

Namrata Paikekar

Data Scientist

+91-7499980416
[linkedin.com/in/namrata-paikekar-66046b268](https://www.linkedin.com/in/namrata-paikekar-66046b268)

namratapaikekar@gmail.com
Pune, India

SUMMARY

I am an experienced Data Scientist specializing in data mining, statistical modeling, machine learning, NLP, deep learning, time series forecasting and generative AI. Proficient in Python, SQL and visualization libraries like matplotlib, with a strong track record of delivering insights that drive business impact.

SKILLS

Python | MySQL | ML Algorithms | Artificial Intelligence | AWS | Clustering | Data Mining | Data Science | NLTK | Machine Learning | Deep Learning | Linear/Logistic Regression | Random Forest | SVM | Decision Trees | NLP | Spacy | Naive Bayes | EDA | Data Visualization | Flask | Keras | K-means | Matplotlib | MSSQL | Numpy | Pandas | Plotly | Pytorch | VSCode | Scikit-learn | Scipy | Gen AI | LLMs | Langchain | Hugging Face | transformer models | AWS | Git

EXPERIENCE

MD India Health Insurance TPA Pvt. Ltd.

Oct 2024 – Present

Pune, India

Data Scientist

- Understand problem statements, collect relevant data, perform exploratory data analysis (EDA), develop predictive models, fine-tune algorithms, evaluate model performance, and deploy solutions.
- Collaborate closely with the development team for seamless API integration of machine learning models into production systems.
- Implement classical machine learning and deep learning algorithms to address complex business challenges effectively.
- Currently working on cutting-edge technology Generative AI projects, focusing on Retrieval-Augmented Generation (RAG) systems.

Neosoft Systems and Cloud Services

Apr 2021 – Oct 2023

Pune, India

Junior Data Scientist

- Designed and implemented a supervised binary classification model using Python, SQL, and machine learning (Logistic Regression, SVM, Random Forest and XGBoost) to predict potential loan customers.
- Conducted extensive data cleaning, preprocessing, and exploratory data analysis (EDA), ensuring data quality and enhancing model performance.
- Developed a deep learning model utilizing Python, CNN, and Computer Vision to detect fraudulent signatures in legal documents.
- Employed Siamese network architecture to analyze semantic similarity between signatures, enabling accurate fraud detection with limited data availability.

PROJECTS

Tariff Digitization

Technologies: Python, OCR (Unstructured, PyTesseract, PaddleOCR, Tabula), Django REST API, Next.js, MSSQL

Project Description:

Developed a tariff digitization solution to extract and digitize the scanned SOC documents and store extracted data into structured and standardized format into MSSQL database for easy access and analysis. Leveraged different OCR libraries and text-processing techniques to convert non-searchable PDFs into structured SQL tables. Standardized hospital room categories by mapping them to unified formats, enabling accurate and efficient processing.

Intelligent RAG System for ICD-PCS Medical Coding

Technologies: Python, MSSQL, Pandas, FAISS, Sentence Transformers, BioGPT (Hugging Face Transformers), OpenAI, Pickle, Numpy

Project Description:

Developed an AI-powered Retrieval-Augmented Generation (RAG) pipeline for healthcare domain data. The project integrated ICD-PCS codes and procedure descriptions stored in MSSQL with semantic search and generative AI. This solution enabled faster and more accurate medical coding assistance, treatment understanding, and decision support for healthcare professionals.

Fraud detection for health insurance claims

Technologies – Python, SQL, Scikit-learn, XGBoost, Random Forest, Logistic Regression, KMeans Clustering, Pandas, NumPy, Matplotlib, Seaborn

Project Description:

Developed a hybrid fraud detection framework for insurance claims by combining unsupervised clustering, rule-based anomaly triggers, and supervised machine learning. Claims were first grouped into clusters to identify behavioral patterns, followed by anomaly detection using medical and operational triggers. Filtered high-risk claims were then passed into supervised ML models to minimize false negatives and ensure higher fraud capture rates. The system significantly improved fraud detection accuracy while maintaining business interpretability.

Key Contributions:

- Built end-to-end AI/ML solutions spanning OCR text extraction, fraud detection, semantic search, and RAG-based summarization in healthcare and insurance domains.
- Designed scalable data pipelines with SQL and Python for ingestion, validation, and preprocessing, ensuring high data quality and consistency.
- Developed and optimized supervised, unsupervised, and deep learning models (Random Forest, XGBoost, Clustering, SentenceTransformers, BioGPT, OpenAI API) to solve business-critical problems.
- Enhanced fraud detection accuracy by integrating clustering, anomaly triggers, and supervised models with a focus on minimizing false negatives.
- Delivered interpretable, production-ready solutions through Explainability techniques, APIs, and interactive dashboards for business stakeholders.

EDUCATION

Bachelor of Engineering

Ahilyadevi Holkar Solapur University

2013 - 2017

Pune, India

LANGUAGES

English
Advanced

Hindi
Advanced