Manish Kumar Mahto

Machine Learning Engineer

Github | LinkedIn | Portfolio | manishcode123@gmail.com | +91 7667056502 | Ranchi , Jharkhand, India

SUMMARY

Entry-level Data Scientist with strong foundations in Python, ML, Deep Learning, and Generative AI frameworks (LangChain, RAG, Transformers). I bring the ability to build and deploy intelligent AI solutions that improve efficiency, enhance decision-making, and deliver measurable business value.

EXPERIENCE

KagglePro LLC   **Sept 2024 – present**
Data Science intern   LATHROP, CA

- **Model Development:** Designed and optimized 5+ ML/DL models for structured and unstructured datasets, achieving up to 12% improvement in predictive accuracy compared to baseline.
- **Pipeline & Collaboration:** Contributed to end-to-end data pipelines, 20+ code reviews, and internal bootcamp notes, improving team productivity and reducing debugging time by 15%.
- **Advanced AI Prototyping:** Applied Transformers, LangChain, and RAG pipelines to develop 2 intelligent prototypes, showcasing practical applications of retrieval-augmented generation in real-world use cases.

TECHNICAL KNOWLEDGE

- **Programming Languages:** Python, SQL, Java
- **Data Science Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, XGBoost, Streamlit
- **Deep Learning Frameworks:** TensorFlow, Keras, PyTorch, Sklearn
- **Machine Learning Techniques:** Regression, Classification, Clustering, Feature Engineering, Model Deployment, Time Series Forecasting
- **Deep Learning Techniques:** ANN, CNN, RNN, LSTM, GRU, Transformers, GANs
- **Databases:** MySQL, ChromaDB
- **Data Engineering:** Data Cleaning, Data Pipelines, Exploratory Data Analysis (EDA)
- **AI Frameworks:** LangChain, RAG Systems, Retrieval-Augmented Generation, Generative AI
- **Tools & Platforms:** Git, GitHub, Jupyter Notebooks, Excel
- **Mathematics & Statistics:** Probability, Statistics, Linear Algebra, Calculus

PROJECT

Movie Recommendation System   Github   (Python, Pandas, Scikit-learn, Streamlit, and Google Colab)

- **Content-Based Recommendation Development:** Built a personalized movie recommendation engine using text vectorization and cosine similarity, improving memory efficiency by 75%.
- **Optimization & Deployment:** Designed and deployed an interactive Streamlit web app that delivers real-time recommendations with detailed movie insights, enhancing user experience and engagement.

Quora Duplicate Question Pair Github (Numpy, Pandas, Matplotlib, Seaborn, nltk , Scikit learn, Xgboost)

- **Duplicate Question Classification:** Developed a machine learning model to detect duplicate question pairs from the Quora dataset, achieving 78% accuracy using a Random Forest Classifier.
- **Feature Engineering & Text Processing:** Engineered custom linguistic features (e.g., common word ratios) and applied advanced text processing techniques such as fuzzy string matching and text vectorization, improving model robustness.

EDUCATION

Ranchi University   **Sept 2022 - Nov 2025**
Bachelor of Science in Information Technology   Ranchi,Jharkhand, India