Kishan Yadav

Hyderabad, Telangana
+91 8826495998

Objective

Motivated and detail-oriented data science graduate with hands-on experience in AI-powered analysis, machine learning, and data visualization. Skilled in Python, predictive modeling, and building interactive applications using Streamlit and Gradio. Proficient in leveraging large language models like Mistral-7B (Ollama) and statistical tools such as Pandas, Seaborn, and Matplotlib for automated EDA and insightful analytics. Passionate about developing intelligent, data-driven solutions that bridge the gap between complex data and strategic decision-making. Eager to contribute to innovative teams in AI development, data analytics, or business intelligence to drive meaningful impact.

Education

Aravali College of Engineering & Management **Aug 2020 – Aug 2024**
B.Tech in Computer Science (AI & ML) *Faridabad, Haryana*

Shree Ram Model School **Apr 2019 – Apr 2020**
Senior Secondary (12th) *Faridabad, Haryana*

Projects

AI-Powered EDA App | *Python, Pandas, Seaborn, Matplotlib, Mistral-7B, Gradio* **GitHub**

- Built an automated EDA tool for CSV files using Pandas, Seaborn, and Matplotlib.
- Integrated Mistral-7B (Ollama) to generate AI-driven data insights and summary.
- Enabled real-time analysis through an intuitive Gradio-based web interface.
- Visualized key metrics using histograms, heatmaps, and correlation plots.
- **Client Approach:** Designed to assist analysts and business teams in performing quick, automated EDA.

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, PySpark, Databricks

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
- **Web & GUI**: Streamlit, Gradio, Tkinter
- **Others**: PyAutoGUI, Pycaw, SpeechRecognition, BeautifulSoup

Concepts & Techniques:

- **Machine Learning (Supervised)**: Linear & Logistic Regression, KNN, SVM, SVR, Naive Bayes, Decision Trees, Random Forest, Boosting, PCA
- **Unsupervised Learning**: K-Means, Hierarchical Clustering, DBSCAN
- **Deep Learning**: CNN, RNN, ANN, LSTM
- **NLP Techniques**: Tokenization, Lemmatization, NER, Word Embeddings, Sentiment Analysis, Text Classification
- **Transformers & LLMs**: BERT, GPT-3/4, Gemini, LLaMA 3, Mistral-7B, LangChain, Prompt Engineering
- **Generative Models**: GANs, VAEs
- **Data Science**: EDA, Feature Engineering, Data Preprocessing, Regularization (L1/L2)

Soft Skills: Analytical, Problem Solver, Adaptable, Collaborative, 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