

# Kshitij Chaturvedi

Data Scientist | ML Engineer

Delhi | +91 9818627198 | kshitijhrms@gmail.com | [Portfolio](#) | [GitHub](#) | [LinkedIn](#) | [Medium](#) | [Twitter](#)

## EDUCATION

Guru Gobind Singh Indraprastha University (ADGIPS), Delhi	
B. Tech. (Electronics and Communication Engineering): Overall GPA - 9.0 (8th Sem.)	2021 - Present
Hansraj Model School, New Delhi	
12th (Physics, Chemistry, and Mathematics): Percentage - 84%	2020 - 2021
10th: Percentage - 88%	2018 - 2019

## WORK EXPERIENCE

**DLC Group of Companies: Data Analyst Intern** (Jan 2025 - April 2025)

- Conducted data cleaning, transformation, and statistical analysis using Python (NumPy, Pandas, SciPy, Scikit-Learn) and SQL to deliver actionable insights.
- Implemented ML pipelines for trend prediction in sales and occupancy data using time series models and regression analysis for strategic planning.
- Built and deployed a RAG workflow using SBERT + FAISS + BART/T5 (using Sentence Transformers, Hugging Face Transformers, LangChain, DVC, FastAPI, and Docker) to automate internal and client query resolution leveraging in-house knowledge base.
- Collaborated with business teams to analyze customer/property data and automated reporting workflows, improving insight delivery and reducing manual effort by ~30%.

**Creative Ai Solutions [NYNEXA]: Backend Dev. (Intern)** (April 2024 - May 2024)

- Developed initial backend architecture for an AI/Web3 Marketplace, boosting system efficiency by 20% using AI dev tools.
- Established a scalable backend foundation with NextJS, Tailwind CSS, MonoRepo - TurboRepo, MongoDB, & Prisma.

**SyntHeim [Radian Arc]: SDE Intern** (Feb 2024 - Aug 2024)

- Contributed to developing an AI-driven health application using Machine Learning & NLP.
- Utilizing libraries like Numpy, Pandas, Scipy, Scikit-Learn, TensorFlow, SpaCy, NLTK, Open-Ai, & Flask.
- Led a team of 8 developers to implement a virtual AI assistant for real-time personalized and multilingual responses.

## PROJECTS

**Domain-Specific RAG:** [Research Paper]

- Built a low-latency, high-accuracy RAG pipeline for enterprise support systems with SBERT (semantic embeddings), FAISS (dense retrieval) & BART/T5 (contextual response generation).
- To perform effectively in resource-constrained environments, outperforming baselines (GPT-3.5, BART-only, BM25) in **F1**: 0.627, **ROUGE-L**: 0.427, **BLEU**: 0.678, **Faithfulness**: 88.7%, with 28% reduced **latency** (620 ms/query).
- Processed and embedded 10K+ real estate/legal documents and 2K+ user queries using spaCy for NLP pre-processing and Tesseract OCR for digitizing scanned PDFs; embeddings stored and indexed with FAISS.
- Implemented using LangChain (pipeline orchestration), Transformers (Sentence & HuggingFace), and deployed FastAPI (API backend), DVC (pipeline management), and Docker (containerization), ensuring scalability and modularity.

**Cloudy** (A heavy rainfall & cloudburst prediction system): [Research Paper]

- Designed and implemented a prediction pipeline using **LSTM**, **GRU**, **CNN-1D**, and **TFT** (Temporal Fusion Transformer) models with advanced data reprocessing (DateTime transformations, Temporal Cyclic & Seasonal features) on historical data.
- Applied **EEMD** (Ensemble Empirical Mode Decomposition) to decompose the original time series signal into **multiple IMFs** (Intrinsic Mode Functions).
- Integrated IOT sensors (ESP32 with DHT11 & beta rain sensor) to collect real-time weather data.
- Leveraged tools: PyTorch, TensorFlow, Keras, Scikit-learn, Pandas, Matplotlib and DVC for model and prediction pipeline management & Flask and Docker for model deployment.

## SKILLS

- Programming Languages:** Python, JavaScript.
- Object Oriented Programming, DBMS, Data Structures & Algorithms, Computer Networks.**
- AI:** Data Analytics, Visualizations and Preprocessing, Machine Learning and Deep Learning, Computer Vision.
- Frameworks & Libraries:** NumPy, Pandas, Matplotlib, Seaborn, Scipy, Scikit-Learn, Tensorflow, Keras, PyTorch, OpenCV, Flask, FastAPI, ReactJS, Tailwind CSS.
- Tools & Technologies:** MySQL, MongoDB, Version Control (Git, GitHub), MLOps (DVC, Docker), Microsoft Excel.
- Soft Skills:** Effective Communication, Time Management & Attention to Detail.

## ACADEMIC ACHIEVEMENTS

- Judged and Mentored BVP Hex Hackathon - Nov'24
- IBM Certified: Artificial Intelligence Analyst.
- Deeplearning.ai Certification by Andrew NG: Machine Learning Specialization in Supervised ML, Advanced Learning Algorithms, and Unsupervised ML.
- Top 10 Teams in HackMait-22 by MAIT and EmpowHer Hack by IEEE-IIITD. Additionally, Top 25 Vihan-2.0 by IEEE-DTU-23.
- Held leadership roles as Management Head in ECELL-ADGITM and Instrumental Head in Swaranjali-ADGITM.