

Aakshat Agarwal

Data Analyst

Jaipur, Rajasthan, India | +91 79768 07739 | aakshatarawal2004@gmail.com
linkedin.com/in/aakshatarawal | github.com/Aaksh26at | portfolio-seven-lemon-63.vercel.app

PROFESSIONAL SUMMARY

Data analyst with hands-on experience turning raw data into reports and decisions using Python, SQL and Excel. Built data pipelines, dashboards and machine-learning models across academic and work projects. Seeking a data analyst role where I can deliver clear, actionable insights.

TECHNICAL SKILLS

Languages & Tools: Python, SQL, Advanced Excel, Google Sheets, Git, GitHub, Jupyter, VS Code

Data Analysis: Pandas, NumPy, Exploratory Data Analysis, Data Cleaning, Feature Engineering, Statistical Analysis

Visualization & BI: Power BI, Matplotlib, Seaborn, Streamlit, KPI & Dashboard Reporting

Machine Learning: Scikit-learn, Regression, Classification, Model Evaluation, Cross-Validation, Threshold Tuning

Databases & Backend: MySQL, PostgreSQL, SQL Query Optimization, Flask, REST APIs

Web Scraping: Requests, BeautifulSoup, Selenium, OCR-based PDF extraction (ddddocr)

EXPERIENCE

Data / CRM Executive — Kalash Group, Jaipur

May 2026 – Present

- Built and maintained CRM lead-tracking reports in Excel, consolidating scattered lead data into structured trackers for sales follow-up and reporting.
- Developed a Python automation tool to retrieve and digitize electricity bill data, cutting a manual, repetitive task down to an automated pipeline that outputs clean Excel records.

PROJECTS

Electricity Bill Data Scraper & Extractor | *Python, Selenium, pdfplumber, dddddocr, Pandas*

- Built an end-to-end automation that logs into a utility portal by consumer (K) number, downloads PDF bills, and extracts billing data into a consolidated Excel dataset.
- Used OCR (ddddocr) to automate the login step and pdfplumber to parse three different bill layouts, removing hours of manual data entry.

Breast Mass Classification — Wisconsin Diagnostic | *Python, Scikit-learn, Streamlit*

- Classified breast masses from 30 cytology measurements, selecting an RBF-kernel SVM on malignant recall — $97.07\% \pm 3.27$ against a majority-class baseline that catches none — over 50 fits of repeated stratified cross-validation.
- Moved the decision threshold off the 0.5 default to 0.225 by costing a missed malignancy at $10\times$ a false alarm, trading nine extra false positives for two more cancers caught, and shipped the threshold inside the model artifact.

Customer Churn Prediction & Retention Dashboard | *Python, Scikit-learn, Pandas, Streamlit*

- Built an ML pipeline on a 7,043-customer telecom dataset reaching 0.845 ROC-AUC and catching 80.4% of churners, against a never-churn baseline of 0% recall.
- Reported the operating cost honestly — 1,501 churners caught against 1,422 false alarms — and shipped a Streamlit app translating predictions into retention actions.

Fire Risk Classification from Weather | *Python, Scikit-learn, Flask*

- Classified fire occurrence in Algerian forest data from weather variables alone, reaching 0.938 ROC-AUC against a 0.500 baseline.
- Identified and removed the Fire Weather Index features that were reproducing the label, and validated with GroupKFold over region-month to prevent correlated days leaking across folds.

EDUCATION

Master of Computer Applications (MCA), Computer Science — University of Rajasthan, Jaipur (2024–2026)

Bachelor of Computer Applications (BCA), Computer Science — University of Rajasthan, Jaipur (2021–2024)

CERTIFICATIONS

- AWS Solutions Architecture — Job Simulation, Forage (2024)
- Data Analytics — Job Simulation, Deloitte Australia via Forage (2024)
- Software Engineering — Job Simulation, JPMorgan Chase & Co. via Forage (2024)
- Data Science 101 Certification (2023)