Manasbi Parajuli
Professor Amruth Kumar
AI
April 15, 2018

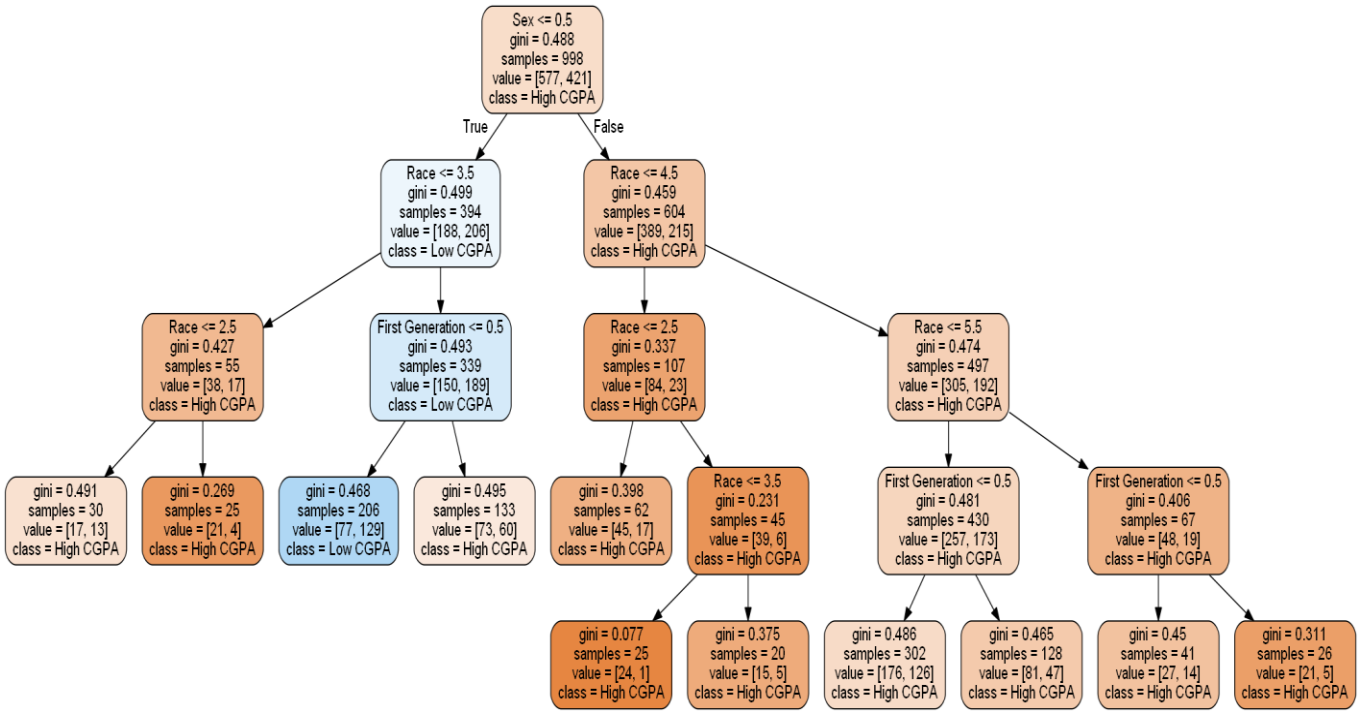
Project 4

Hypothesis 1: Who has a high CGPA?

Input Parameters:

a) Sex: Female or Male (0 or 1)
b) First Generation: 0 or 1
c) Race (Encoded as a number)

Outcome Variable: Cumulative GPA (**High:** Greater than 3.0, **Low:** Less than or equal to 3.0)
Total Samples: 1427
Train Samples: 998
Test Samples: 429
Maximum Depth: 5
Decision Tree Classifier Model Score: 62.47%



Analysis using Decision Tree Classifier Algorithm:

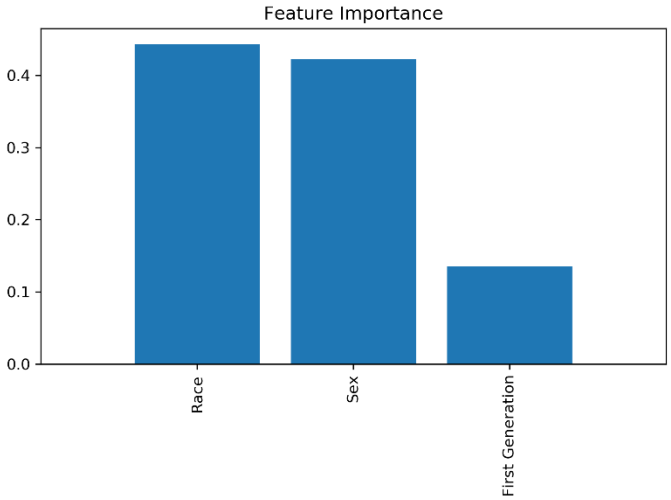
Of the **998 train samples** under consideration,

- **High CGPA**
  a. 30 females who had race encoded less than or equal to 2.5
  b. 25 females who had race encoded greater than 2.5
  c. 133 females who had race encoded greater than 3.5 and were first generation students
  d. 62 males who had race encoded less than or equal to 2.5
  e. 25 males who had race encoded less than or equal to 3.5
  f. 20 males who had race encoded greater than 3.5

g. 302 males who had race encoded less than or equal to 5.5 and were not first-generation students
h. 128 males who had race encoded less than or equal to 5.5 and were first generation students
i. 41 males who had race encoded greater than 5.5 and were not first-generation students
j. 26 males who had race encoded greater than 5.5 and were first-generation students

- **Low CGPA**
  a. 206 females who had race encoded greater than 3.5 and were not first-generation students

Random Forest Classifier:



Weightage of Input Parameters on the Random Forest Classification Algorithm

i) Race: 44.29%
ii) Sex: 42.21%
iii) First Generation: 13.5%

Train Accuracy: 62.73%

Test Accuracy: 63.4%

From our analysis and from the figure where we calculated the weightage of an input parameter, we deduce that the criterion of being a first-generation student does not have a significant decision on a high or low CGPA while the criteria race and sex impact equally.

Hypothesis 2: Do First Generation students graduate quickly in college?

Input Parameters:

a) Sex: Female or Male (0 or 1)
b) First Generation: 0 or 1
c) Cumulative GPA: 0 if less than or equal to 3.1, 1 if greater than 3.1

Outcome Variable: Semesters taken to graduate (0 if student took more than 15 semesters, 1 if a student took less than or equal to 15 semesters to graduate)

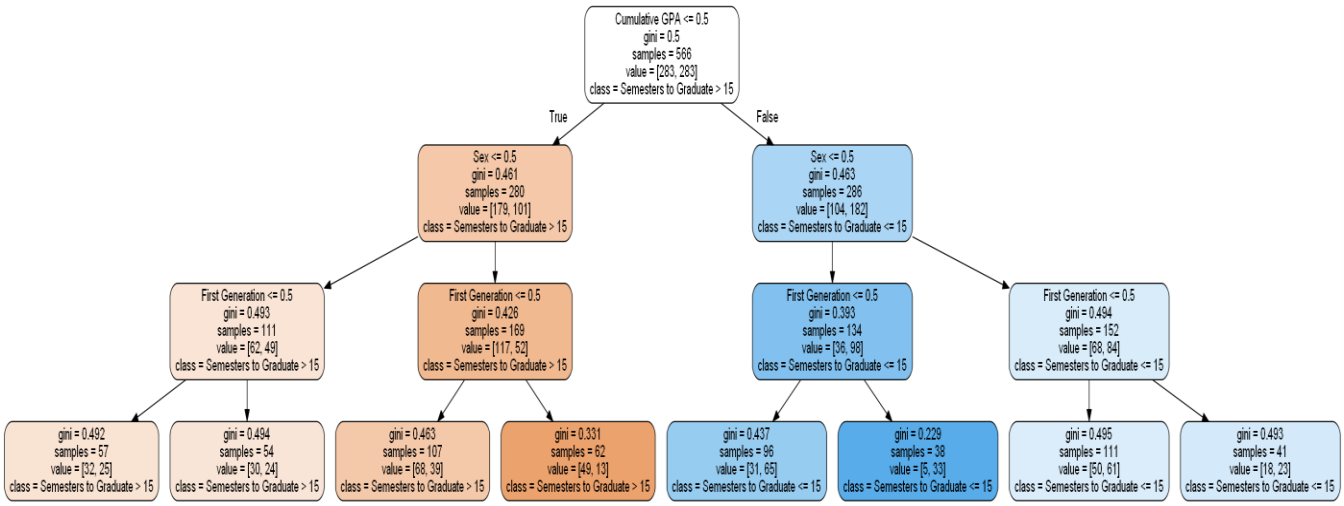
Total Samples: 787

Train Samples: 566

Test Samples: 221

Maximum Depth: 6

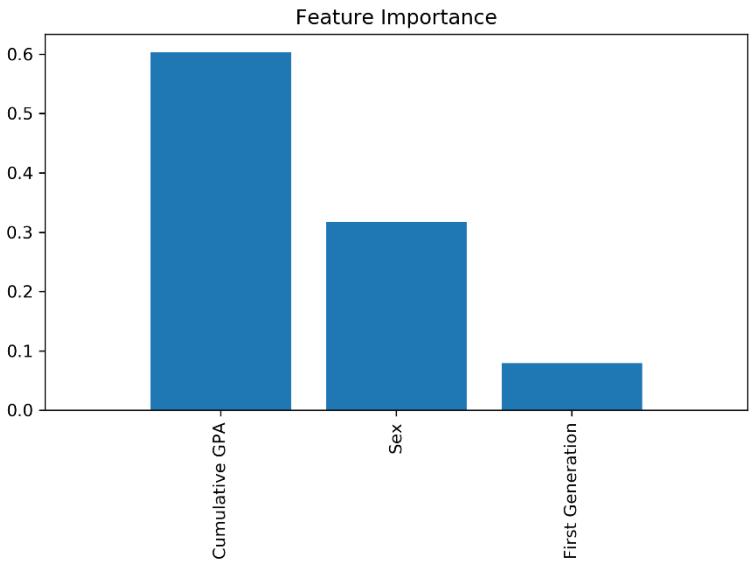
Decision Tree Classifier Model Score: 57.01%



Analysis using Decision Tree Classifier Algorithm:

Of the **566 train samples** under consideration,

- **Graduated quickly (Took less than or equal to 15 semesters)**
  a) 96 female students who had CGPA greater than 3.1 and were not first-generation students
  b) 38 female students who had CGPA greater than 3.1 and were first-generation students
  c) 111 male students who had CGPA greater than 3.1 and were not first-generation students
  d) 41 male students who had CGPA greater than 3.1 and were first-generation students
- **Did not graduate quickly (Took greater than 15 semesters)**
  a) 57 female students who had CGPA less than or equal to 3.1 and were not first-generation students
  b) 54 female students who had CGPA less than or equal to 3.1 and were first-generation students
  c) 107 male students who had CGPA less than or equal to 3.1 and were not first-generation students
  d) 62 male students who had CGPA less than or equal to 3.1 and were first-generation students



Random Forest Classifier:

Weightage of Input Parameters on the Random Forest Classification Algorithm

i) Cumulative GPA: 60.34%

ii) Sex: 31.74%

iii) First Generation: 79.2%

Train Accuracy: 63.78%

Test Accuracy: 57.01%

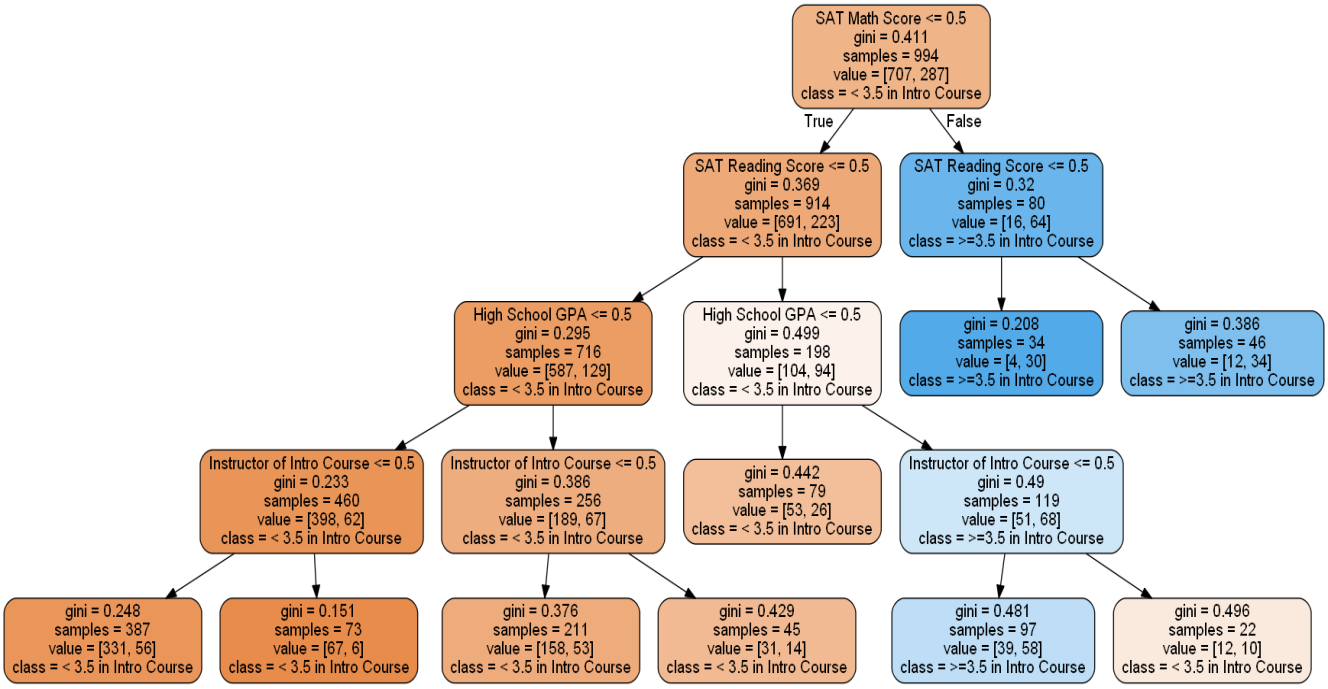
From our analysis and from the figure where we calculated the weightage of an input parameter, we deduce that being a first-generation student significantly influences on whether a student graduates quickly (takes less than or equal to 15 semesters to graduate). Also, a very high CGPA of greater than 3.1 influences if a student graduates quickly as compared to whether a student is male or female.

Hypothesis 3: Who does well in Introductory Course?

Input Parameters:

a) SAT Reading Score (1 if reading score > 600, 0 otherwise)
b) SAT Math Score (1 if math score > 715, 0 otherwise)
c) High School GPA (1 if high school GPA > 3.4, 0 otherwise)
d) Instructor of Intro Course (1 if instructor's rating > 6, 0 otherwise)

Outcome Variable: Grade in introductory course (1 if >= 3.5, 0 otherwise)
Total Samples: 1381
Train Samples: 994
Test Samples: 387
Maximum Depth: 6
Decision Tree Classifier Model Score: 77.78%

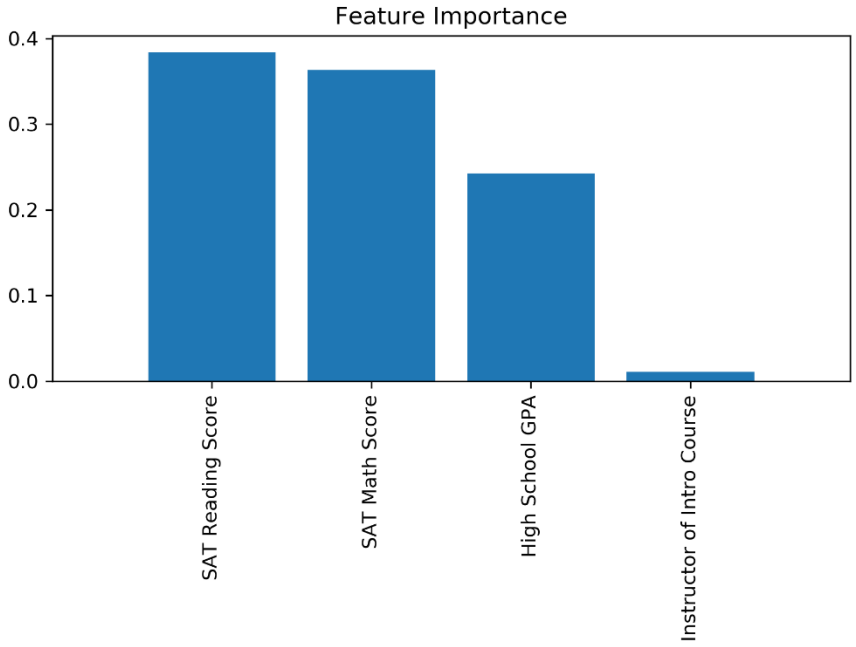


Analysis using Decision Tree Classifier Algorithm:

Of the **994 train samples** under consideration,

- **Does well in Introductory course (Gets more than 3.5 grade in the course)**
  a) 97 students who scored less than or equal to 715 in SAT Math, scored less than or equal to 600 in SAT Reading, had a high school GPA greater than 3.4 and whose instructor in the intro course had a rating of less than or equal to 6
  b) 34 students who scored greater than 715 in SAT Math and scored less than or equal to 600 in SAT Reading
  c) 46 students who scored greater than 715 in SAT Math and scored greater than 600 in SAT Reading
- **Did not do well in Introductory course (Gets less than or equal to 3.5 grade in the course)**
  a) 387 students who scored less than or equal to 715 in SAT Math, scored less than or equal to 600 in SAT Reading, had a High School GPA of less than or equal to 3.4 and whose instructor in the intro course had a rating of less than or equal to 6

b) 73 students who scored less than or equal to 715 in SAT Math, scored less than or equal to 600 in SAT Reading, had a High School GPA of less than or equal to 3.4 and whose instructor in the intro course had a rating greater than 6
c) 211 students who scored less than or equal to 715 in SAT Math, scored less than or equal to 600 in SAT Reading, had a High School GPA greater than 3.4 and whose instructor in the intro course had a rating of less than or equal to 6
d) 45 students who scored less than or equal to 715 in SAT Math, scored less than or equal to 600 in SAT Reading, had a High School GPA greater than 3.4 and whose instructor in the intro course had a rating greater than 6
e) 79 students who scored less than or equal to 715 in SAT Math, scored greater than 600 in SAT Reading, and had a High School GPA of less than or equal to 3.4
f) 22 students who scored less than or equal to 715 in SAT Math, scored greater than 600 in SAT Reading, had a High School GPA greater than 3.4 and whose instructor in the intro course had a rating greater than 6



Random Forest Classifier:

Weightage of Input Parameters on the Random Forest Classification Algorithm

i. SAT Reading Score: 38.39%
ii. SAT Math Score: 36.31%
iii. High School GPA: 24.22%
iv. Instructor of Intro Course: 1.08%

Train Accuracy: 77.67%
Test Accuracy: 78.55%

From our analysis and from the figure where we calculated the weightage of an input parameter, we deduce that the instructor does not have a significant influence in how a student will perform in the intro course. On the other hand, other factors such as the SAT scores have equal weightage in the performance of the student while high school GPA affects slightly less than SAT scores and more than the rating of the intro course instructor.

Hypothesis 4: Who does well in Software Course?

Input Parameters:

a) Grade in Intro Course (1 if grade >= 3.5, 0 otherwise)

b) Grade in Follow-up Course (1 if grade >=3.0, 0 otherwise)

c) Grade in Fundamentals Course (1 if grade >= 3.0, 0 otherwise)

Outcome Variable: Grade in Software Course (1 if grade > 3.2, 0 otherwise)

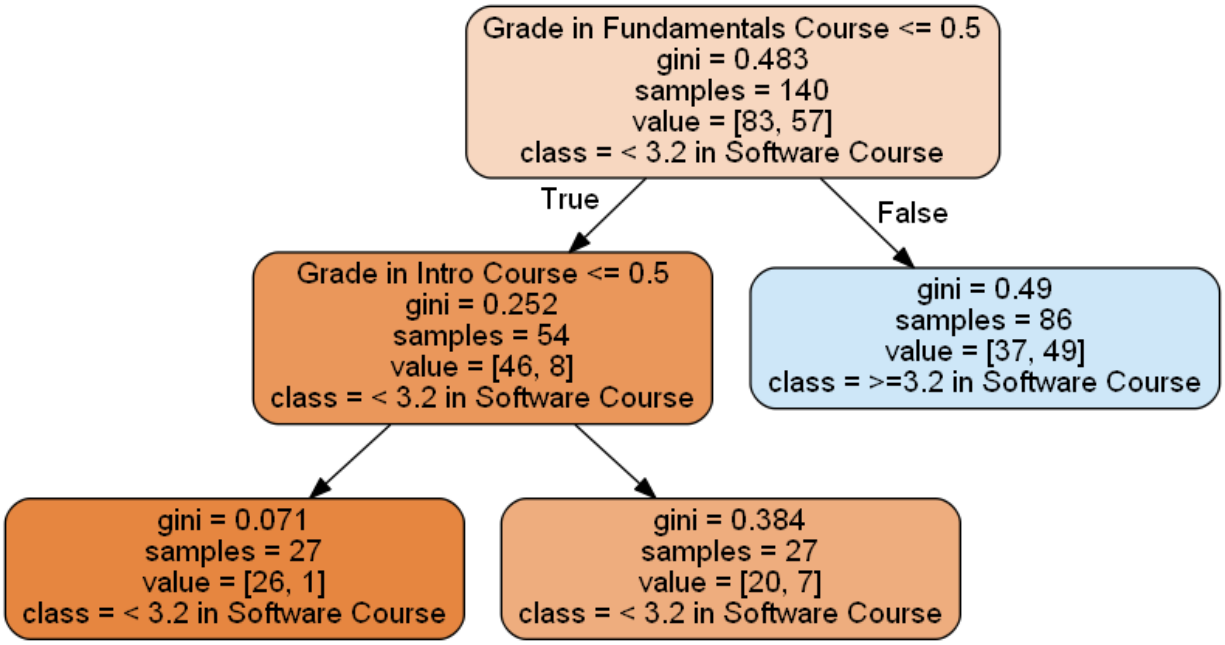
Total Samples: 195

Train Samples: 140

Test Samples: 55

Maximum Depth: 6

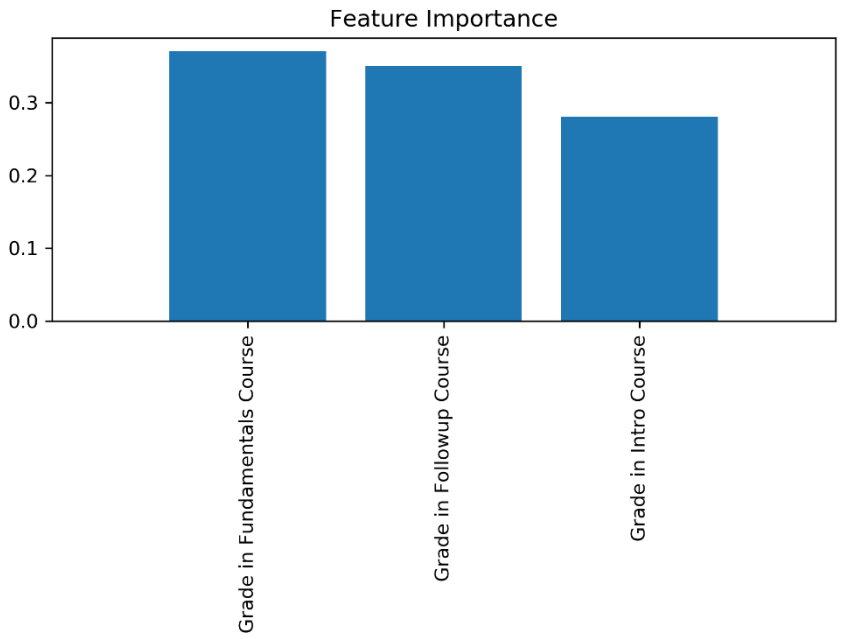
Decision Tree Classifier Model Score: 60%



Analysis using Decision Tree Classifier Algorithm:

Of the **140 train samples** under consideration,

- Did well in Software Course (Gets greater than or equal to 3.2 grade in the course)
  a. 86 students who got a grade of greater than or equal to 3.0 in Fundamentals Course
- Did not do well in Software Course (Gets less than 3.2 grade in the course)
  a. 27 students who got a grade of less than 3.0 in Fundamentals Course and got a grade of less than 3.5 grade in Intro Course
  b. 27 students who got a grade of less than 3.0 in Fundamentals Course and got a grade of greater than or equal to 3.5 grade in Intro Course



Random Forest Classifier:

Weightage of Input Parameters on the Random Forest Classification Algorithm

i. Grade in Fundamentals Course: 37%
ii. Grade in Follow-up Course: 35%
iii. Grade in Intro Course: 28%

Train Accuracy: 68.57%

Test Accuracy: 65.45%

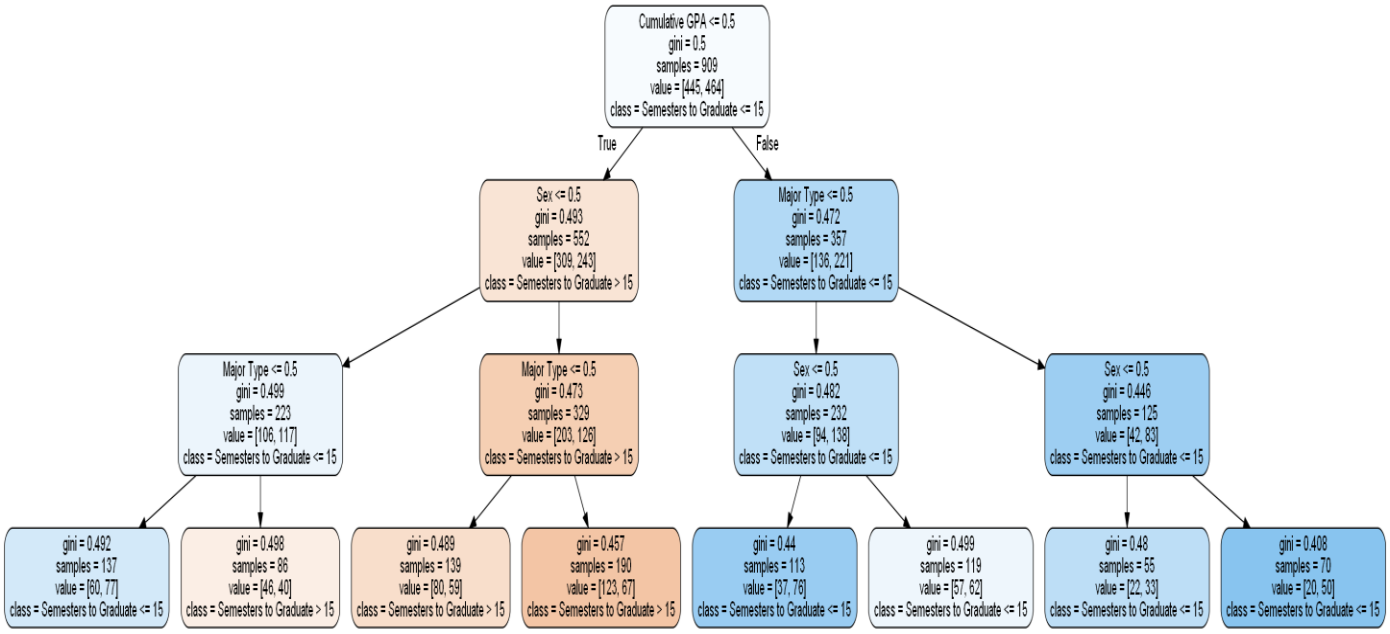
From our analysis and from the figure where we calculated the weightage of an input parameter, we deduce that a student who does considerably well in Fundamentals Course and Follow-up Course will get a grade greater than 3.2 in Software Course. However, the data does not have enough test samples to verify the claim. Also, our decision tree model gives us a very low score of 60% and our decision tree has only 3 leaves implying that our hypothesis may not be good enough.

Hypothesis 5: How quickly do Sciences or Non-Sciences students graduate?

Input Parameters:

a) Sex (Female: 0, Male: 1)
b) Major Type (1 if Non-Science, 0 if majors of CS, Math and Science)
c) Cumulative GPA (1 if CGPA > 3.2, 0 otherwise)

Outcome Variable: Semesters taken to graduate (0 if student took more than 15 semesters, 1 if a student took less than or equal to 15 semesters to graduate)
Total Samples: 1263
Train Samples: 909
Test Samples: 354
Maximum Depth: 6
Decision Tree Classifier Model Score: 58.19%



Analysis using Decision Tree Classifier Algorithm:

Of the **909 train samples** under consideration,

- **Graduated quickly (Took less than or equal to 15 semesters)**
  - 137 female students who had a CGPA less than or equal to 3.2 and who are Sciences major
  - 113 female students who had a CGPA greater than 3.2 and who are Sciences major
  - 119 male students who had a CGPA greater than 3.2 and who are Sciences major
  - 55 female students who had a CGPA greater than 3.2 and who are Non-Sciences major
  - 70 male students who had a CGPA greater than 3.2 and who are Non-Sciences major
- **Did not graduate quickly (Took greater than 15 semesters)**
  - 86 female students who had a CGPA less than or equal to 3.2 and who are Non-Sciences major
  - 139 male students who had a CGPA less than or equal to 3.2 and who are Sciences major
  - 190 male students who had a CGPA less than or equal to 3.2 and who are Non-Sciences major

Random Forest Classifier:

Weightage of Input Parameters on the Random Forest Classification Algorithm

i) Cumulative GPA: 58.38%

ii) Sex: 25.90%

iii) Major Type: 15.72%

Train Accuracy: 60.18%

Test Accuracy: 58.19%

From our analysis and from the figure where we calculated the weightage of an input parameter, we deduce that students who had a high CGPA tend to graduate quickly with about 58.19% accuracy. Students' choices of a major do not have much of a significance in whether they will graduate quickly. Also, we can infer with some degree of accuracy that female students tend to graduate quickly compared to the men.