

KARTIKAY TAUNK

Data Science Student · Python · Machine Learning · AI Systems · End-to-End ML Pipelines
kartikaytaunk@gmail.com +91 89558 59044 github.com/kartikay2005

Jaipur, Rajasthan

EDUCATION

Year	Degree / Board	Institute	Status
2023–Present	BS Data Science & Applications	IIT Madras (Online Degree Programme)	Pursuing
2023–Present	B.Tech – Computer Science Engineering	JECRC Foundation, Jaipur	Pursuing
2023	CBSE (XII)	St. Xavier's Sr. Sec. School, Jaipur	Completed

PROJECTS

Customer Churn Prediction & Retention Scoring | Python · XGBoost · SHAP · Flask 2025

- **Built an end-to-end churn prediction pipeline** on a 50,000-row telecom dataset — applied SMOTE to fix the 85:15 class imbalance that makes accuracy a meaningless metric, then compared Logistic Regression, Random Forest, and XGBoost via stratified 5-fold CV, selecting XGBoost (AUC-ROC 0.89, F1 0.84, Recall 0.88 on churners)
- **Engineered 25+ behavioural features** (tenure buckets, usage-trend slopes, support-ticket frequency) via Pandas and SQL window functions, lifting AUC-ROC from 0.71 → 0.89; used **SHAP waterfall plots** to explain top churn drivers to non-technical stakeholders, then deployed the scorer as a **Flask REST API** with JSON endpoints

Stock Market Share Analysis | Python · Pandas · Scikit-learn · Seaborn 2024

- **Cleaned and engineered 5+ years of OHLCV data** for Indian equities — built 15+ technical indicators (RSI, MACD, Bollinger Bands, EMA crossovers), trained a Random Forest classifier achieving ~72% directional accuracy on out-of-sample 6-month validation windows
- **Resolved class imbalance** via SMOTE (downside recall 41% → 63%), visualised sector-level return correlations via Seaborn heatmaps, and documented the full EDA-to-prediction workflow on GitHub with reproducible notebooks

AI Medical Assistant Platform | SIH 2024 | NLP · Python · Docker · PostgreSQL 2024

- **Integrated an NLP symptom-to-diagnosis classifier** on free-text patient input and a demand-forecasting module that flagged medicine stock-outs 48 hours ahead; **designed RBAC-secured PostgreSQL schema** with encrypted records and audit logging — directly mirrors production DS data governance requirements
- **Shipped via GitHub Actions + Docker** across three services (REST API, React PWA, React Native mobile), cutting release cycles to under 10 minutes — demonstrating reproducible ML deployment beyond notebook prototypes; selected for SIH 2024 national round from 75,000+ registrations

Inventory Demand Forecasting | BDM Capstone | Regression · Random Forest · Pandas 2024–25

- **Modelled 12 months of real business data** — invoices, batch logs, raw-material records for a Jaipur manufacturer; EDA surfaced overproduction and stockout patterns costing ~₹2L/month, resolved by a Random Forest forecaster that cut RMSE 18% over naive baseline
- **SHAP feature importance** identified top-3 leakage drivers (seasonal demand spikes, supplier lead-time variance, SKU proliferation) and translated model outputs into concrete procurement decisions 3 weeks ahead of cycle — a real business outcome, not a Kaggle score

IIT MADRAS BS DATA SCIENCE — RELEVANT COURSEWORK

- **Completed:** Machine Learning Foundations · ML Techniques · ML Practice · Tools in Data Science · Business Data Management · Business Analytics
- **Currently pursuing:** Database Management Systems · Intro to Deep Learning & Generative AI · Programming, Data Structures & Algorithms (Python) · Modern Application Development II · System Commands

TECHNICAL SKILLS & TOOLS

Languages	Python, SQL	ML / AI	Scikit-learn, XGBoost, Random Forest, Logistic Regression, NLP, SHAP, SMOTE, Feature Engineering, Model Evaluation (AUC-ROC, F1, RMSE)
Data & EDA	Pandas, NumPy, Matplotlib, Seaborn, Hypothesis Testing, Statistics, EDA	Dev & Tools	Git, Docker, GitHub Actions, Flask REST APIs, Jupyter; Salesforce CRM (basics), Cloud Computing (basics)

· Dual Degree: B.Tech CSE + IIT Madras BS Data Science · All projects live on GitHub