

Vidit Parashar

✉ vparashar67@gmail.com | 📞 +91 9654124615

🐙 [GitHub](#) | 💼 [Linkedin](#)

Education

Manipal University Jaipur

2021 - 2025

B.Tech (Hons.) in Computer Science and Engineering (AI & ML)

Skills

Languages: Python, SQL

Data Engineering & Cloud: Azure Data Factory, Azure Databricks, Azure Synapse Analytics, Azure Data Lake Gen2, Apache Kafka, Apache Spark, Apache Airflow, dbt, Docker

ML & Tools: PySpark, Scikit-learn, Keras, LangChain, FinBERT, OpenAI, Librosa, YOLOv5

BI & Deployment: Power BI, Streamlit, FastAPI, Docker, Git, REST APIs

Work Experience

Abacus Insights, Mysore

Oct 2025 - Present

Data Engineering Intern

- Undergoing training currently, covered Python, SQL, Data Engineering, SDLC, Front-End, Soft-Skills and FHIR modules.
- Developed a Front-End for a client and hosted it on vercel. Test Scores till now: Python: 87.77%, SQL: 96.67%, Data Engineering: 93.33%, SDLC: 100%

Zee Media Corporation, Noida

Jan 2025 - Sept 2025

Software Engineering Intern

- Built an internal RAG chatbot (FastAPI, LangChain, document reranking, LLMs) projected to **reduce journalist lookup time by ~60%** based on pilot evaluations with internal documents.
- Trained a YouTube viewership predictor (RandomForest) using features from titles, audio signals, and visual density; model achieved $R^2 = 0.94$. **Increasing average viewership by 20%**
- Prototyped a real-time financial sentiment analyzer (FinBERT + scraping LiveMint) estimated to improve narrative detection for market analysis during trials.
- Deployed a content-based Movie Recommender (cosine similarity, 5K+ dataset, 5 features) as a Streamlit proof of concept for personalization experiments.

HCL Technologies, Noida

Jun 2024 - Aug 2024

Academic Trainee

- Prototyped a cloud-based supply chain analytics pipeline on Azure (ADF, Synapse, Power BI) using sample ERP-like datasets to model procurement and logistics KPIs.
- Dashboards and automated extracts were projected to reduce manual reporting effort by ~70% based on benchmarked manual workflows and test runs.
- Gained practical exposure to SAP data structures and enterprise data engineering patterns.

DRDO, Solid State Physics Lab, New Delhi

Mar 2023 - Jul 2023

AI/ML Intern

- Developed an ML-based anomaly detection prototype for sensor time-series data; model estimated a ~15% improvement in prediction accuracy over baseline methods on held-out validation sets.
- Built a lightweight Flask+Plotly web dashboard to visualize model outputs.

Project Work

- **Azure ETL Pipeline | Azure Data Factory, Databricks (PySpark), Synapse, Power BI:** Developed and deployed a Medallion-architecture ETL pipeline on Azure to ingest GitHub API data, perform transformations in Databricks, and curate datasets in Synapse for BI consumption. Delivered a scalable data solution ready for analysis in Power BI.
- **Fraud Detection System | Python, Scikit-learn, XGBoost, FastAPI, Docker, AWS:** Developed an end-to-end ML pipeline for proactive fraud detection on transactional data. Engineered domain features, mitigated class imbalance with SMOTE and scale-pos-weight, and optimized with XGBoost. Produced an inference service using FastAPI and Docker and validated via Postman. Deployed on AWS. **View Github**