Vaibhav Kumar

Computer Science Student — AI Data Science Enthusiast

vaibhavkkr24@gmail.com — +91 7004931269
LinkedIn — GitHub — Portfolio

PROFESSIONAL SUMMARY

Computer Science student with hands-on experience in **Machine Learning, Deep Learning, and Generative AI**. Built and deployed real-world projects including **Ransomware Detection Systems, Crime Analytics, Spam Email Classifier, Teen Phone Addiction Analysis, and a Movie Recommendation System**. Skilled in Python, data preprocessing, and model optimization, with practical exposure from multiple AI/ML internships. Strong analytical mindset, problem-solving ability, and passion for turning data into intelligent solutions.

TECHNICAL SKILLS

- **Languages:** Python, C, C++, SQL, HTML, CSS, JavaScript
- **Libraries & Frameworks:** Scikit-learn, TensorFlow, Keras, Pandas, NumPy, Matplotlib, Seaborn, OpenCV, Streamlit, FastAPI, Flask, XGBoost
- **Generative AI:** LangChain, Llama3, Ollama, FAISS, ChromaDB, Hugging Face Transformers, OpenAI API
- **Databases & Version Control:** MySQL, MongoDB, Git, GitHub
- **Cloud & MLOps:** Oracle Cloud Infrastructure (OCI), AWS, Docker

PROJECTS

Ransomware Anomaly Detection (Isolation Forest) [GitHub] *Python, Scikit-learn, Isolation Forest, FastAPI*

- Simulated realistic backup logs with injected ransomware attack patterns like mass deletions.
- Engineered time-based features to capture activity pace and context for anomaly detection.
- Trained an **Isolation Forest** model and deployed via FastAPI to flag suspicious behavior in real time.

Metadata Threat Radar (Ransomware Detection) [GitHub] *Python, Scikit-learn, RandomForest, FastAPI*

- Built a real-time ransomware detection system analyzing file metadata for suspicious patterns.
- Trained a **RandomForestClassifier** with custom logs, achieving **91% recall** on injected threats.
- Deployed as a FastAPI service with explainable alerts (e.g., "Mass Deletion suspected: 35 deletes in 5 mins").

Smart Crime Analytics [GitHub] *Python, Scikit-learn, XGBoost, TensorFlow*

- Built to analyze and predict crime trends for public safety and policy insights.
- Utilized real-world dataset from the National Crime Records Bureau (NCRB), with socio-economic feature engineering.
- Achieved **90.9% accuracy (XGBoost)**; Random Forest at **87.6%**; surfaced regional and temporal patterns.

Spam Email Classification *Python, Scikit-learn, Random Forest, Gradient Boosting*

- Built to detect and classify spam emails for improved digital communication security.
- Used UCI Spambase dataset (4601 samples, 57 features) with preprocessing, scaling, and feature engineering.
- Achieved **95.1% accuracy, F1 0.9367, ROC-AUC 0.9834** using tuned Random Forest; top signals: `char_freq_!`, `char_freq_$`, `word_freq_free`.

AI Agent for Teen Phone Addiction Analysis [GitHub] *Python, LangChain, Ollama, ChromaDB*

- Built an interactive agent to analyze teen phone usage patterns and addiction levels.
- Used a custom dataset (age, daily usage, sleep hours, addiction level) stored in a Chroma vector DB with Ollama embeddings.
- Enabled context-aware Q&A with LangChain and the Tinydolphin LLM to surface behavioral insights.

Movie Recommendation System [GitHub] *Python, Scikit-learn, Streamlit, NLP*

- Built a content-based **movie recommendation system** using TF-IDF vectorization and cosine similarity.
- Processed metadata (genre, cast, keywords, and overview) to generate personalized movie suggestions.
- Deployed interactive web app using **Streamlit** for real-time recommendations and search functionality.

WORK EXPERIENCE

Summer Intern – Data Science *Celebal Technologies*
June 2025 – August 2025

- Assisted in data preprocessing, exploratory analysis, and visualization tasks.
- Gained exposure to production-level data pipelines and real-world business datasets.
- Contributed to model development and validation for internal analytics tools.

Intern – Artificial Intelligence & Machine Learning *YBI Foundation*
June 2024 – December 2024

- Developed and evaluated predictive models using deep learning and ML frameworks.
- Collaborated on AI-based research projects with focus on classification and forecasting.
- Documented project lifecycle including data handling, experiments, and results.

CERTIFICATIONS

Oracle Certified Professional: Oracle Cloud Infrastructure 2025 Data Science Professional
Oracle University
Issued: October 07, 2025 – Valid until: October 07, 2027

Oracle Certified Professional: Oracle Cloud Infrastructure 2025 Generative AI Professional
Oracle University
Issued: October 14, 2025 – Valid until: October 14, 2027

EDUCATION

- **Integrated M.Sc. in Computer Science**, Central University of Rajasthan *2021–2026*
- **12th (CBSE)**, British English Gere, Gaya, BR *2021* – 73.4%
- **10th (CBSE)**, National Public School, Gaya, BR *2019* – 80.8%