Rahul Singh

Pune, India
+91-8809109905 — singh205184@gmail.com — GitHub — LinkedIn — Portfolio

Profile

AI & Data Science undergraduate focused on building end-to-end machine learning systems and data driven applications. Experienced in EDA, feature engineering, model evaluation, and deploying scalable ML pipelines for real world business problems.

Education

Dr. D. Y. Patil Institute of Engineering, Management & Research (DYPIEMR), Pune
B.E. in Artificial Intelligence & Data Science 2023 – 2027
CGPA : 8.55 / 10

Technical Skills

Languages : Python, C++, JavaScript
ML / Data : Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn
Tools : TensorFlow, FastAPI, Streamlit, Git, GitHub, Jupyter Notebook
Core : Machine Learning, Data Science, EDA, Feature Engineering, Pipelines, Model Evaluation

Experience

Machine Learning Intern — CodeAlpha Jan 2026 – Feb 2026

- Built ML models for Credit Scoring (multi-class) and Breast Cancer prediction (binary) using real-world datasets.
- Performed data preprocessing, feature engineering, and trained models (Logistic Regression, Random Forest, SVM).
- Evaluated performance using Accuracy, F1-score, and ROC-AUC, achieving ~**98% accuracy**.
- Worked with Python, Scikit-learn, and end-to-end ML workflows.

Projects

Customer Churn Prediction System (End-to-End ML, Deployed) 2026

- Built end-to-end churn prediction system on Telco data (7k+ rows, 26 features) using pipelines and ColumnTransformer, reduced 6000+ features to 48.
- Merged multi-source datasets, removed data leakage features, fixed incorrect data types, and handled missing values using **Scikit-learn Pipelines** and **ColumnTransformer**.
- Trained Logistic Regression, Random Forest, Gradient Boosting; selected tuned Logistic Regression (class_weight=balanced) with **ROC-AUC ≈ 0.84**.
- Interpreted model to identify key churn drivers and deployed full pipeline as a **Streamlit web app** with probability and risk explanation.

IPL 50+ Score Prediction System (EDA → ML → Deployment) 2026

- Converted IPL batting dataset (15k+ rows) into ML problem; performed EDA, feature engineering, and data leakage prevention.
- Trained Logistic Regression, Random Forest, Gradient Boosting; final Gradient Boosting achieved **F1 ≈ 0.80, ROC-AUC ≈ 0.98**.
- Built pipeline-based training workflow and deployed using **Streamlit** for real-time predictions.

Drone Surveillance System for Anomaly Detection (AI Hackathon) 2025

- Built AI-based drone monitoring system for fire, flood, and anomaly detection using ML.
- Developed a simulation and real-time analytics pipeline using FastAPI and PyBullet.

Achievements

- Qualified Round 2 (Top 120 out of 240+ teams) in Inno-Hackathon VIT 2025.
- Selected at internal college-level round of Smart India Hackathon (SIH) for AI-based drone surveillance project (VIRA).
- Built and deployed multiple end-to-end machine learning systems involving data preprocessing, feature engineering, model training, evaluation, and interpretation.