Summary

The model is built for Education institute to determine the probability of person taking admission to the course. For this various feature are provided upon which we need to determine various factors contributing to the conversion of person into taking the admission.

I followed these steps to build a model and draw conclusions:

1 Importing Libraries:

Imported necessary libraries for EDA and data processing.

2 Data cleaning and processing

Checked data for null values and eliminated the null values as well as replaced it with some other values wherever needed.

3 EDA

I divided this process into univariate and categorical data visualization. I understood that it is difficult to determine feature importance mere based on visualization for elimination. Hence decided to go with other methods to feature engineering.

4 Dummy variables

Created dummy variables for categorical data with multiple slots and removed the original columns.

5 Train test split

Then I divided data into train and test set with train data as 70% and test data as 30%.

6 Model building and evaluation

Did RFE to decide top 15 features and removed the less relevant columns based on VIF and p value. Initially kept cutoff 0.5 and later using ROC curve decided optimum cutoff as 0.35 and got accuracy, sensitivity and specificity of around 80%.

7 Prediction on test set

Did prediction on test set and got similar accuracy.