

# Predicting Success for Computer Science Students

Sulav Malla, Jing Wang, William Hendrix, Ken Christensen  
 sulavmalla@mail.usf.edu, jw@mail.usf.edu, whendrix@usf.edu, christen@cse.usf.edu  
 Department of Computer Science and Engineering



## Introduction

Admission into the Computer Science and Computer Engineering programs at USF requires:

- A minimum AdmitGPA of 3.1 in six entry-level courses
- A minimum of 'B' in the first programming course (CS1)

We analyze the grades in CS1 and six entry-level courses (Calculus 1 & 2, Physics 1 & 2, and English Compositions 1 & 2), to determine predictors for success ( $\geq$ 'B') in CS2.

## Methodology

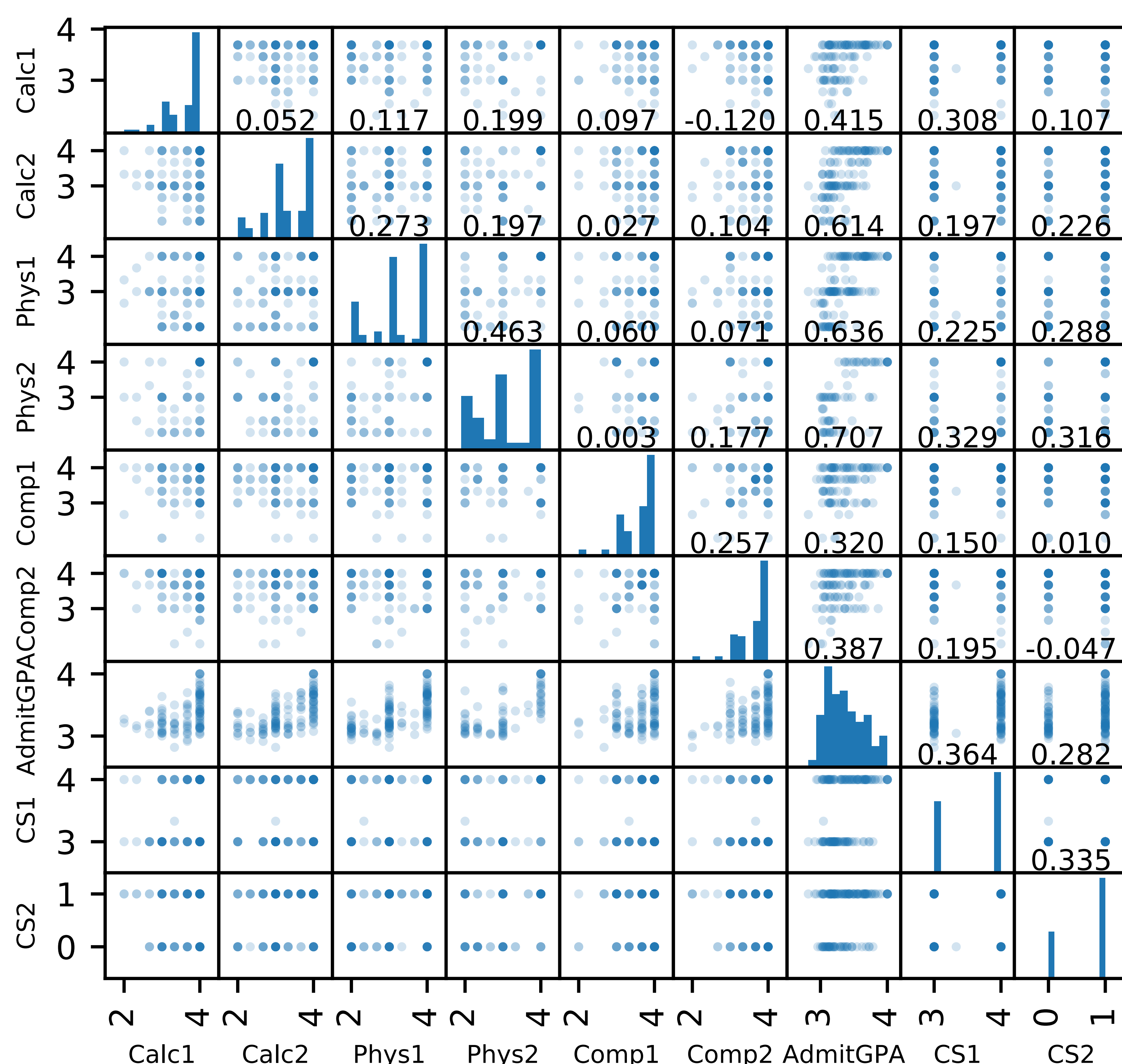
- **Dataset:** gender, major, department admission status, grades in six entry-level courses, grades in CS1, and grades in CS2 of 148 students who took the CS2 course in the semester under study.
- We use readily available off-the-shelf statistical and data mining tools in Python and WEKA for generating summary statistics, calculating correlations, testing statistical significance, and creating decision trees.

## Findings

- Grade in CS1 is the most important predictor of a student's success in CS2, as found previously [1, 2].
- Grades in Physics and Calculus were more indicative of success in CS2 while the two English Composition courses showed no effect on CS2.
- All six students who had a 'B' in CS1 and a 'C' in Calc2, failed in CS2.

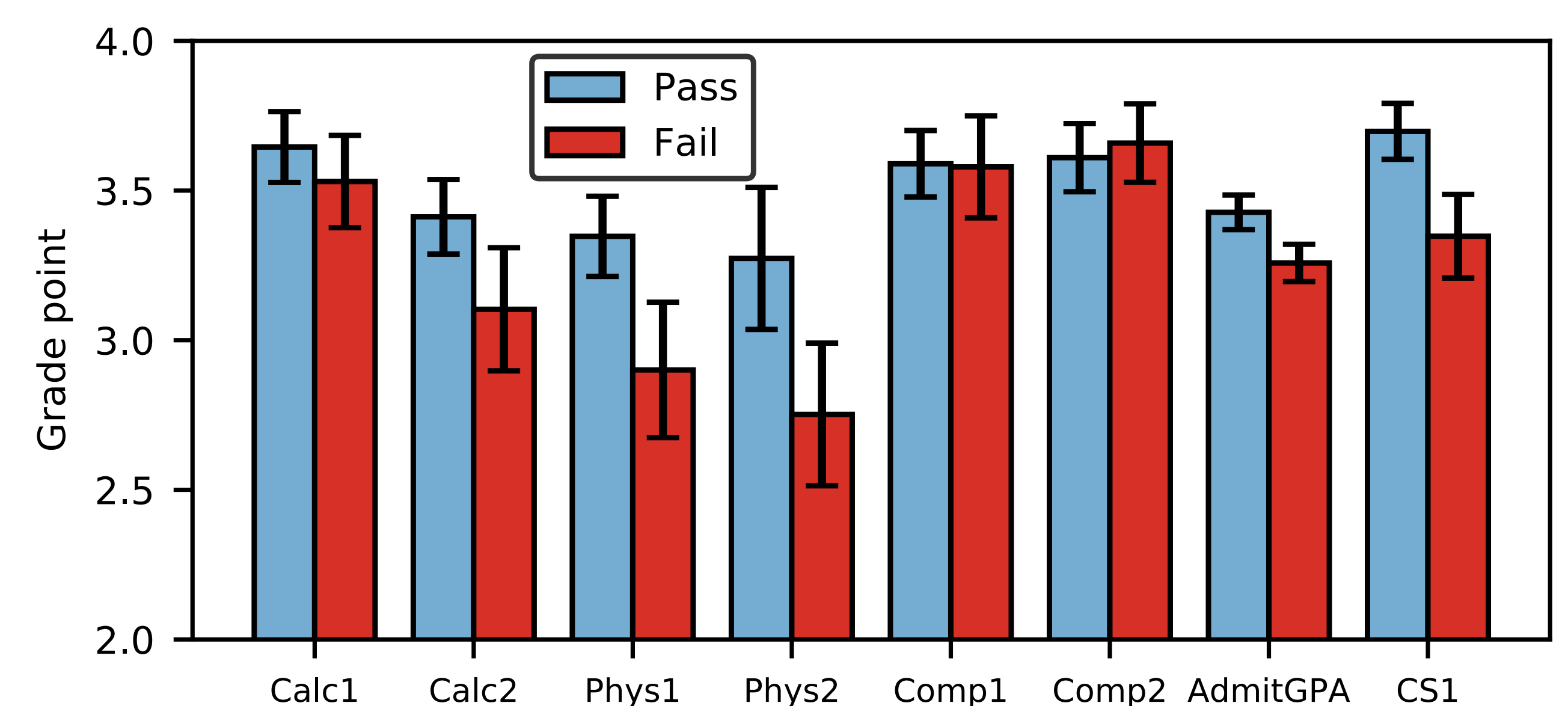
## Analysis and Results

### 1) Grade distribution and correlation:



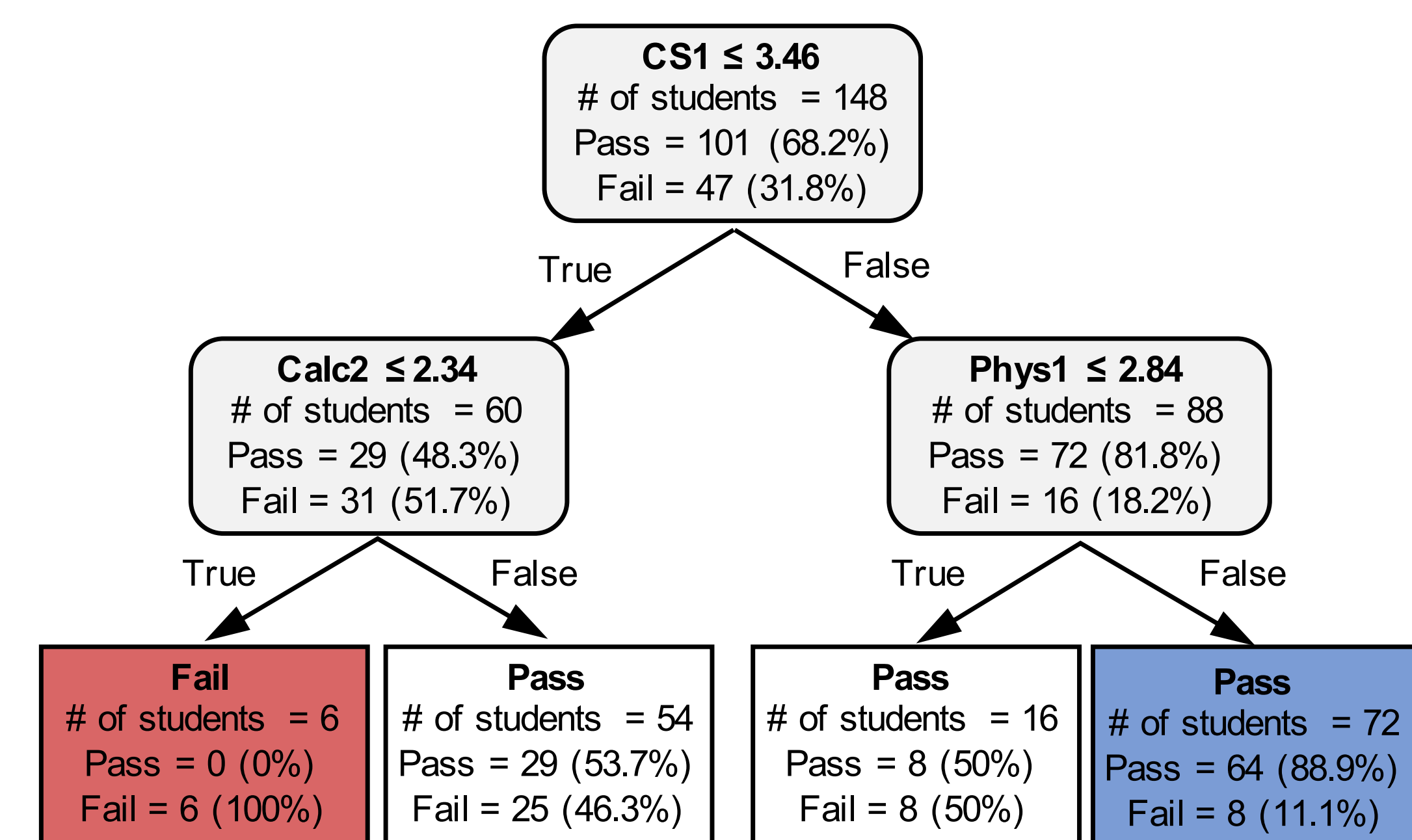
**Figure 1:** Scatter plot matrix between grade points in courses along with histogram in the main diagonal and correlation coefficient in the upper triangle.

### 2) Average grade points and t-test:



**Figure 2:** Average grade points in different courses with 95% confidence interval for two groups of students who pass/fail CS2. Statistically significant difference found for Calc2, Phys1, Phys2, AdmitGPA, and CS1.

### 3) Decision tree classifier:



**Figure 3:** Decision tree to predict if a student will pass or fail in CS2.

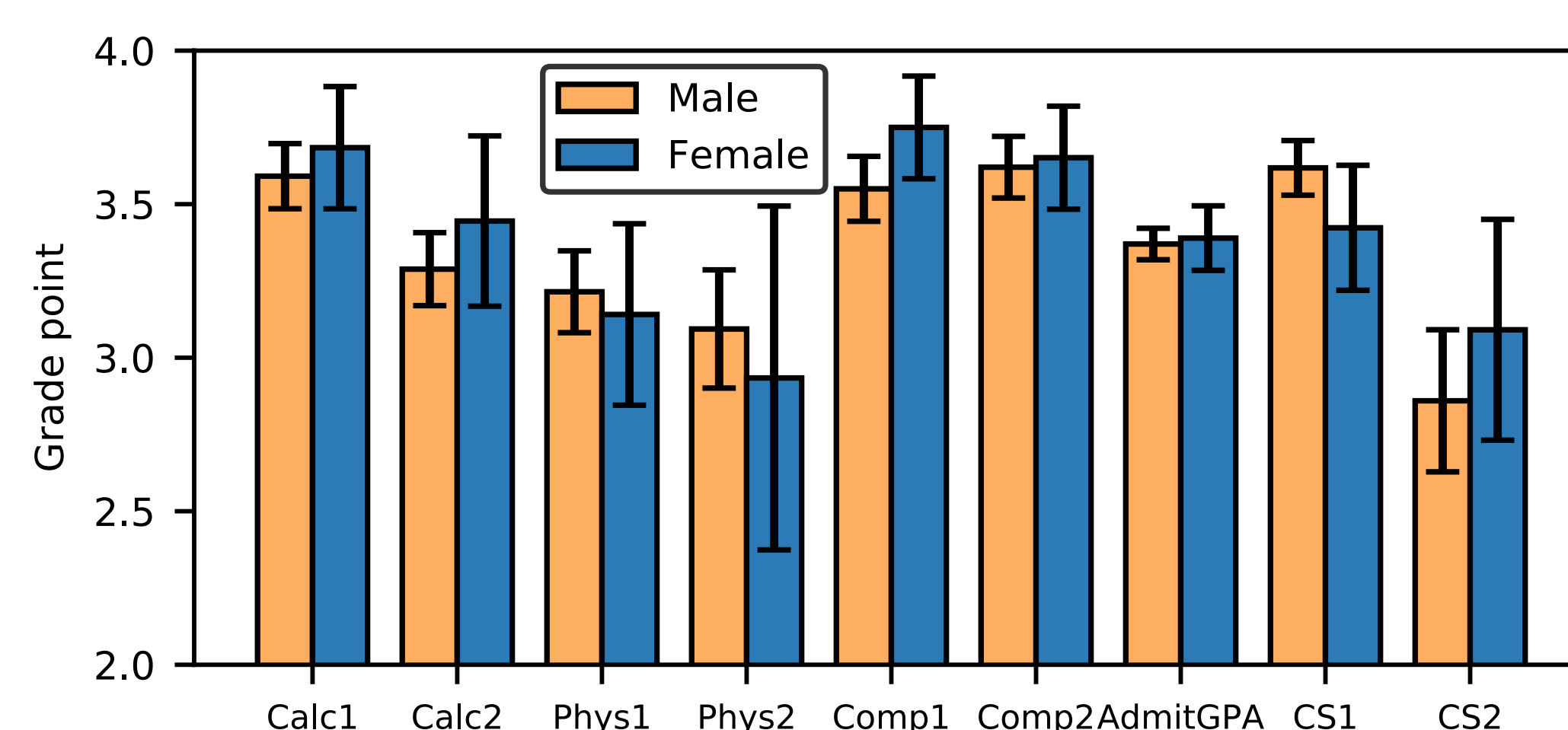
- **RULE #1: IF CS1 ≤ B+ AND Calc2 ≤ C+ THEN Fail.**  
All 6 students who satisfied this rule in fact failed.
- **RULE #2: IF CS1 > B+ AND Phys1 > B- THEN Pass.**  
88.9% of the students (64 out of 72) who satisfied this rule passed CS2.

### 4) Course importance ranking:

Attribute	Chi-Squared	Gain Ratio	Info Gain	Relief-F	Correlation	Average rank
CS1	1	1	1	1	1	1
Phys1	2	2	2	2	3	2.2
Phys2	3	3	3	6	2	3.4
Calc2	4	4	4	3	5	4
AdmitGPA	5	5	5	4	4	4.6
Comp1	6	6	6	5	7	6
Comp2	7	7	7	7	8	7.2
Calc1	8	8	8	8	6	7.6

**Table 1:** Ranking of importance of each course grade in predicting success in CS2 by different algorithms.

### 5) Effect of Gender:



**Figure 4:** Average grade points in different courses with 95% confidence interval for male and female students. No statistically significant difference found, except for Comp1.

## Summary and Future Work

We have shown how we can automatically analyze student grade data to find meaningful information. In the future, we would also like to study the impact of grades in core required computer science courses on the student's success in our programs.

## References

- [1] D. A. Trytten and A. McGovern, "Moving from managing enrollment to predicting student success," in *IEEE FIE*, Oct 2017, pp. 1–9.
- [2] A. N. Kumar, "Predicting student success in computer science – a reproducibility study," in *IEEE FIE*, Oct 2018, pp. 1–6.

This work is Institutional Review Board (IRB) exempt. IRB#: Pro00038240