Kunal Pathak

Mumbai, Maharashtra, India | +91-8108375850 | kunalpt2004@gmail.com
linkedin.com/in/kunal-pathak30 | github.com/Kunalpathak2004 | kunalpathak-portfolio.vercel.app

PROFESSIONAL SUMMARY

Data Scientist with expertise in Python, SQL, and machine learning, specializing in fraud detection systems and predictive analytics. Delivered 20 percent cost reduction through automated ML pipelines and 30 percent improvement in forecast accuracy across retail and financial sectors. Track record includes processing 500K plus daily transactions, preventing 2M dollars in fraud losses, and publishing peer-reviewed IEEE research. National hackathon qualifier with demonstrated ability to translate complex data into actionable business insights.

TECHNICAL SKILLS

Programming: Python, SQL, R, MySQL
Machine Learning: Scikit-learn, TensorFlow, Predictive Modeling, Classification, Regression, Clustering
Data Analysis: Pandas, NumPy, SciPy, Statistical Analysis, A/B Testing, Hypothesis Testing, EDA
Data Visualization: Power BI, Tableau, Matplotlib, Seaborn, Excel, Dashboards
Data Engineering: ETL Pipelines, Data Cleaning, Feature Engineering, Data Preprocessing
Tools: Git, Jupyter, Streamlit, VS Code

EDUCATION

Lokmanya Tilak College of Engineering — Navi Mumbai, India
Bachelor of Technology in Computer Science (Data Science Specialization) — *Nov 2022 – Present*
- Relevant Coursework: Machine Learning, Statistical Analysis, Data Structures, Algorithms, Database Systems

PROFESSIONAL EXPERIENCE

Data Scientist – Fraud Detection System | *Python, Scikit-Learn, Pandas*
- Architected machine learning pipeline achieving 95 percent accuracy on 500K financial transactions daily
- Cut manual transaction review workload by 20 percent, recovering 200 hours monthly for compliance teams
- Engineered 50 plus behavioral and transactional features identifying suspicious payment patterns
- Deployed Flask API with sub-100ms latency, preventing 2M dollars in fraudulent transactions
- Optimized ensemble models (Random Forest, XGBoost, Gradient Boosting) reducing false positives by 35 percent

Data Analyst – Retail Sales Analytics | *Python, SQL, Power BI, ETL*
- Developed ETL pipeline processing 100K daily retail transactions from 5 regional data sources
- Increased sales forecast accuracy by 30 percent applying ARIMA and Prophet time series models
- Automated executive reporting with Power BI, reducing dashboard generation time from 5 hours to 3 hours
- Executed K-means clustering on customer base, uncovering segments driving 60 percent of total revenue
- Performed regression analysis correlating promotional strategies to 18 percent increase in product margins
- Built executive KPI dashboards displaying revenue, regional sales, and inventory turnover metrics

Data Engineer – Invalid Traffic Analysis | *Python, Pandas, Streamlit, Analytics*
- Created real-time detection system analyzing 1M daily ad impressions, raising fraud identification by 25 percent
- Developed 30 plus analytical features via correlation matrices and time series decomposition
- Deployed Streamlit web application allowing fraud teams to upload datasets and detect threats in real-time
- Programmed automated EDA framework revealing geographic, temporal, and device-based traffic anomalies
- Authored technical documentation detailing feature methodology for future predictive model iterations

ACHIEVEMENTS AND LEADERSHIP

IEEE Research Publication 2024: Published peer-reviewed research paper on data analysis methodologies in IEEE conference proceedings

Smart India Hackathon – Team Lead 2023: Managed 6-person technical team building ML solution, qualifying for national finals among 10,000 plus teams

Orchestrated technical architecture, model training, and final pitch delivery within 36-hour time constraint

CERTIFICATIONS

Microsoft Data Analyst | Microsoft Generative AI | Microsoft Responsible AI | Microsoft Power BI