems, LLM-powered agents and integrating them into real-world business applications.

WORK EXPERIENCE

Symplocos Solution, Bengaluru

July 2021-Present

- Designed a Gen-AI solution using GPT-4, LLaMA-3, Pinecone and LangChain to classify return comments and extract product-wise insights, helping reduce return rates and improve product quality.
- Built forecasting models using XG-Boost and Random Forest on time-series features to predict warehouse, fleet, and manpower demand, enabling efficient logistics planning and reducing delays.
- Developed ML and DL models using customer policy and claims data to predict churn risk. Generated churn segments (High/Medium/Low) to drive targeted retention strategies and improve renewals.

TECHNICAL SKILLS

- Languages & Frameworks:** Python, FastAPI
- Databases:** SQL
- Libraries:** NumPy, Pandas, TensorFlow, Keras, NLTK, SpaCy, LlamalIndex
- Deep Learning:** ANN, RNN, LSTM, Encoder-Decoder, Bert
- Natural Language Processing:** Tokenization, Embeddings, Stemming & Lemmatization, Language Modeling, Text Classification
- Machine Learning:** Feature Engineering, Exploratory Data Analysis (EDA), Linear Regression, Logistic Regression, SVM, Decision Trees, Random Forest, K-Nearest Neighbors, Clustering, PCA
- Data Visualization:** Matplotlib, Seaborn
- Cloud Servers:** AWS
- Version Control System:** Git, GitHub
- Generative AI :** Hugging face Transformer, LangChain, OpenAI API, Prompt Engineering, RAG

EDUCATION

Bachelor of Computer Science

Dr. Babasaheb Ambedkar Marathwada University, Chh. Sambhaji Nagar.

PROJECTS

1) Client : Papier Nov 2024 - July 2025

- Used GenAI models like GPT-4 and LLaMA 3 with few-shot prompting to automatically classify customer return comments.
- Applied sentence embeddings and FAISS to group similar return comments and discover hidden patterns using semantic clustering.
- Created prompt templates in LangChain to extract structured outputs from LLMs for dashboards and reports.
- Integrated Retrieval-Augmented Generation (RAG) to generate product-wise summaries backed by real customer return data, enabling fact-based and explainable insights.
- Generated product-wise summaries of return reasons to help teams improve product quality, listings, and reduce returns.
- Compared traditional ML (TF-IDF + SVM) with LLM-based methods and chose GenAI for better accuracy and flexibility.
- Delivered insights through Python pipelines and structured reports that support business teams in decision-making.

2) Client : Marshmallow April 2023 - Aug 2024

- Collected and processed large-scale logistics datasets using Pandas, extracting key features such as order volumes, delivery zones, SKUs, and promotional events for forecasting.
- Engineered time-based features (e.g., day-of-week, sales events), lag variables, and rolling statistics to capture seasonality and trends in warehouse/fleet demand.
- Built and evaluated machine learning models such as XGBoost and Random Forest Regressors to forecast short-term and long-term logistics capacity.
- Tuned forecasting models to predict regional spikes or drops in demand with high accuracy, supporting proactive logistics planning and resource allocation.
- Developed reusable training and evaluation pipelines for demand forecasting using Scikit-learn, ensuring scalable deployment and repeatability across zones.
- Enabled actionable planning through structured outputs that feed into dashboards for weekly/monthly capacity forecasts, helping reduce delays, overstocking, and last-minute escalations.

3) Client : Schenker Sep 2021 – Feb 2023

- Collected and cleaned customer, policy, and claims data using Python (Pandas) to understand which customers might stop renewing their policy.
- Created useful features like how often claims were rejected, how recently the customer interacted, and if their premium changed over time.
- Built machine learning models like Logistic Regression, XGBoost, and Random Forest to predict if a customer is likely to leave.
- Found the most important reasons behind customer churn, so the business could take action based on clear insights.
- Assigned churn risk levels (High / Medium / Low) to each customer and shared results in dashboards, helping teams focus on retaining at-risk customers.