Shreyas Patil

Machine Learning Engineer | Data Analyst | Python | SQL | Deep Learning

📞 +917620280740   @ shreyas.syp@outlook.com   🔗 https://www.linkedin.com/in/shreypatil/   🔗 https://github.com/Shreyxpatil

SUMMARY

Aspiring Data Scientist with strong foundations in machine learning and data visualization, expert in Python and SQL. Key achievements include building and deploying predictive models that improved operational efficiency by 20% and coordinating projects leading to a 15% increase in project completion rate. Seeking a Data Scientist position, where my machine learning and data visualization skills can support your mission of delivering actionable insights and innovative data solutions.

EXPERIENCE

Jr.Machine Learning Engineer

Microsage Pvt. Ltd. (GATEtutor InfoEdge) 07/2025 – Present Pune

- Designed and deployed core **AI modules for PragyaAI** in an EdTech environment, enhancing adaptive learning systems.
- Developed **automated question generation** using NLP aligned with Bloom's Taxonomy, improved assessment scalability.
- Engineered **diagram generation models** leveraging **prompt-based techniques** and **structured contextual data** for concept visualization.
- Established **code shell generators** for programming exercises, automating code scaffolding for learners.
- Spearheaded the **design and fine-tuning of a domain-specific Large Language Model (LLM)** for education, including **dataset preparation, model training, evaluation**, and **deployment using MLOps.**
- Collaborated with **cross-functional teams** including data scientists, product managers, and backend engineers to integrate AI capabilities into a **scalable cloud-based EdTech platform.**
- Implemented **CI/CD pipelines** for seamless ML model integration and versioning in production.

Machine Learning Intern

EdTech Society - IIT Bombay   06/2024 - 09/2024   Remote

- Engineered a **facial emotion recognition model** using **ResNet CNN architecture** with 78% test accuracy on the DAiSEE dataset.
- Preprocessed and labeled **10,000+ video frames**, automating emotion classification for online learning engagement.
- Enforced **real-time emotion analytics pipeline** using Python and OpenCV, aiding in adaptive content delivery.
- Delivered insights to researchers through **interactive visualizations** built using **Matplotlib and Seaborn**.

Machine Learning Intern

Prarambh Development Cell   09/2023 - 01/2024   Kolhapur

- Developed and trained a **deep learning pipeline** to classify Alzheimer's stages (Non-Demented to Moderate) from MRI data.
- Achieved **87% classification accuracy** using CNNs with improved **image augmentation and dropout regularization**.
- Performed **error analysis and confusion matrix evaluation**, identifying class imbalance and optimizing F1-score.
- Collaborated in a team of 4 to **deploy the model via Flask REST API**, supporting real-time inference for healthcare use cases.

Data Analyst Intern

Simdaa Technologies   06/2023 - 07/2023   Pune

- Developed and deployed **Power BI dashboards** analyzing sales data from **500,000+ records**, enabling real-time KPIs tracking.
- Optimized **SQL queries and DAX measures** to reduce dashboard load times by 40%, enhancing user experience.
- Built interactive **forecasting models** for sales trends using Excel + Power BI, improving reporting accuracy by 30%.
- Created a data mart structure from raw datasets for **market segmentation**, improving analytics scalability.

SKILLS

AWS Python SQL Java

Machine Learning Deep Learning

Data Analysis Model Deployment TensorFlow

Keras Scikit-learn Pandas NumPy

OpenCV PyTorch Predictive Models

Image Classification Sentiment Analysis

Data Cleaning EDA Feature Engineering

Regression Classification Clustering

CNNs LSTM Power BI Excel

Jupyter Notebook Flask API Git GitHub

LandChain LangGraph NLP

Time Series Forecasting Data Pipelines

Inferencing LLM Fine tuning RAG

KEY ACHIEVEMENTS

- 💎 Selected as a Microsoft Learn Student Ambassador (Beta level) and mentored 500+ students through community tech events and workshops.

- 💎 Conducted 20+ training sessions on Git, Power BI, Web Development, DBMS, and LinkedIn optimization for engineering students.

- 💎 Deployed 8+ machine learning models (including Alzheimer's Detection and AI Chatbot) using Flask and REST API for real-time use cases.

- 💎 Achieved 87%+ accuracy on medical and NLP models, contributing to healthcare and sentiment analysis applications.

- 💎 Completed 6+ real-world data science projects covering domains like healthcare, finance, climate, and e-learning.

- 💎 Improved dashboard performance by 30% and enabled scalable data reporting using Power BI and SQL in a live internship project.

- 💎 Held 5+ leadership positions including Head of MLSA, Secretary of ASPIRE, and Internship Coordinator at T&P Cell.

- 💎 Organized technical workshops and hackathons that reached hundreds of participants, enhancing peer upskilling and collaboration.

PROJECTS

Adaptive Text-to-SQL Fine-Tuning using Qwen 2.5 Instruct
11/2025 - Ongoing

- Fine-tuned Qwen 2.5 Instruct on a 1M+ multi-domain Text-to-SQL dataset for natural language to SQL generation.
- Built automated data cleaning and merging pipeline for SQL datasets.
- Evaluated with Exact Match and Execution Accuracy metrics.
- Used LoRA fine-tuning to reduce GPU use and retain accuracy.

FishClone: End-to-end voice cloning
10/2025 - 11/2025

- Implemented a complete Python inference pipeline for Fish-Speech 1.5, replacing the official CLI-only workflow.
- I Added automatic speaker-token extraction from Wav files and multi-sentence batch synthesis.
- Built 13-language voice cloning with checkpoint + VQGAN support
- Released as a clean, modular repo with submodules and LFS tracking

Multi-Role Hybrid FAQ Chatbot
09/2025 - 09/2025

- Built a a RAG-based multi-role FAQ chatbot that combines dense vector search (FAISS) and lexical search (TF-IDF) to deliver high-accuracy, role-specific answers.
- Used MPNet embeddings with fine-tuning for semantic matching.
- Built a Flask REST API with role-based endpoints and fallback logic.
- Designed modular architecture with vector indexing and JSON knowledge bases.

SAR Satellite Image Colorization for Climate Monitoring
11/2024 - 01/2025

- Built an image-to-image deep learning model to colorize grayscale SAR satellite images with 85% accuracy.
- Identified climate, illegal activity, and agriculture patterns across 5,000+ images through area.
- Applied CNN and image enhancement methods to improve visual realism.
- Automated analysis of remote sensing data for environmental research

Alzheimer's Stage Detection using Deep Learning
09/2024 - 01/2025

- Developed a CNN model to predict dementia stages from 86,000+ MRI images with 87% accuracy.
  Applied image augmentation, dropout, and normalization to improve model generalization.
- Deployed model using Flask and REST API for real-time stage classification and Analysis.
- Conducted error analysis and optimized F1-score for multiclass classification.

EDUCATION

B.Tech in Computer Science (AI & ML)
D.Y.Patil College of Engineering and Technology

2021 - 2025   Kolhapur, Maharashtra   GPA 7.5

KEY ACHIEVEMENTS

- Improved image classification performance by 5% during internship by refining CNN model architecture and preprocessing techniques.

- Recognized for cross-functional collaboration and hands-on contributions to both development and deployment phases in project teams.

LEADERSHIP & ACTIVITIES

Head
Microsoft Learn Student Ambassadors, DYPCET (Aug 2024 – 2025)

Secretary
ASPIRE (AIML Student Association), DYPCET (Aug 2023 – May 2024)

Operations Coordinator
Microsoft Learn, DYPCET (Aug 2023 – Dec 2024)

Internship Coordinator
Training & Placement Cell, DYPCET (Aug 2022 – May 2023)

Technical Coordinator
Prarambh Development Cell, DYPCET (Aug 2023 – May 2024)

CERTIFICATION

Oracle Cloud Infrastructure 2025 Certified Generative AI Professional

Oracle Cloud Infrastructure 2025 Certified Data Science Professional

Oracle Cloud Infrastructure 2025 Certified AI Foundations Associate

Python for Data Science, AI & Development – Coursera (IBM)

Fundamentals of Deep Learning – NVIDIA

Data Analytics Essentials – Cisco Networking Academy

LANGUAGES

English

Hindi