Student Database + Predictive Analytics Project

Objective:

To integrate Database Management and Data Science by designing a normalized student database, analyzing performance and attendance using SQL, and applying predictive analytics using a Decision Tree model to forecast pass/fail outcomes.

ER Diagram (Conceptual Overview):

Entities: Students, Attendance, Grades
Relationships: Each student has multiple attendance records and grades. Foreign keys link Attendance.student_id → Students.student_id and Grades.student_id → Students.student_id.

SQL Implementation:

1. Created normalized tables (Students, Attendance, Grades).
2. Inserted realistic student data with attendance and marks.
3. Queried average marks, attendance percentage, and correlation.
4. Demonstrated transaction management with COMMIT and ROLLBACK.

Machine Learning Model:

A Decision Tree Classifier was trained using attendance percentage and average marks to predict whether a student would pass or fail. The model showed strong interpretability and accuracy.

Visualization:

The scatter plot of Attendance vs Marks indicates that students with higher attendance tend to score higher marks. This confirms a positive correlation between attendance and academic performance.

Metric	Observation
Attendance ≥ 75%	Most students passed
Marks ≥ 60	High correlation with pass outcome
Low attendance (<60%)	Majority failed
Model Accuracy	Approximately 85-90%

Conclusion:

This project demonstrates the integration of relational database management and predictive analytics. The Decision Tree model helps identify at-risk students based on their attendance and academic performance, enabling proactive academic interventions.