

KETAN GUPTA

✉ 12212041ketanguptait@gmail.com

☎ +91 7067711850

🌐 [KetanGupta45-1](#)

in [ketangupta41](#)

Education

Indian Institute of Information Technology, Sonapat
B.Tech in Information Technology
CGPA: 8.44

Aug 2022 – May 2026
(Expected)

India International Public School, Dabra
Higher Senior Secondary **93%**

2022

Achievements

Amazon ML Challenge

Sept 2024

- Developed a **CNN + OCR pipeline** to extract product attributes from noisy e-commerce images, processing over **1M images**, and secured a **Top 2,500 rank out of 20,000+ participants**.

IPL Per Ball Prediction (Research-Based)

Apr 2025 – Present

- Developing DL models on **1,170 IPL matches** with **30+ features**, benchmarking **BiLSTM**, and **Transformers** in PyTorch to improve next-ball prediction accuracy using **f1-score**, **precision**, **recall**.
- Solved **350+ DSA Questions** on **LeetCode**, **Geek4Geeks**.

Projects

Online Fraudulent Transaction Detection (ML Classification)

Apr 2025

- Analyzed **6.3M+ transactions**, addressing class imbalance (80% non-fraud, 20% fraud).
- Engineered features using **VIF**, **PPS**, **mutual information**; removed outliers with **Isolation Forest**.
- Applied **SMOTE**; trained **Random Forest**, **CatBoost**, **Voting**, and **Stacking models**.
- Achieved **Precision: 99%**, **Recall: 88%**, **F1-score: 92%**.

Loksabha Elections 2024 Dashboard (Power BI/SQL)

July 2025

- Queried SQL Server schemas to extract KPIs from **543 constituencies**.
- Designed an interactive **Power BI dashboard** segmented by candidate, party, and region.
- Enabled drill-through and filter analysis to uncover vote shares, win margins, and swing constituencies.

Jigsaw Comment Classification Challenge (NLP/LLM)

Aug 2025 – Sep 2025

- Tackled a **highly imbalanced multi-label NLP task** on **160K toxic comments** across six categories.
- Improved rare-class recall by **10%** using **focal loss**, **weighted sampling**, and **threshold tuning**.
- Progressed from **BiGRU (75%/73%)** to **DistilBERT (80%/80%)** and **RoBERTa (75%/85%)** in **precision/recall**.
- Deployed as an **interactive Streamlit app** with **FastAPI** backend.

Skills

Languages: Python, C++, SQL

ML & AI: Scikit-learn, PyTorch, Transformers, NLP, Generative AI

Tools & Frameworks: Power BI, Docker, Git, GitHub, Jupyter, Google Colab, FastAPI

Soft Skills: Communication, Presentation, Leadership, Problem Solving

Core Concepts: Data Structure, Operating System, Computer Network, OOP, Database Management System