

# MAHADEV CHAVAN

+91 8149287361 [chmahadev321@gmail.com](mailto:chmahadev321@gmail.com) [LinkedIn](#)

## Summary

Data Science Engineer with experience in building scalable ML and AI solutions across computer vision, NLP, and predictive analytics. Skilled in Python, SQL, cloud platforms, and deploying production-grade models that improve efficiency and decision-making. Strong background in data-driven problem solving and cross-functional collaboration.

## Skills

- **Languages & Frameworks:** Python, SQL, TensorFlow, Keras, OpenCV, Pandas, NumPy, RAG, Generative AI, CNN, NLP
- **Tools:** AWS (S3, Glue, Redshift, Lambda), Docker, Kubernetes, FastAPI, Flask, Git, Linux, Tableau, PySpark, Jupyter, VS Code

## Professional Experience

### Data Science Engineer, ([Ksolves India Limited](#))

Pune, India 04/2024 - Present

- Designed and delivered production-grade ML solutions across computer vision, NLP, and predictive analytics to solve real-world operational challenges.
- Computer Vision – Applicant Monitoring (POC): Designed a computer vision-based monitoring system using facial landmark detection, gaze estimation, and geometric analysis to assess candidate attention and engagement during coding evaluations, demonstrating feasibility for real-time behavioral analysis.
- Predictive Maintenance – HVAC Systems: Developed and deployed an AI-powered predictive model to monitor HVAC system performance, enabling proactive maintenance and operational optimization through data-driven insights.
- Intelligent Inventory Automation (NLP-Based): Engineered an NLP-driven automation system using semantic similarity techniques to map customer requirements to inventory items, reducing manual effort by ~ 75% and improving estimation accuracy; deployed as a production-ready API using FastAPI.
- Automated Ticket Routing System (Salesforce Integration): Built an ML-driven ticket assignment system integrated with Salesforce, leveraging historical resolution data, ticket metadata, and engineer workload to optimize routing. Deployed as a RESTful API using Django, containerized with Docker, and orchestrated via Kubernetes, ensuring scalability and reliability.

### Trainee Engineer, ([Neosoft Technologies Pvt Ltd](#))

Pune, India 05/2022 - 01/2023

- Assisted senior data scientists in collecting, cleaning, and preprocessing structured and unstructured data from databases, APIs, and external sources; performed data wrangling and EDA using Pandas, SQL, and Matplotlib to extract insights.
- Contributed to end-to-end data pipelines, including computer vision tasks using OpenCV and CNNs, as well as model training, evaluation, and performance analysis.

## Education

MSc Computer Science (70.28%) [Savitribai Phule Pune University](#)

Pune, India 2018 - 2020

PG Diploma In AI & ML (Distance) (First Class) [University Of Hyderabad](#)

Pune, India Feb 2021 - Feb 2022

PG Diploma in Big Data Analytics (69.37%) [CDAC Pune](#)

Pune, India Sept 2023 - Mar 2024

## Projects

### Intelligent Document Query & Knowledge Retrieval System (Generative AI)

- Designed a Generative AI-powered document intelligence system to ingest, index, and retrieve information from PDFs and structured/unstructured files (documents, logs, code files).
- Implemented RAG (Retrieval-Augmented Generation) with agent-based workflows for accurate context-aware querying and reasoning.
- Built efficient document parsing, embedding generation, and similarity search pipelines for fast information retrieval.
- Enabled scalable, modular agent interaction for query understanding and response generation.

**Technologies:** Python, LangChain, LLM APIs, Vector Databases (FAISS / Chroma), Embeddings, Generative AI, REST APIs.

### Hindi Character Recognition (Deep Learning)

- Developed a CNN-based deep learning model for multi-class Devanagari character recognition, achieving 96% training and 94% test accuracy.
- Performed EDA, preprocessing, and visualization using Pandas, NumPy, Matplotlib, and Keras.

**Technologies:** Python, TensorFlow, CNN, OpenCV.

### Nationwide Distributed Energy Resource Modeling & Analysis [Github Link](#)

- Designed and implemented a scalable data pipeline using PySpark and AWS (S3, Glue, Redshift) for large-scale energy analytics.
- Streamlined ingestion, transformation, and analytics workflows using Boto3 and Tableau.

**Technologies:** PySpark, AWS, Boto3, Tableau.

## Online Courses & Certifications

- Generative AI with Large Language Models (Aug. 2025) - Coursera [Certificate Link](#)
- Fundamental course in the AWS Machine Learning Scholarship (2020)- Udacity. [Certificate Link](#)
- Microsoft Technology Associate(MTA) (2019)- Microsoft [Certificate Link](#)