

Rakshit Kumar

jrakshitk35@gmail.com — 07795613146 — Bangalore, 560100

Portfolio: rakshitstatuswindow.netlify.app — linkedin.com/in/jrakshitkumar — github.com/Rakshitk2223

Summary

Data Analyst with experience in the end-to-end data analysis lifecycle, from data wrangling and statistical modeling to data visualization and stakeholder reporting. Proficient in Python, R, and SQL, with a proven ability to translate complex datasets into actionable business insights and KPIs for data-driven decision-making.

Technical Skills

Languages: Python, R, SQL, Java, Git

Libraries & Frameworks: Pandas, NumPy, Matplotlib, Scikit-learn, TensorFlow, XGBoost

Databases & Tools: Power BI, Jupyter Notebook, VS Code

Methodologies: Statistical Analysis, Predictive Modeling, Data Cleaning

Projects

Amazon Sales Data Analysis

Python, Pandas, Seaborn, Matplotlib

- Analyzed a dataset of 1,465 products and over 26.8 million reviews to identify key market segments, discovering that electronics (led by USB cables at 15.9%) dominate over 60% of the product portfolio.
- Engineered features and applied statistical analysis, identifying a strong 0.962 correlation between actual and discounted prices and detecting that 14.81% of products were price outliers.
- Developed pricing and rating analytics, revealing an average discount of 47.7% and identifying top-performing products like a USB-C cable with a perfect 5.0 rating from over 18,000 reviews.

Credit Score Classification

Python, Scikit-learn, Pandas, XGBoost

- Processed and cleaned a financial dataset of 100,000 credit records with 28 features, preparing a high-quality dataset for robust model training and evaluation.
- Benchmarked 5 classification models, selecting Random Forest as the optimal performer by achieving 81.50% accuracy, 80.50% precision, and 80.95% recall.
- Optimized the final model by analyzing feature importance to identify the most influential predictors of credit risk, delivering a robust tool for data-driven loan assessment.

Early Prediction of Chronic Kidney Disease (CKD)

Python, XGBoost, Scikit-learn

- Engineered and tuned an XGBoost classifier, achieving 93.33% accuracy and a critical 94.44% recall, significantly minimizing false negatives for early-stage disease detection.
- Conducted a comparative analysis of 4 machine learning models, including Random Forest and Logistic Regression, to scientifically validate the selection of XGBoost as the most effective predictor.
- Implemented a robust data pipeline using Pandas and Imblearn to cleanse data and resolve class imbalance, ensuring high-quality inputs for reliable and precise model performance.

Education

Vellore Institute of Technology - AP University

Bachelors of Technology — Cumulative GPA: 8.13/10.0

2026

St. Francis De Sales Public School, Bengaluru

Senior Secondary [XII] — 92%

Secondary School [X] — 95%

2022

2020

Leadership & Extracurricular Activities

GeeksForGeeks Student Chapter, VIT-AP

Engaging with a community focused on competitive programming and technical skill development.

Member (2025 - Present)

Melange Cultural Club, VIT-AP

Developed and maintained a comprehensive database for 2 major club events and member participation, increasing organizational efficiency by an estimated 30%.

Documentation Team Member (

Certifications & Professional Development

Certifications: Oracle Cloud Infrastructure 2025 Certified Generative AI Professional; Artificial Intelligence using Google TensorFlow; Data Analysis and Programming with R.

Virtual Experience Programs (Forage): British Airways (Data Analyst); Deloitte (Data Analytics).

Languages

English (Professional) — Hindi (Native) — French (Beginner)