

# Digvijay Singh Shekhawat

+91-8455025660 | tiku12004@gmail.com | LinkedIn | GitHub: ShekhawaTTiku

## Summary

---

Artificial Intelligence and Machine Learning undergraduate with hands-on experience developing AI-powered data analytics applications, machine learning pipelines, and deep learning solutions. Experienced with Python, SQL, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), NLP, feature engineering, and predictive analytics. Interested in Generative AI, computer vision, and building end-to-end AI systems that transform structured and unstructured data into actionable insights.

## Experience

---

**Anmoul Infomatics, ML Intern**

Aug 2025 – Jul 2026

- Automated cleaning, validation, and preprocessing of structured Excel/CSV datasets using Python, preparing analysis-ready data for exploratory data analysis and visual reporting.
- Built document processing workflows using Tesseract OCR, NLP, and document classification to extract text and structured fields from scanned PDFs and multi-format files.
- Developed embedding-based semantic search and an LLM knowledge assistant using RAG; refined document chunking and similarity-based retrieval to improve response relevance and grounding.

## Projects

---

**Fantasy Cricket Prediction & Team Optimization**

[GitHub](#) | [Live Demo](#)

- **Won 1st Place at FIFS Sports Data Gameathon 2025 2.0.** *Certificate*
- Engineered player-form, venue, and matchup features using pandas and NumPy for IPL fantasy-point prediction with PyCaret models.
- Implemented PuLP integer programming to optimize 11-player teams within 100 credits and 15-player squads within 145 credits; deployed a Streamlit app with API-based data ingestion and CSV export.

**DataBuddy – LLM-Powered Conversational Data Analyst**

[GitHub](#) | [Live Demo](#)

- Built a Streamlit analytics assistant that converts natural-language queries into pandas code for CSV/Excel analysis and interactive Plotly visualizations.
- Implemented dataset profiling, AST-based code validation, and subprocess execution with a configurable 5-second timeout.
- Integrated Groq, Mistral, and Gemini APIs for LLM routing, provider fallback, and conversation summarization to support contextual follow-up queries.

**NLP-Based Automated Fake News Detection**

[GitHub](#)

- Developed a LIAR2 classification pipeline using claim text, fact-checker justifications, and speaker credibility, with 18,369 training and 2,296 test examples.
- Fine-tuned bert-base-uncased with a hybrid text-credibility classifier in PyTorch, achieving **0.8830 weighted F1** versus **0.8684** for SVM.
- Benchmarked SVM, Random Forest, and BiLSTM; engineered 50,000 TF-IDF features and applied SHAP to explain SVM predictions across 200 test samples.

## Technical Skills

---

**Languages:** Python, SQL, Java, C, C++

**Machine Learning & Deep Learning:** Scikit-Learn, PyTorch, TensorFlow, PyCaret, XGBoost, Gradient Boosting, CNNs, Vision Transformers, Transfer Learning, Fine-Tuning, Feature Engineering, Model Evaluation, PuLP

**Generative AI & NLP:** LLMs, Hugging Face Transformers, BERT, TF-IDF, SHAP, LangChain, LlamaIndex, RAG, Sentence Transformers, Embeddings, Semantic Search, Prompt Engineering, AI Agents, Function Calling, Tool Calling

**Data Engineering:** pandas, NumPy, PostgreSQL, Apache Spark, PySpark, ETL, Data Pipelines, Data Cleaning, Data Wrangling, Data Validation

**Development & Visualization:** FastAPI, Streamlit, Gradio, Git, GitHub, Plotly, Matplotlib, Seaborn

## Education

SOA - Institute of Technical Education and Research

B.Tech in Computer Science (Data Science Specialization) | GPA: 7.5/10.0

2022 – 2026

## Key Achievements

FIFS Sports Data Gameathon 2025 1.0: Qualified for the final round and secured a Top 20 finish among national participants.

FIFS Sports Data Gameathon 2025 2.0: Achieved 1st Place (Winner) among top university teams in a national-level competition. *Certificate*

Amazon ML Challenge 2024: Ranked in the Top 500 teams out of approximately 10,000 teams on the public leaderboard.

## Certifications

IBM Data Analyst Professional Certificate — Coursera | Data Science using Python — Coursera

Google AI Essentials | Hugging Face NLP Course

Kaggle: Python, Pandas, Machine Learning, Feature Engineering & Deep Learning