Vinayak Pawar

+91 9960990969   Pune   vnpawar94@gmail.com

Portfolio: https://vinayakpawar94.github.io/
GitHub: https://github.com/vinayakpawar94
LinkedIn: https://linkedin.com/in/vinayak-pawar94

SUMMARY

Highly motivated Data Scientist with hands-on experience in extracting insights from complex datasets, building robust ML models, and deploying scalable data-driven solutions. Proven ability to translate business challenges into analytical problems, collaborate cross-functionally, and communicate insights effectively. Proficient in predictive modeling, NLP, data visualization, and workflow automation.

EDUCATION

Data Science & AI Certification
INTTRVU.AI   2025

Post Graduate: Master's in Statistics
Dr. BAMU, Aurangabad   2019

Graduation: Bachelor's in Science
SRTMU, Nanded   2016

PROJECTS

Financial Fraud Detection 🔗
- Built a pipeline to detect fraud using XGBoost on 6M+ transactions.
- Engineered features like anomaly score using Isolation Forest and a tuned threshold to optimize recall.
- Automated batch prediction and model evaluation with ROC, F1, and confusion matrix visualizations.

Insurance Premium Prediction 🔗
- Predicting Health Insurance Premiums for Data.
- The dataset consisted of 1 million records containing information such as Age, Gender, BMI, Smoking Status, Region, Occupation, Medical History, and other relevant details.
- Built a Regression Model using XGBoost along with feature importance for presentations.
- Created a Batch Job for predicting live data

Chatbot - Generative AI 🔗
- It answers questions related to financial & operational risk guidelines provided by the RBI.
- Used RAG as it provides relevant information in the form of a prompt to LLM.
- Used Langchain as it provides tools & abstraction to improve customization, accuracy & relevancy.
- Created an API & a Front-End Chatbot that would be interactive. We can ask questions to it and it would provide prompt responses.

PROFESSIONAL EXPERIENCE

Data Analyst
iResearch Services Pvt. Ltd. | Aug 2023 – Present
- Developed machine learning models (classification, clustering) on survey and behavioral datasets.
- Applied NLP to derive sentiment insights from unstructured text fields.
- Built automated ETL pipelines for data cleaning, preprocessing, and transformation.
- Visualized insights through dashboards and collaborated with stakeholders to deploy solutions.
- Continuously improved model performance and integration with business workflows.

Data Quality Analyst
District Government | Jun 2019 – Jul 2023
- Managed and validated data collection for 300,000+ farmers under the PM-KISAN scheme.
- Preprocessed large-scale datasets using Excel, ensuring high data accuracy and completeness.
- Collaborated with officials to resolve anomalies and improve decision-making with detailed reports.

SKILLS

- Programming Languages - Python, SQL
- Machine Learning Libraries - Scikit-learn, XGBoost, Keras, Langchain, Shap, Lime, Gensim, NLTK
- Deep Learning & AI Expertise - CNN, LSTM, Word2Vec, NLP, Neural Networks
- Generative AI - Langchain, RAG, BERT, LLMs, Prompt Engineering
- Data Visualization - Tableau, Power BI, Matplotlib, Seaborn, Plotly
- Other Tools - AWS (EC2), GitHub, Flask, Docker, SQLite

CERTIFICATIONS & AWARDS

- Certifications: Python Certification, Data Science Professional Certificate (Intervu)
- Awards: Employee of the Quarter (Jun 2024 – Aug 2024), Covid Yodha Award (District Magistrate)