Anshuman Singh

Gurugram, India | anshumanr434@gmail.com | linkedin.com/in/anshumansingh2023

Data & Machine Learning student (B.Tech CS AI/ML, CGPA 9.44/10) with hands-on experience delivering end-to-end solutions: data collection, feature engineering, model training, evaluation, lightweight deployment. Proven leadership in hackathons & internship projects. Strengths: reproducible analysis, clear documentation, structured experimentation, collaboration. Seeking roles in Data Science, ML Engineering, Analytics, or Software.

EDUCATION

SGT University, Gurugram - B.Tech Computer Science (AI & ML Specialization), 2023 - 2027
CGPA: 9.44/10.00 | Relevant Coursework: Data Structures & Algorithms; Machine Learning; Deep Learning; NLP; Databases; Statistics; Linear Algebra

EXPERIENCE

Data Science Intern - TechnoHacks Edutech (Remote) Jun - Aug 2024
- Analyzed 10K+ loan customer records; engineered interaction features boosting ensemble accuracy to 92.8% while improving targeting efficiency.
- Developed XGBoost, LightGBM, Random Forest models with SHAP & feature selection; delivered decision rules reducing acquisition cost.
- Produced concise technical report (EDA, assumptions, evaluation split) ensuring reproducibility & stakeholder clarity.

ML Team Lead - CodeSangam Hackathon (Healthcare Track, 3rd Place) Feb 2024
- Led 4-person team to ship Flask prototype in 48h: data collection, labeling, model training, evaluation & deployment pipeline.
- Curated Ayurveda symptom/remedy dataset (web/PDF) & generated synthetic samples improving minority class balance.
- Documented labeling guidelines & evaluation notes to speed iteration & reduce ambiguity.

Data Science Lead - AIFusion Hackathon (Special Recognition) Sep - Oct 2024
- Designed synthetic benchmarking datasets & validated multi-source data consistency (schema, range & outlier checks).
- Built concise visuals (trend & distribution charts) informing campaign strategy; recognized for data quality rigor.

PROJECTS

StreaxBot (Reddit Quiz Platform) - React, Node.js, Express, Firestore, Gemini API Oct 2024
- Implemented end-to-end quiz pipeline: user topic input - web scraping - data cleaning - question generation - delivery.
- Added real-time dedup & guardrail prompts; sampling review reduced low-quality question incidence.

Sentiment Modeling with Synthetic Augmentation - PyTorch, Transformers, LoRA, ADASYN
- Fine-tuned RoBERTa with LoRA adapters; ADASYN augmentation raised macro F1 by ~12% on imbalanced multi-class task.
- Published reproducible assets: dataset, preprocessing script, training notebook & concise model card.

LocalLibrary Metadata Scraper - Python, Playwright, asyncio
- Built asynchronous scraping pipeline (rate limits, UA rotation, session reuse) extracting multi-source book metadata.
- Normalized ISBN/title fields & produced data dictionary enabling downstream search & analysis.

LEADERSHIP & ACTIVITIES

- Self-directed paper summaries on model evaluation, interpretability & data quality; integrated learnings into projects.
- Maintained lightweight experiment logs (config, metrics, observations) improving iteration discipline.
- Created benchmarking scripts to compare models & calibration approaches on small structured & text datasets.

SKILLS

Languages: Python, JavaScript, SQL
ML & Data: PyTorch, scikit-learn, Hugging Face Transformers, XGBoost, LightGBM, Random Forest, Pandas, NumPy, SHAP, LoRA, ADASYN, Feature Engineering, Model Evaluation (accuracy, F1, calibration)
Data Engineering: Web Scraping (Playwright), REST APIs, ETL, Data Cleaning, Deduplication, JSON
Tools: Git, Linux, Docker (basic), Jupyter, Experiment Tracking (logs & scripts)
Concepts: Prompt Design, Feature Attribution, Reproducibility, Documentation, Agile Collaboration

HONORS & ACHIEVEMENTS

- CGPA 9.44/10 (AI & ML specialization)
- 3rd Place - CodeSangam Hackathon (Healthcare Track)
- Special Recognition - AIFusion Hackathon (Data Quality)
- Built multiple original datasets (NLP sentiment, healthcare symptoms) enabling ML prototypes.