Problem Statement:

An education department in the US needs to analyse the factors that influence the admission of a student into a college.

Variables in the Dataset:

• GRE (Graduate Record Exam scores)

• GPA (grade point average)

• Rank refers to prestige of the undergraduate institution. The variable rank takes on the values 1 through 4. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest.

• Admit is a response variable; admit/don't admit is a binary variable where 1 indicates that student is admitted and 0 indicates that student is not admitted.

• SES refers to socioeconomic status: 1 - low, 2 - medium, 3 - high.

• Gender male (0, 1) = 0 -> Female, 1 -> Male

• Race – 1, 2, and 3 represent Hispanic, Asian, and African-American

Analyse the historical data and determine the key drivers.

Analysis information:

• Run logistic model to determine the factors that influence the admission process of a student (Drop insignificant variables)

• Transform variables to factors wherever required

• Calculate accuracy of the model

• Try a decision tree and select a champion model

• Determine the accuracy rates for each model

• Select the most accurate model