

# Logistic Regression vs Tree based Model

Md Mominul Islam  
101009250

# About Data

- SDSU Fictitious Data with 350 Observations and 7 Columns

Dependent Variables: Graduated

Independent Variables : Enrolled, HSGPA and ACT Score

# Objective

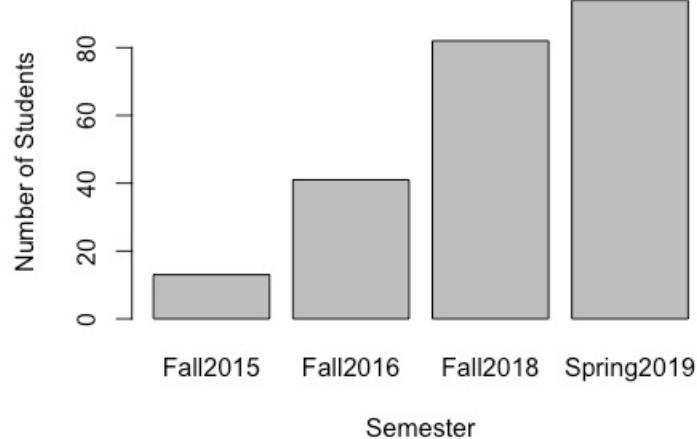
Understand and compare logistic regression model and Tree based model.

## *Exploratory Analysis of Categorical Data*

### Cohort Column

Fall 2015 Session	Fall 2016 Session	Fall 2018 Session	Spring 2019
13	41	82	94

**Cohorts throughout the Semester**



If we look at our cohort column, we can see that from 2015 to 2019, no. of enrollment increases.

## *Exploratory Analysis of Categorical Data*

### Enrollment and Graduated Column

0 (Otherwise)	1 (Being Enrolled)
91	139

Not Graduated	Graduated
139	91

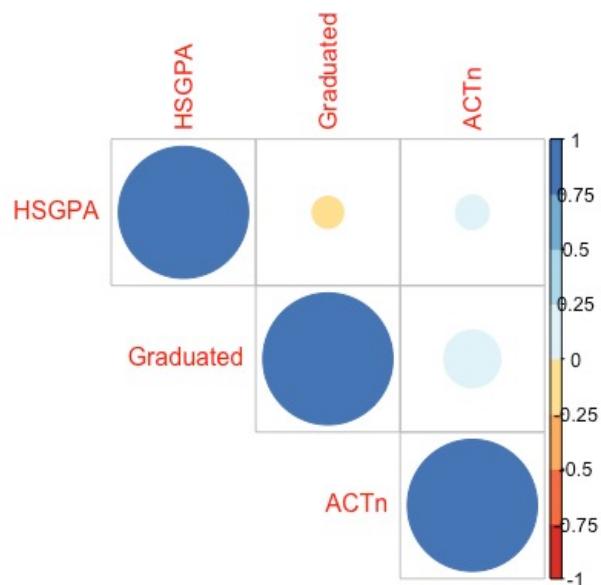
- Out of 230 admitted students, 139 are being enrolled.
- We have 230 students, out of which 91 of them graduated and 139 didn't.

### Summary Statistics

Parameters	HSGPA	ACT
Minimum Value	0.0	14.00
First Quartile	3.235	22.00
Median	3.460	25.00
Mean	3.435	25.34
Third Quartile	3.770	29.00
Maximum Value	4.00	34.00

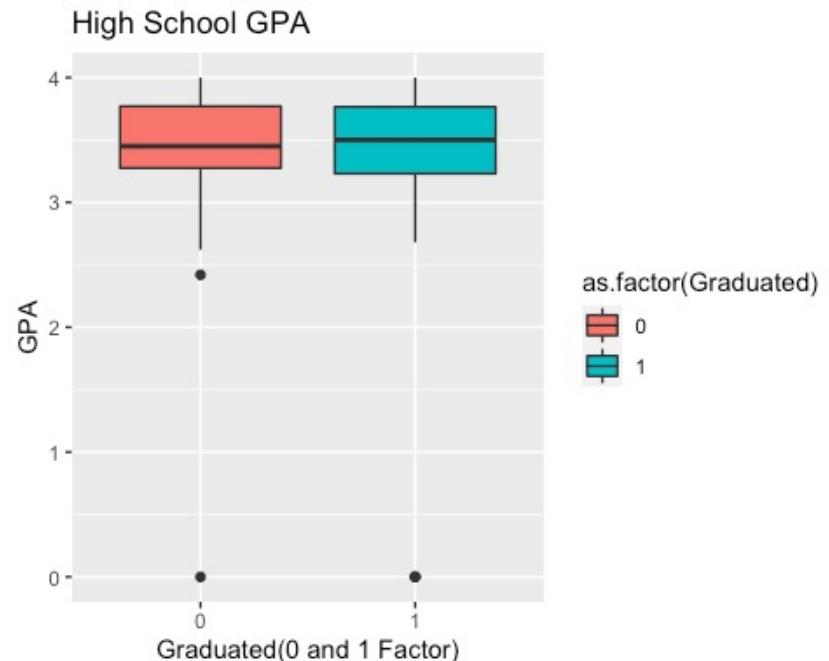
Average high school GPA of the students in South Dakota State University is 3.435, the average ACT score is 25.34.

# Correlation Plot



- weak correlation in between HSGPA and Graduated variables.
- slightly positive correlation between ACT score and Graduated variables.

# Boxplot of Numeric Variables



50% of the students have GPA in between 3.00~4.00. There are some extreme values in our data set.

# Logistic Model

$$\text{logit}(P) = \log\left(\frac{P}{1 - P}\right)$$

Here, P means probability of being graduated.

	Estimate	Std. Error	t-value	Pr(> t )
Intercept	-1.76770	1.11539	-1.585	0.11301
HSGPA	-0.27025	0.25402	-1.064	0.28739
ACT	0.08906	0.03061	2.909	0.00362 **
Enrolled	20.30775	1093.46	0.019	0.98518

## Parameter Interpretation

- $b_{HSGPA} = -0.270$  and  $\exp(-0.270) = 0.76$ . So, with unit increase of high school GPA, the odd of a student being graduated will increase by 76.33%.
- $b_{ACT} = -0.08906$  and  $\exp(0.08906) = 1.093$ . So, with unit increase of ACT score, the odd of a student being graduated will increase by 109.3%

### Error rate

		Predicted Class
Original Class	0	1
0	123	16
1	71	20

- 0.5 as our cutoff.

Any value less than 0.5 will be classified as not Graduated (0), and the others will be classified as Graduated (1)

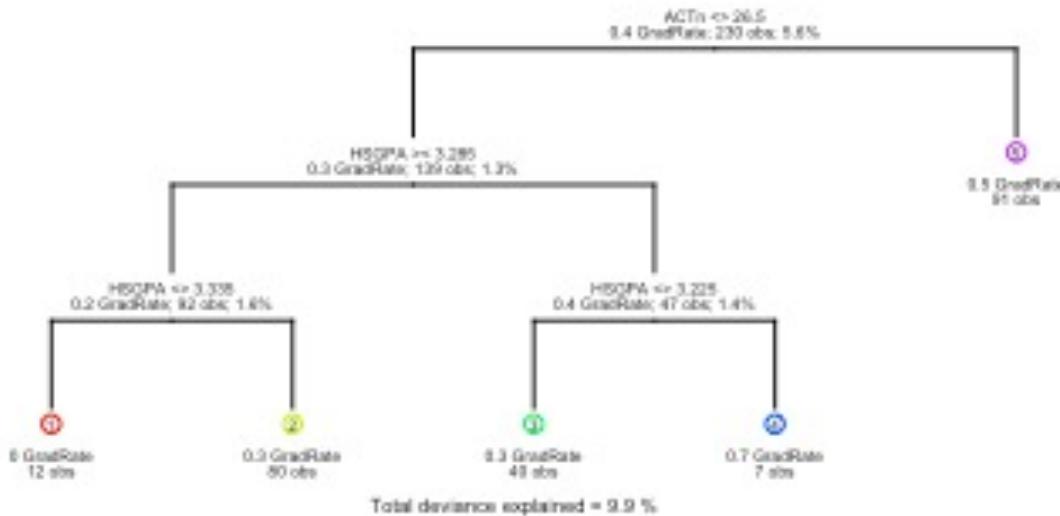
- Error rate: 37.82%.

## Validation of model: Training and Testing

- 29 were predicted wrong out of 69. So the error rate here is  $29/69 = 42.02\%$ .

		Predicted Class
Original Class	0	1
0	34	8
1	21	6

# TREE REGRESSION MODEL



# Comparison Between Two Models

