Anurag Indora

+91 7056717349 | anuragindora474@gmail.com | LinkedIn | GitHub | Portfolio

Education

Indian Institute of Technology (ISM) May 2025
Integrated M.Tech in Mathematics & Computing *Dhanbad, Jharkhand*

Experience

CodSoft May 2024 – June 2024
Machine Learning Intern

- Developed a fraud detection pipeline with **SMOTE**, **ADASYN**, and **XGBoost**, achieving **97.2% accuracy** on 284K+ transactions.
- Reduced false positives by **23%**, improving operational efficiency and reducing business risk exposure.
- Designed modular ML components (*loader*, *transformer*, *trainer*, *evaluator*) enabling scalable, repeatable workflows.
- Integrated CI/CD with **GitHub Actions**, ensuring robust deployment and minimizing downtime.
- Collaborated with cross-functional teams to align technical outputs with business KPIs, supporting data-driven decision-making.
- Optimized pipeline efficiency by automating preprocessing workflows, reducing manual overhead by 40%.
- Produced comprehensive performance reports for stakeholders, translating technical insights into actionable **business intelligence**.

Projects

Movie Recommender System | ***Python, TF-IDF, Cosine Similarity, Sparse Matrix, TMDb API, Streamlit***

- Developed a **content-based movie recommendation system** leveraging **TF-IDF vectorization** and **cosine similarity** across 5K+ titles.
- Optimized similarity storage using **compressed sparse matrices** (`.npz`) and **gzip-compressed pickles**, reducing model size by over 80%.
- Integrated **TMDb API** to dynamically fetch posters and metadata for the selected and recommended movies.
- Built an interactive and responsive interface using **Streamlit**, displaying real-time recommendations under one second.
- Enhanced system maintainability through modular architecture and improved inference efficiency with **on-demand sparse retrieval**.

Transformer-Based Text Summarizer | ***Python, HuggingFace, BART, FastAPI, Streamlit, Accelerate***

- Built a scalable abstractive text summarization pipeline using `facebook/bart-base`, fine-tuned on 50K+ samples from `cnn_dailymail v3.0.0` to generate concise summaries from long-form news articles.
- Structured the ML system into 6 modular stages (ingestion, validation, transformation, training, evaluation, prediction), improving development speed and maintainability by 40%.
- Achieved a **20.67 ROUGE-L** score and **3.6 eval samples/sec** on 4GB GPU by leveraging `fp16` `mixed-precision` and `gradient accumulation`, reducing memory usage by 50%.
- Developed dual-mode interfaces: CLI for retraining and a production-ready API via FastAPI + Streamlit UI to support real-time summarization and demo integration.
- Delivered a production-ready API, enabling integration into enterprise workflows and improving **knowledge management efficiency**.

Achievements

Kaggle Profile: kaggle.com/evilak09

Silver Medal (50th Rank) – NeurIPS Open Polymer Challenge (Kaggle, 2025).

Secured top academic standing in Mathematics & Computing coursework, demonstrating analytical rigor and quantitative problem-solving skills.

Technical Skills

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

Libraries/Frameworks: scikit-learn, XGBoost, PyTorch, TensorFlow, HuggingFace, FastAPI, Streamlit, Docker

Tools/DevOps: Git, GitHub Actions, Makefile, CI/CD, Azure

Databases: MySQL, PostgreSQL, SQLite

Core Concepts: ML Pipelines, Deployment, Feature Engg., Attention, RNNs (LSTM, GRU), EDA

Publications

Indora, A. (2025). *Facial Expression Recognition in Children using Graph Neural Networks*. Master's Thesis, IIT (ISM) Dhanbad.