Adham Mahmoud

professional

Student Performance Prediction

Machine-Learning Project

The problem

- **Many students are at risk of underperforming or dropping out due to low academic results, absenteeism, or lack of support.**

- **Teachers and schools often lack early predictive tools to identify these students, leading to delayed interventions.**

Dataset Overview

Source: UCI Student Performance Dataset (650 records, Portugal)

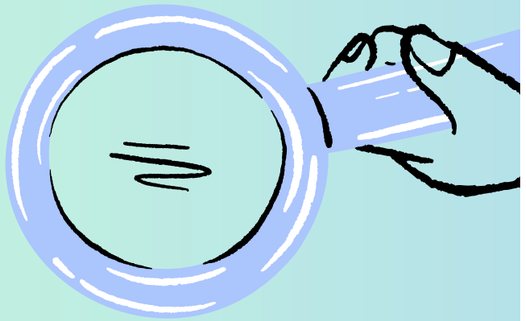
Demographics → school, sex, age, address, famsize, parents' education/jobs

Study & Support → studytime, failures, schoolsup, famsup, paid, higher, internet

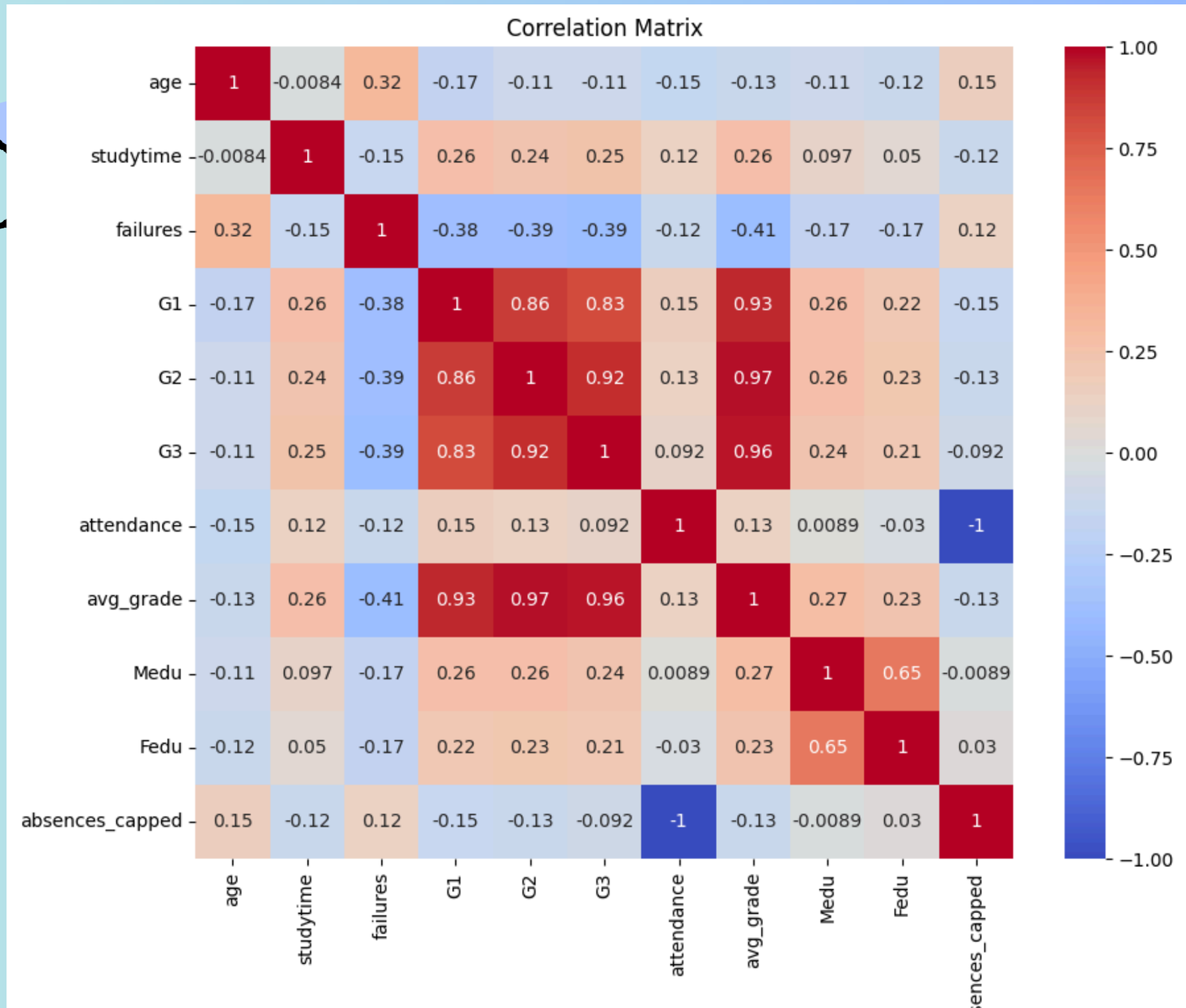
Lifestyle & Behavior → goout, freetime, Dalc, Walc, health, romantic

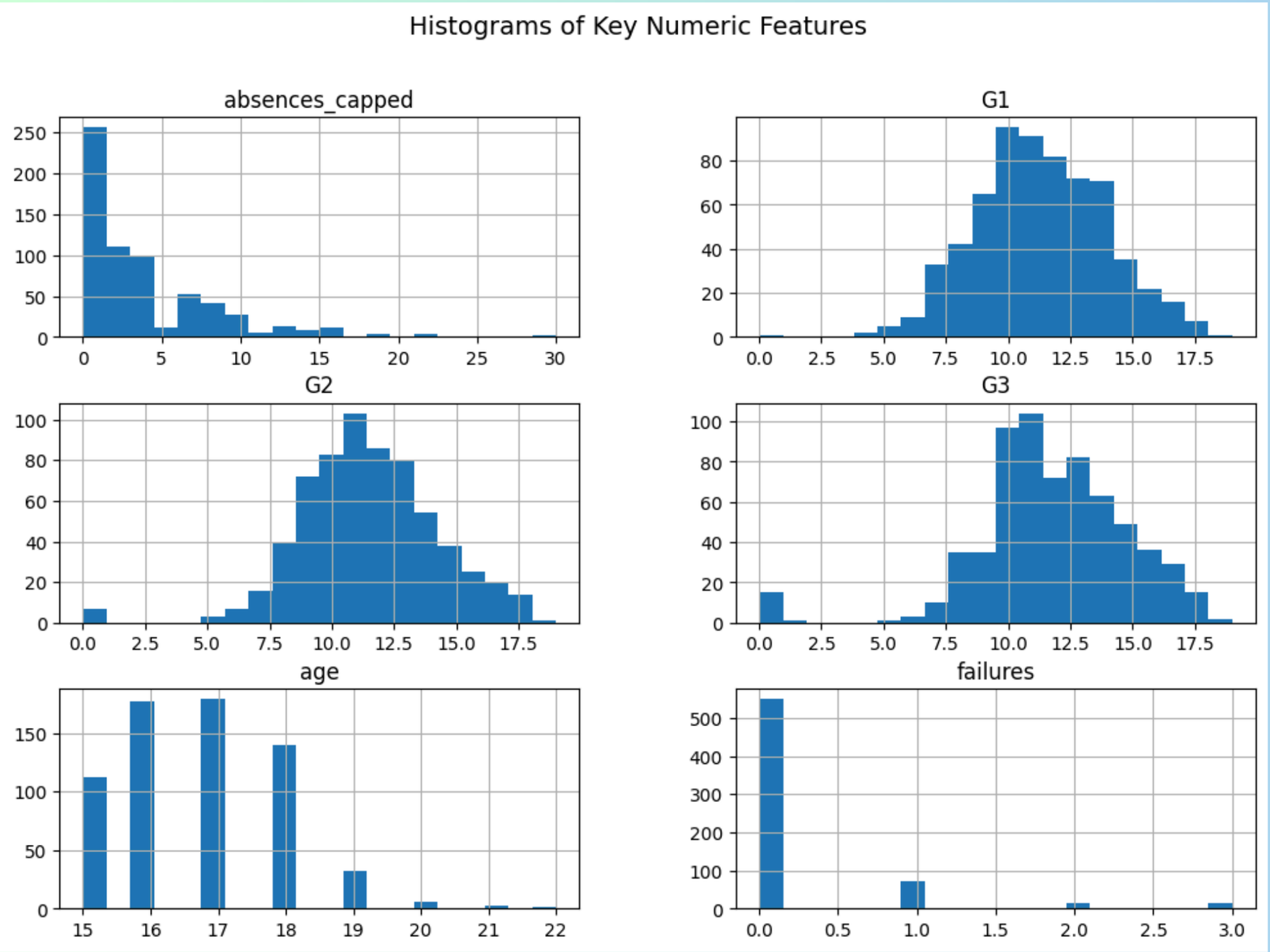
Performance Indicators → G1, G2, G3 (grades), absences

Exploratory Insights



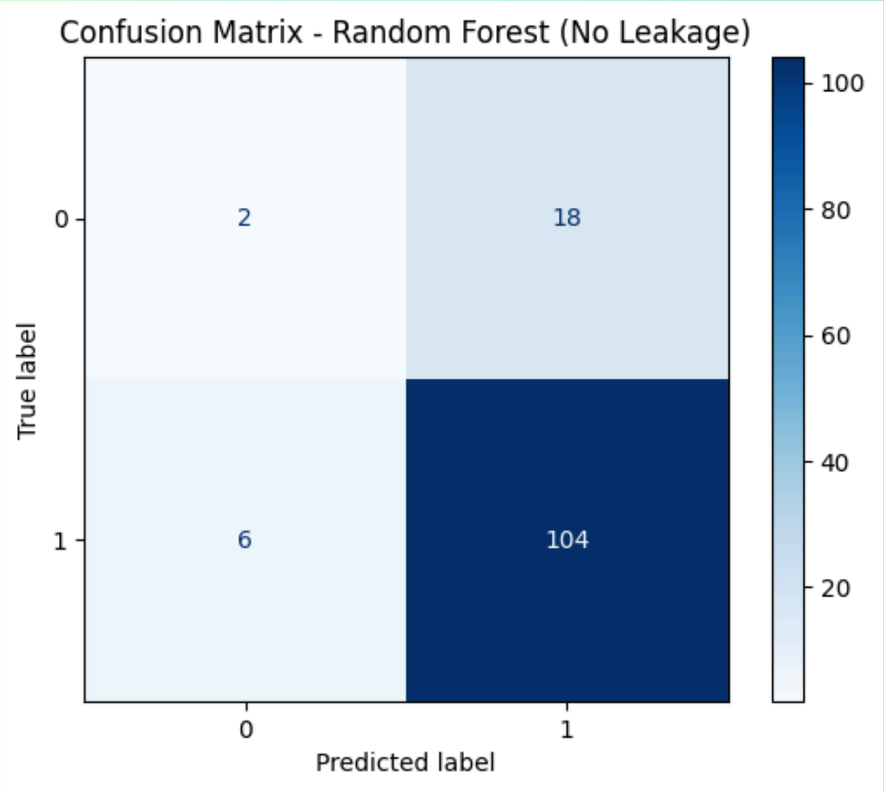
Final grade (G3) is strongly driven by early grades (G1, G2, avg_grade), while failures and higher absences negatively affect performance





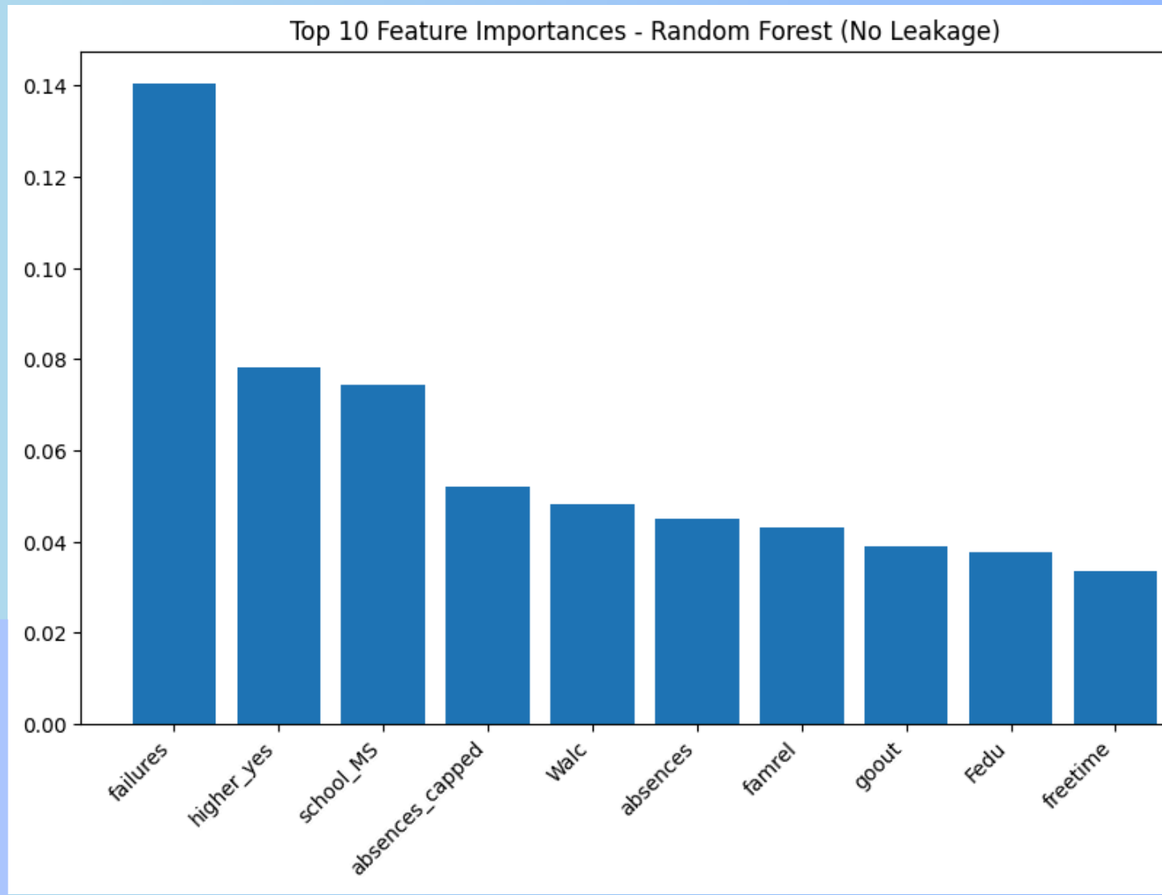
Most students are 15–18 years old with 0–1 failures and low absences. Grades (G1, G2, G3) cluster around passing, but a smaller group shows clear risk of failure.

Model Evaluation – Random Forest



Model achieves strong performance without leakage

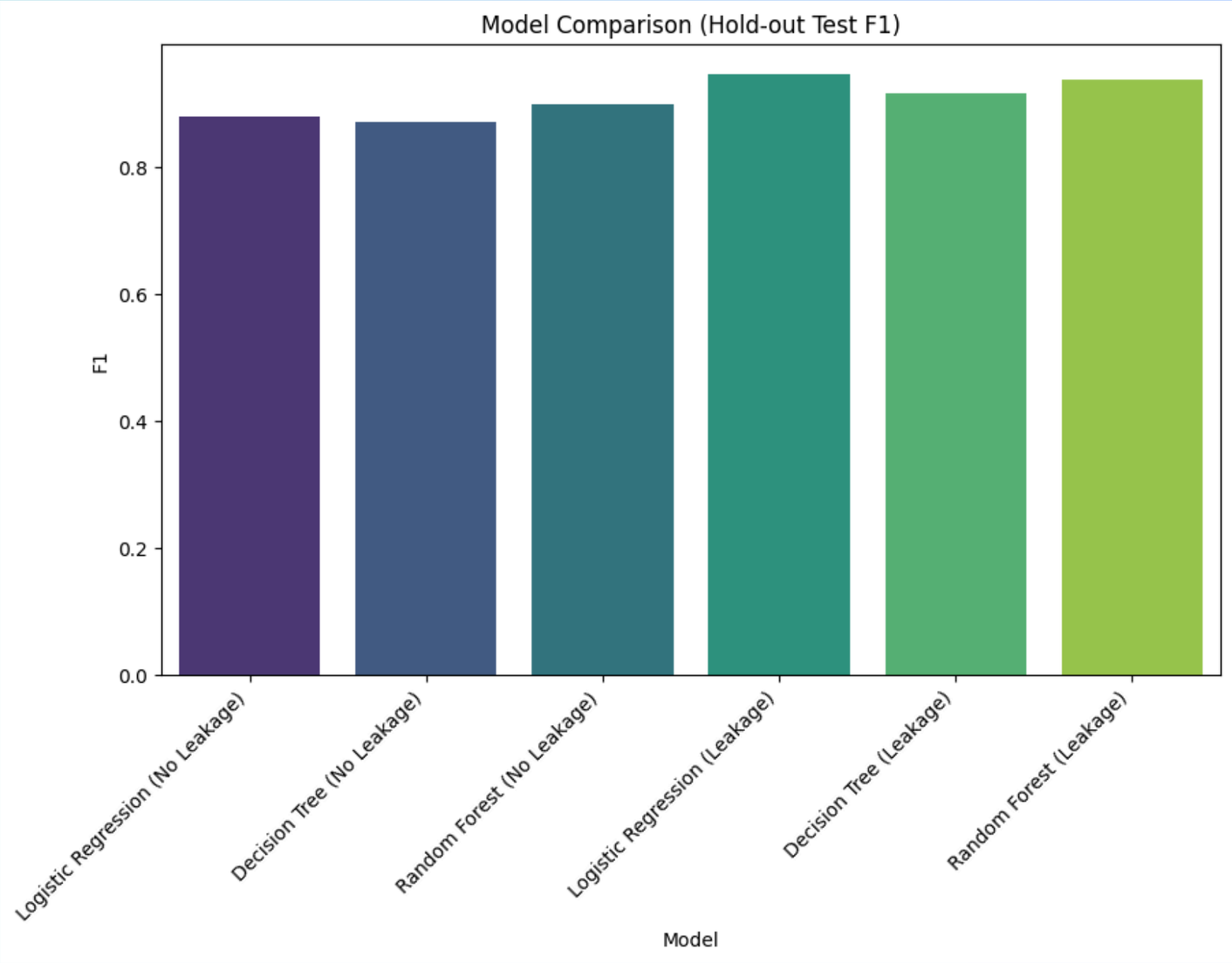
Top 10 Features



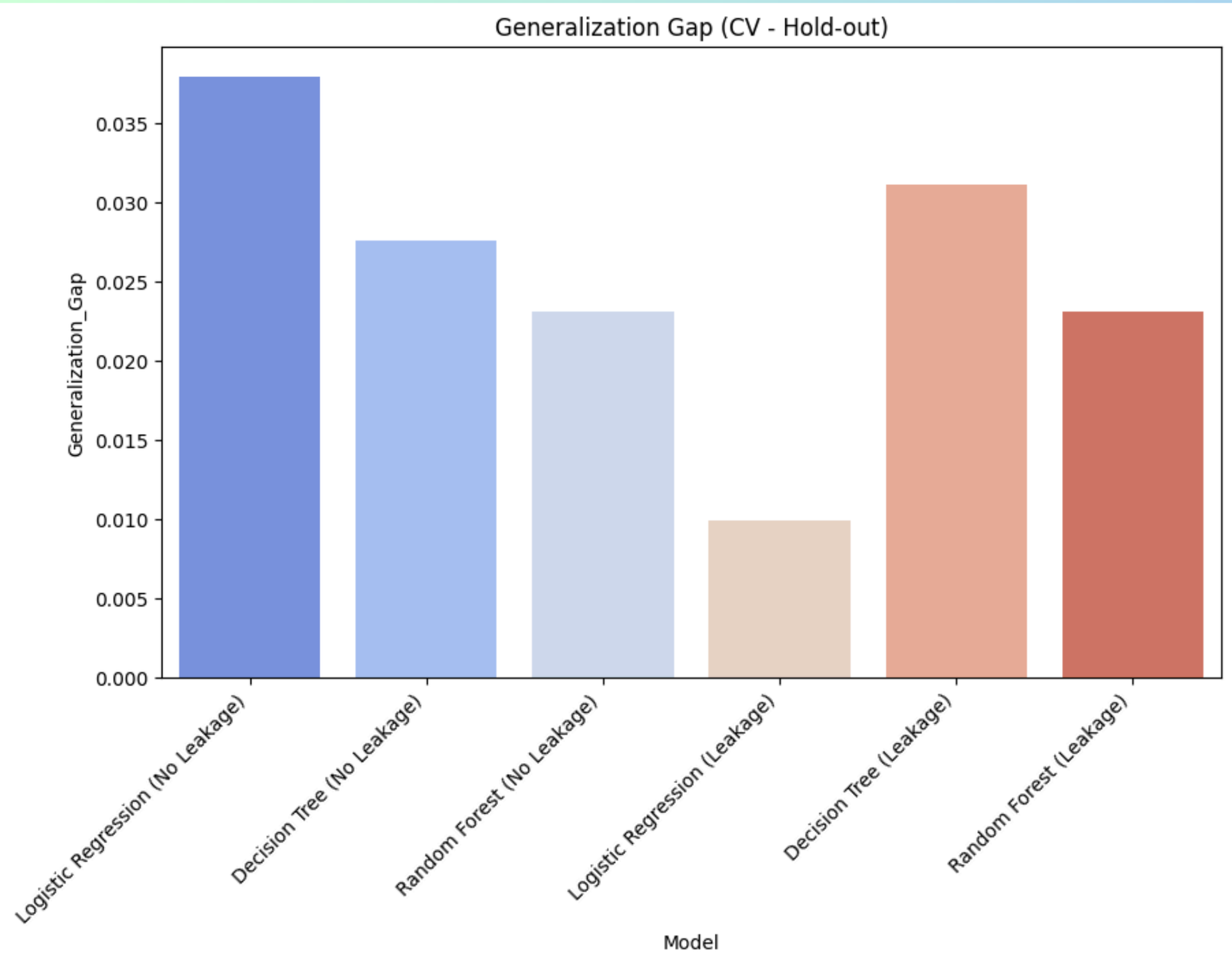
Study time and family support are key drivers of success, highlighting the value of consistent habits and parental encouragement.

Model Comparison (F1 Score)

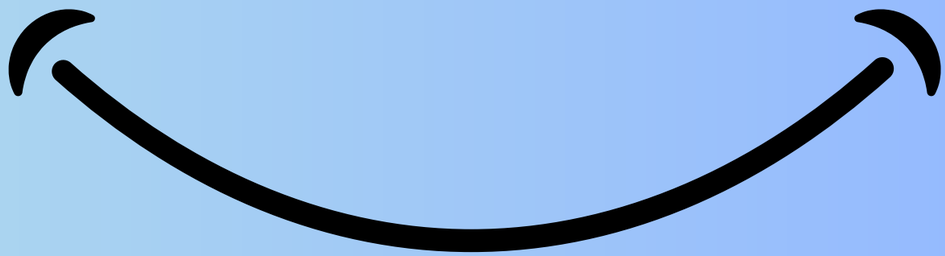
Random Forest consistently outperforms Logistic Regression & Decision Tree. Balances accuracy, recall, and F1 effectively.



Generalization & Robustness



Small gap between CV and holdout scores.
Confirms that the model generalizes well and avoids overfitting.



Family involvement and consistent support are crucial for student success — let’s invest in our kids’ future.

Thank You

click For more

click for more