CHETAN PAWAR

Nashik, Maharashtra • chetanpawar2002official@gmail.com • 9359208098 • GitHub • LinkedIn • Portfolio

TECHNICAL SKILLS

Programming Languages: Python, SQL
ML & Data Science Libraries: Pandas, NumPy, Scikit-learn, XGBoost, TensorFlow, Keras, Hugging Face Transformers, LangChain
Data Visualization & BI: Matplotlib, Seaborn, Plotly, Tableau, Power BI, Excel
Tools & Platforms: Git, GitHub, Docker, Jupyter Notebook, Streamlit, Flask, VS Code
Cloud & Databases: Firebase, SQLite, MySQL
Concepts & Methodologies: Machine Learning, Deep Learning, Natural Language Processing (NLP), Generative AI, Data Wrangling, Feature Engineering, MLOps, Statistical Analysis, A/B Testing

PROJECTS

RepoBot AI GitHub | Live Demo

- **Engineered** a Generative AI-powered web application utilizing **Flask** and **Groq AI** to analyze and interpret over **500 GitHub code files**, significantly enhancing code comprehension capabilities.
- **Implemented** a vector-based search system with **ChromaDB** and **HuggingFace embeddings**, achieving accurate semantic search across **400+ text chunks** for superior code insight retrieval.
- **Automated** the end-to-end pipeline for GitHub repository ingestion and vector storage, improving data processing efficiency by **30%** and enabling scalable code analysis.
- **Integrated** a memory-augmented conversational AI agent to handle complex code-related queries, providing dynamic and context-aware responses to users.

End To End Food Delivery Time Prediction GitHub | Live Demo

- **Developed** and **deployed** an **XGBoost regression model** to predict food delivery ETAs, aimed at optimizing logistics and reducing average delivery time by **15%**.
- **Executed** comprehensive **data cleaning** and **feature engineering**; addressed **outliers**, **missing values**, and **multicollinearity** to improve dataset quality.
- **Evaluated** and tuned multiple algorithms (**Linear Regression**, **Decision Trees**, **Random Forest**); optimized **XGBoost** to achieve an **R-squared of 0.82** and **RMSE of 3.98** on test data.
- **Deployed** the final model as an interactive web application using **Streamlit**, allowing for real-time delivery time predictions.

Airline Passenger Referral Program Development GitHub | Live Demo

- **Built a classification model** to predict passenger referral likelihood, addressing key business objectives for improving **customer retention and satisfaction**.
- **Mitigated class imbalance** using **SMOTE oversampling** techniques and handled missing data through advanced imputation strategies.
- **Achieved 87% accuracy** and **85% precision** for the positive class (referral) using a **Random Forest** classifier.
- **Deployed** the model with **Streamlit** for business user interaction, with analysis indicating a potential to **increase referral rates by 20%**.

EXPERIENCE

Data Analytics Intern (Simulation) | Accenture North America (via Forage) Jan 2024 - Feb 2024

- **Analyzed** 7 datasets for a social media client using **Python (Pandas, NumPy)** to uncover user trends, informing strategic recommendations that were presented to **20+ stakeholders**.
- **Automated** data cleaning processes, improving data reliability by **20%** and streamlining the analysis workflow for the team.

Alfido Tech (Remote) | Data Analytics intern Sep 2023 - Oct 2023

- **Engineered SQL queries** to analyze diverse datasets, leading to a **10% increase in forecasting accuracy** and supporting data-driven decision-making.
- **Optimized** resource allocation plans through data trend analysis, enhancing operational efficiency and customer experience.

CERTIFICATIONS

Career Essentials in Data Analysis (Microsoft & LinkedIn)
Complete Machine Learning, NLP Bootcamp, MLOPS & Deployment (Krish Naik, Udemy)
Introduction to Data Science (Infosys Springboard)

EDUCATION

B.Tech, Computer Science & Eng. (Sandip University, Nashik) | CGPA: 8.3/10 | 2022-2025
Diploma, Computer Technology (MVPS's RSM Polytechnic) | Score: 82% | 2020-2022