

Sahil Rai

Machine Learning & AI Engineer

<https://www.linkedin.com/in/sahil-rai-b4a950352>

sahil.raii234234@gmail.com

<https://sahilrai1225.github.io/sahilrai1225-portfolio/>

Projects

Network Security – Phishing Detection System

- Built classification model using Scikit-learn achieving 88% accuracy on test data.
- Engineered 8+ features for URL analysis and optimized preprocessing pipeline.
- Trained and evaluated classification model on 10K+ samples to improve generalization performance.

Student Marks Predictor

- Developed regression model achieving R^2 score of 0.8655.
- Applied feature scaling, preprocessing, and performance evaluation (RMSE, R^2).
- Built end-to-end ML pipeline and deployed via Flask web interface.

Hate Speech Detection – NLP Classification on Twitter Data

- Built NLP-based classification model to detect hate speech in social media text.
- Performed text preprocessing including tokenization, stopword removal, stemming/lemmatization, and vectorization (TF-IDF / embeddings).
- Trained with Deep Learning RNN Model
- Evaluated performance using Accuracy, Precision, Recall, and F1-score to handle class imbalance.

QA with PDF – Retrieval-Augmented Generation (RAG) System

- Designed RAG pipeline using LangChain and vector embeddings.
- Integrated vector database (DataStax/Pinecone-style) for semantic retrieval.
- Developed Streamlit-based interface enabling contextual document querying.

Movie Recommendation System

- Implemented TF-IDF vectorization and cosine similarity for content-based filtering.
- Designed normalized MySQL schema for scalable movie metadata storage
- Built Flask application serving personalized recommendations.

Technical Skills

- Python
- Pandas
- SQL
- Power BI
- Scikit-learn
- Tensorflow
- Flask
- MySQL
- LangChain
- RAG
- Generative AI

Education

Bachelor of Computer Applications

Institution: Institute of Management Studies Noida

Year of Graduation: June 2026

Certifications

- Supervised Machine Learning – DeepLearning.AI (Stanford)
- Advanced Learning Algorithms – DeepLearning.AI (Stanford)
- Data Analysis with Pandas – Udemy
- Complete Data Science, Machine Learning, Deep Learning, NLP Bootcamp – Udemy
- Generative AI with LangChain & HuggingFace – Udemy
- Advanced SQL & MySQL – Udemy