

Analyzing 'Student Performance' Dataset

< Micah Cheng >

}

Introduction

Attendance 0.5802

Hours_Studied 0.4451

Previous_Scores 0.1742

Access_to_Resources 0.1678

Tutoring_Sessions 0.1568

Parental_Involvement 0.1560

Parental_Education_Level 0.1052

Peer_Influence 0.0991

Family_Income 0.0945

Motivation_Level 0.0885

Distance_from_Home 0.0880

Teacher_Quality 0.0751

Extra_Activities_Yes 0.0630

Internet_Access_Yes 0.0511

Physical_Activity 0.0251

Gender_Male -0.0049

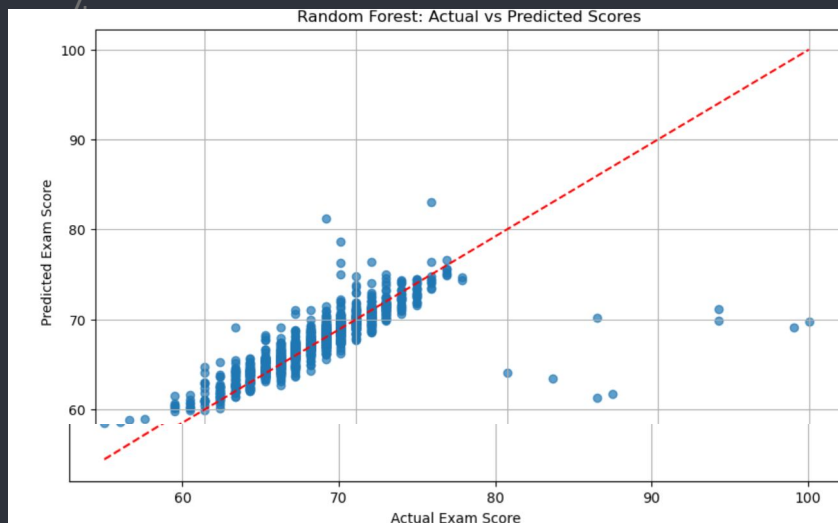
School_Type_Public -0.0108

Sleep_Hours -0.0171

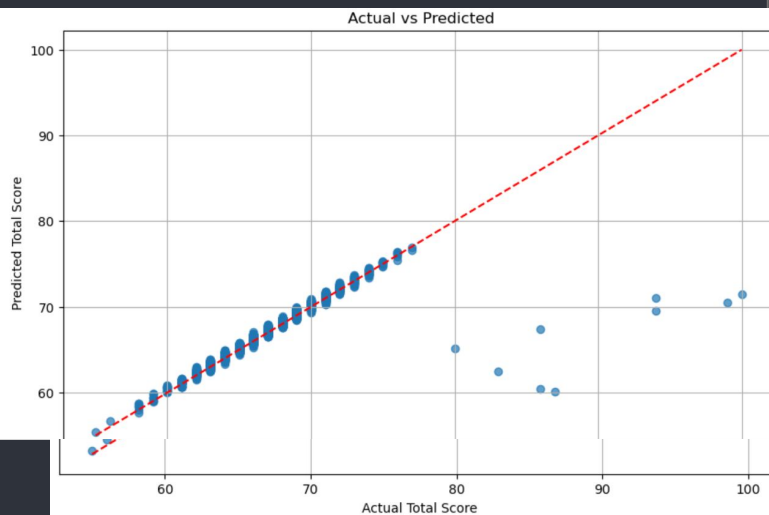
Learning_Disabilities_Yes -0.0839

	Hours_Studied	Attendance	Parental_Involvement	Access_to_Resources
0	23	84	1	3
1	19	64	1	2
2	24	98	2	2
3	29	89	1	2
4	19	92	2	2
	Sleep_Hours	Previous_Scores	Motivation_Level	Tutoring_Sessions
0	7	73	1	0
1	8	59	1	2
2	7	91	2	2
3	8	98	2	1
4	6	65	2	3
	Family_Income	Teacher_Quality	Peer_Influence	Physical_Activity
0	1	2	3	3
1	2	2	1	4
2	2	2	2	4
3	2	2	1	4
4	2	3	2	4
	Parental_Education_Level	Distance_from_Home	Exam_Score	\
0	1	3	67	
1	2	2	61	
2	3	3	74	
3	1	2	71	
4	2	3	70	
	School_Type_Public	Extracurricular_Activities_Yes	Internet_Access_Yes	
0	1	0	1	
1	1	0	1	
2	1	1	1	
3	1	1	1	
4	1	1	1	
	Learning_Disabilities_Yes	Gender_Male	Pass_Fail_N	Pass_Fail
0	0	1	0	Fail
1	0	0	0	Fail
2	0	1	1	Pass
3	0	1	1	Pass
4	0	0	1	Pass

1 Can we build a predictive model that
2 accurately estimates student scores based on
3 various factors?" {



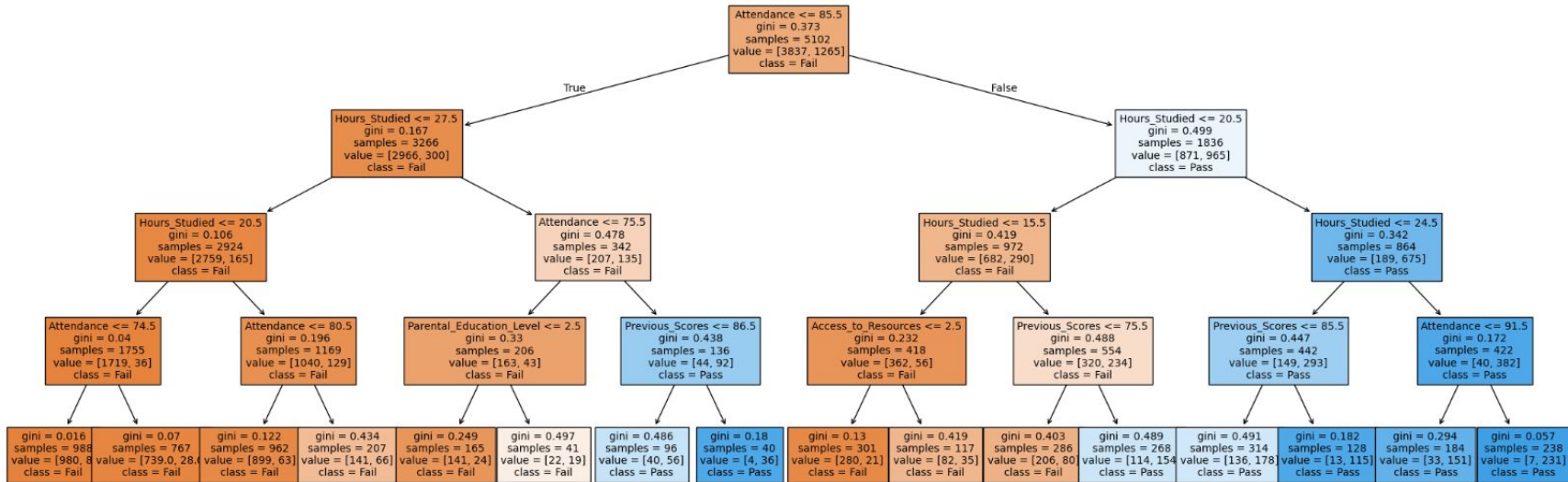
1 / MSE: 5.82 R^2 : 0.63



MSE: 4.15 R^2 : 0.73

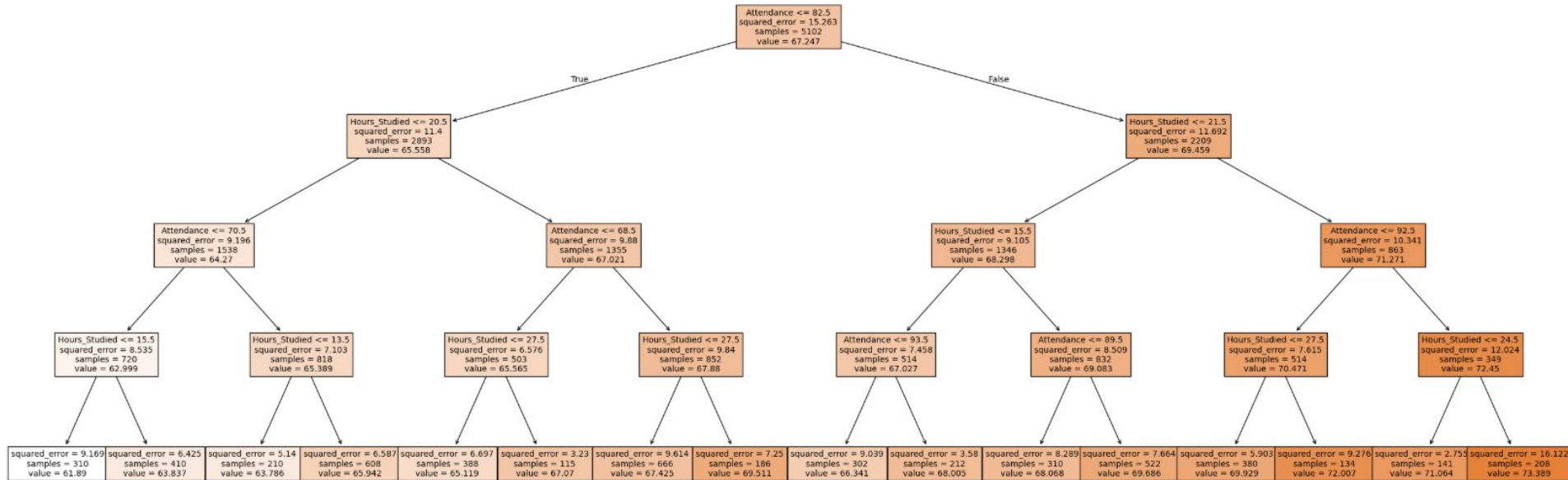
Decision Tress 'Lightbox';

Pass/Fail Decision Tree (Accuracy=0.85)



1 Decision Tress 'Lightbox';

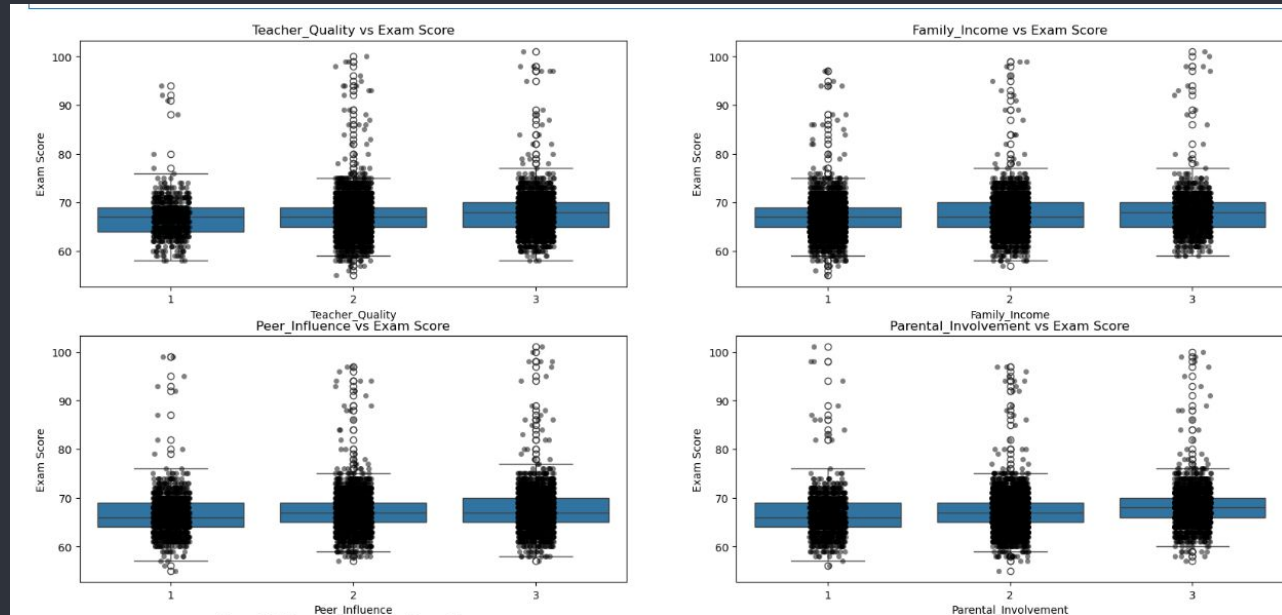
Exam Score Prediction (MSE=7.712363312804017, R²=0.5036873661329333)



Regression MSE: 7.712363312804017, R²: 0.5036873661329333

Q1 { How much can other people influence student's final grade? Can I predict scores from it?

< Features: Teacher Quality, Peer Influence, Parent Involvement >



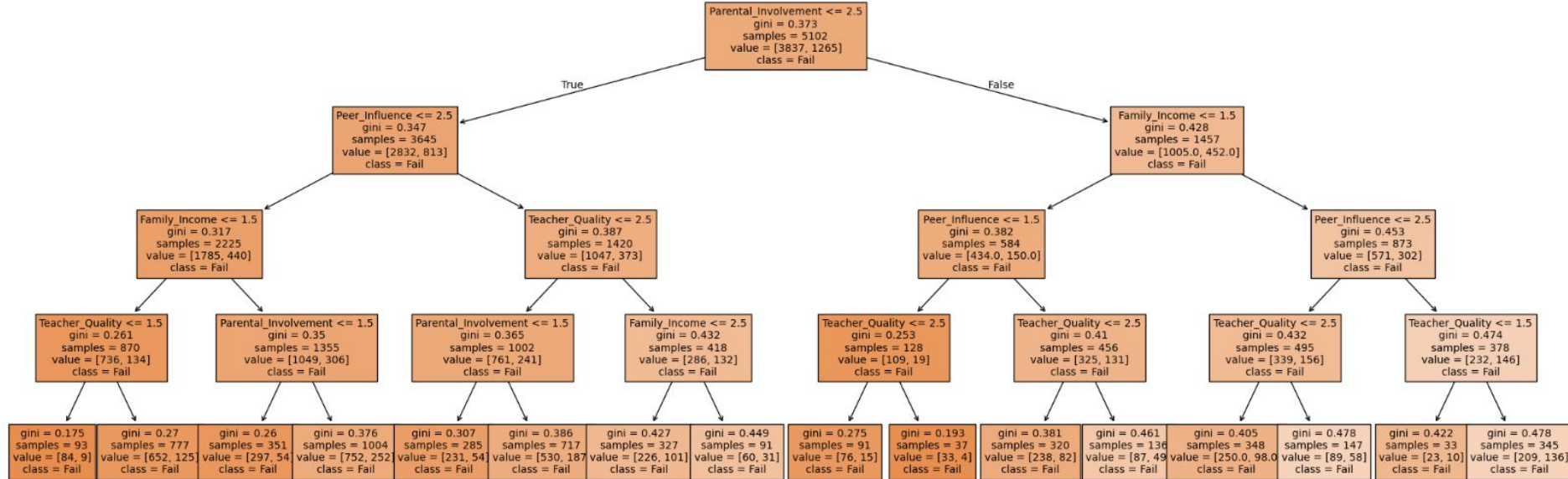
OLS Regression Results

	Feature	Coefficient	P-value
const	Intercept	61.560137	0.000000e+00
Teacher_Quality	Teacher_Quality	0.482052	1.202499e-09
Family_Income	Family_Income	0.482396	5.470385e-14
Peer_Influence	Peer_Influence	0.488966	9.052567e-15
Parental_Involvement	Parental_Involvement	0.860713	8.355593e-36
Parental_Education_Level	Parental_Education_Level	0.529844	4.460957e-18

(<statsmodels.regression.linear_model.RegressionResultsWrapper at 0x1e76701e270>,
Feature Coefficient P-value
const Intercept 61.560137 0.000000e+00
Teacher_Quality Teacher_Quality 0.482052 1.202499e-09
Family_Income Family_Income 0.482396 5.470385e-14
Peer_Influence Peer_Influence 0.488966 9.052567e-15
Parental_Involvement Parental_Involvement 0.860713 8.355593e-36
Parental_Education_Level Parental_Education_Level 0.529844 4.460957e-18)

Accuracy: 0.75

Pass/Fail Decision Tree (Accuracy: 0.75)



Q3: Finding the combination of input feature values that would maximize the predicted exam score

OPTIMAL VALUES FOR MAXIMUM EXAM SCORES:

```
Hours_Studied: 30.0 (Avg score: 70.91184573002755)
Attendance: 98.0 (Avg score: 70.8108108108108)
Parental_Involvement: 3 (Avg score: 68.11220043572985)
Access_to_Resources: 3 (Avg score: 68.10315789473684)
Sleep_Hours: 4 (Avg score: 67.69256756756756)
Previous_Scores: 97.0 (Avg score: 68.13519813519814)
Motivation_Level: 3 (Avg score: 67.74393108848865)
Tutoring_Sessions: 6 (Avg score: 71.66666666666667)
Family_Income: 3 (Avg score: 67.84146341463415)
Teacher_Quality: 3 (Avg score: 67.68188976377952)
Peer_Influence: 3 (Avg score: 67.63650607128868)
Physical_Activity: 6 (Avg score: 68.19354838709677)
Parental_Education_Level: 3 (Avg score: 67.97265625)
Distance_from_Home: 3 (Avg score: 67.51381215469613)
School_Type_Public: False (Avg score: 67.31635802469135)
Extracurricular_Activities_Yes: True (Avg score: 67.45495140530602)
Internet_Access_Yes: True (Avg score: 67.30951976921772)
Learning_Disabilities_Yes: False (Avg score: 67.3644483362522)
Gender_Male: 0 (Avg score: 67.27472118959108)
Pass_Fail_N: 1 (Avg score: 72.00822264389627)
Pass_Fail: Pass (Avg score: 72.00822264389627)
```