



KISHAN YADAV

Hyderabad, Telangana

📞 +91 8826495998 ✉️  

Objective

Innovative and results-driven data science graduate with practical experience in building AI-powered applications, LLM-integrated systems, and real-time intelligent solutions. Skilled in Python, machine learning, and full-stack development using tools like Streamlit, LangChain, and Gradio. Proficient in leveraging large language models such as Mistral-7B (Ollama) to generate automated insights, intelligent feedback, and conversational interfaces. Developed impactful projects including a virtual mock interview assistant, fake job posting detection system, and smart waste classification app. Passionate about solving real-world problems through data-driven automation, and eager to contribute to forward-thinking teams in AI research, product development, or data analytics.

Education

Aravali College of Engineering & Management

B.Tech in Computer Science (AI & ML)

Aug 2020 – Aug 2024

Faridabad, Haryana

Shree Ram Model School

Senior Secondary (12th)

Apr 2019 – Apr 2020

Faridabad, Haryana

Projects

MockMate: Virtual Mock Interview Assistant | *Python, Streamlit, LangChain, Ollama*

GitHub

- Built an AI-powered mock interview web app that generates personalized technical and behavioral questions from resumes using Mistral-7B via Ollama.
- Enabled voice-based answering with in-app recording, transcription using SpeechRecognition, and feedback scoring using LangChain prompts.
- Displayed live feedback, score summaries, and progress tracking for each question in the interview session.
- Exported interview results in both PDF and JSON formats for performance review and future analysis.
- **Client Approach:** Designed to help job seekers practice interviews with AI-generated coaching and evaluation.

JobGuard: Fake Job Posting Detection | *Machine Learning, Streamlit, Sklearn*

GitHub

- Built a Streamlit-based web application to detect fake job postings using machine learning techniques.
- Engineered features from job descriptions, extracted salary patterns, and performed predictive imputation for missing categorical values.
- Trained and optimized a RandomForestClassifier model using GridSearchCV with TF-IDF and OneHotEncoding in a full pipeline.
- Achieved high classification accuracy and provided adjustable detection thresholds for user-driven confidence levels.
- **Client Approach:** Designed as a fraud prevention tool for HR platforms, recruitment agencies, and job seekers.

EcoVision: Intelligent Trash Sorting | *Deep Learning, TensorFlow, OpenCV*

GitHub

- Developed a real-time intelligent trash classification system to promote automated waste segregation in smart bins.
- Trained a MobileNetV2-based CNN on the TrashNet dataset to classify garbage into plastic, paper, metal, glass, and organic categories.
- Achieved over 90% classification accuracy using data augmentation and preprocessing for improved generalization.
- Built a real-time camera interface with OpenCV to detect and classify waste on live video streams.
- **Client Approach:** Positioned as a solution for smart city projects, IoT-driven waste management, and sustainability tech.

Text-to-Image Generation | *Stable Diffusion, Hugging Face, Python*

GitHub

- Created a pipeline to convert natural language prompts into realistic images using Stable Diffusion v2.1.
- Used Hugging Face diffusers for model access and GPU acceleration for performance.
- Explored prompt engineering techniques for style, clarity, and concept generation.
- Enabled flexible output options for resolution, theme, and artistic control.
- **Client Approach:** Demonstrated capabilities for marketing, design, and creative content generation.

Technical Skills

Languages: Python

Tools/Platforms: VS Code, Jupyter Notebook, PyCharm, Spyder, Git, GitHub, MS Excel, Tableau, Streamlit Cloud, Ollama

Cloud & Big Data: Microsoft Azure ML, Azure ML Pipelines, Databricks, PySpark

Database: MySQL

Libraries/Frameworks:

- **Data Science & ML:** NumPy, Pandas, Seaborn, Matplotlib, Scikit-learn, XGBoost, LightGBM, SciPy
- **Deep Learning:** TensorFlow, PyTorch, Keras, Theano
- **Computer Vision:** OpenCV, MediaPipe, YOLO, Haar Cascade Classifier
- **NLP:** SpaCy, NLTK, Gensim, Hugging Face Transformers
- **LLMs & Generative AI:** LangChain, Hugging Face Transformers, GPT-3/4, Mistral-7B (Ollama), LLaMA 3, Gemini, Prompt Engineering, BERT, ChatGPT, AutoGen, Transformers Agents
- **Web & GUI:** Streamlit, Gradio, Tkinter
- **Automation & Others:** PyAutoGUI, Pycaw, SpeechRecognition, BeautifulSoup
- **MLOps:** MLflow, GitHub Actions, Azure ML Pipelines — Model versioning, experiment tracking, training pipeline automation, model registry, lifecycle management
- **LLMOps:** LangSmith — Prompt evaluation, LLM observability, debugging chains, testing prompt flows

Concepts & Techniques:

- **Machine Learning (Supervised):** Linear & Logistic Regression, KNN, SVM, SVR, Naive Bayes, Decision Trees, Random Forest, Gradient Boosting, PCA
- **Unsupervised Learning:** K-Means, Hierarchical Clustering, DBSCAN
- **Deep Learning Architectures:** CNN, RNN, ANN, LSTM
- **NLP Techniques:** Tokenization, Lemmatization, NER, Word Embeddings, Sentiment Analysis, Text Classification
- **Transformers & Agentic AI:** Prompt Engineering, Context-aware Chains, Retrieval-Augmented Generation (RAG), LLM Agents, Tool Use with LangChain
- **Generative Models:** GANs, VAEs
- **Data Science Workflow:** EDA, Feature Engineering, Data Preprocessing, Hyperparameter Tuning, Regularization (L1/L2)

Soft Skills: Analytical Thinking, Problem Solving, Adaptability, Collaboration, Communication, Research-Oriented

Certifications

- **NPTEL** — Soft Skills
- **NPTEL** — Psychology of Stress and Well-being
- **NPTEL** — Principles of Management
- **Naresh i Technologies** — Full Stack Data Science and AI (Ongoing)

Languages

- Hindi – Native
- English – Proficient