**Student Academic Dropout Prediction**

**1.Objectives:**

* To study historical data and understand factors related to success and dropout of students
* To study how these factors contribute to their academic performance
* To identify interdependencies and interactions between variables affecting students’ overall performance and areas of improvements
* To give timely suggestions can be given to educational institutions to implement precautionary measures required to reduce attrition rate of students

**2.Data:**

* A second-hand dataset from Kaggle - <https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention>
* Data consists of 4424 records of 35 variables

**3.Data Description:**

The dataset contains 35 variables which are coded as numeric. For example, Marital status is coded as numbers from 1 to 6, each with different categories as below:

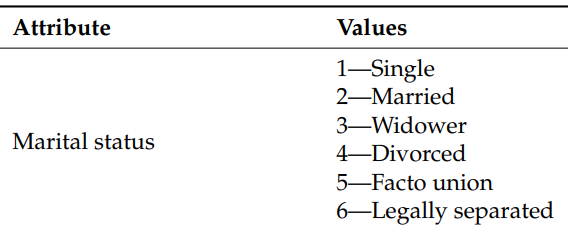


Figure 1: Values for attribute Marital Status

Similarly, gender variable is coded as 1 and 2, with categories as below:

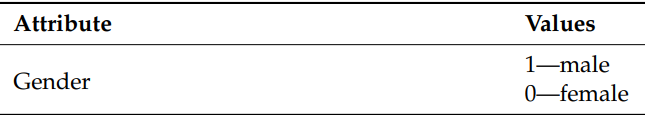


Figure 2: Values for attribute Gender

The main outcome variable is named “Target” in the dataset with categories “Dropout”, “Enrolled” and “Graduate”.

The Description of all the other coded categories is given below:

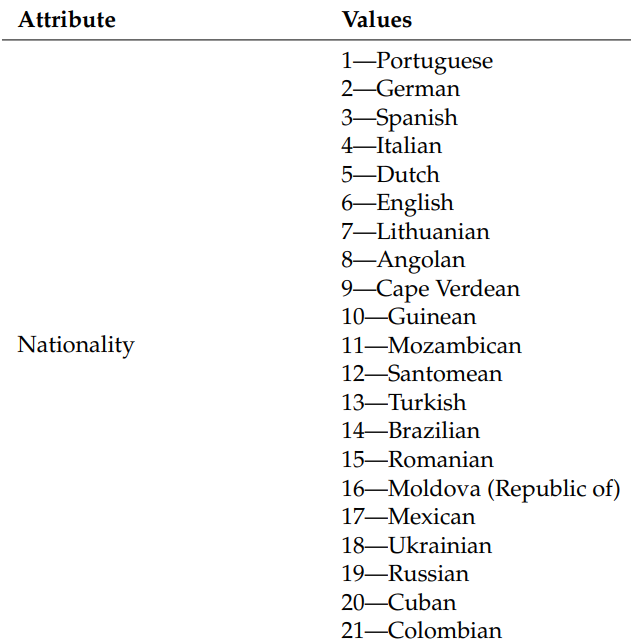
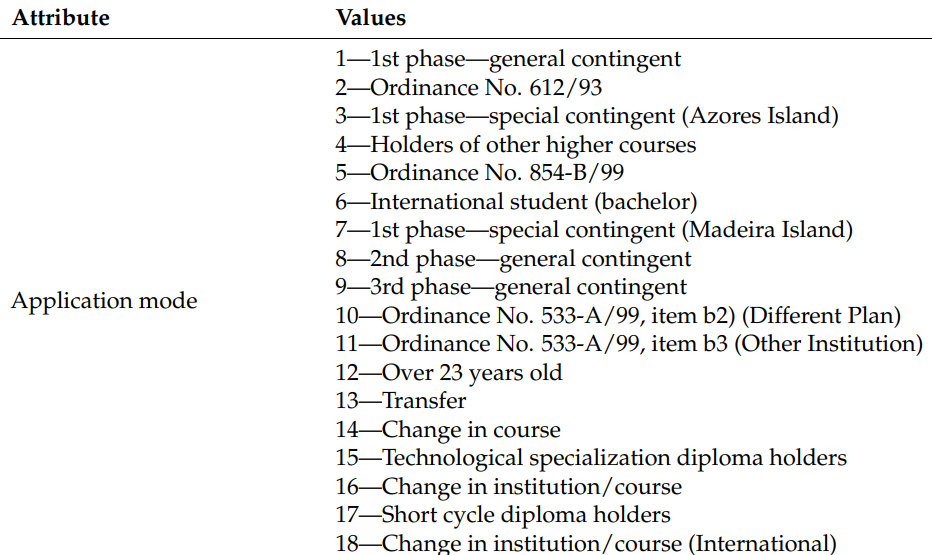


Figure 4: Values for attribute Application mode

Figure 3: Values for attribute Nationality

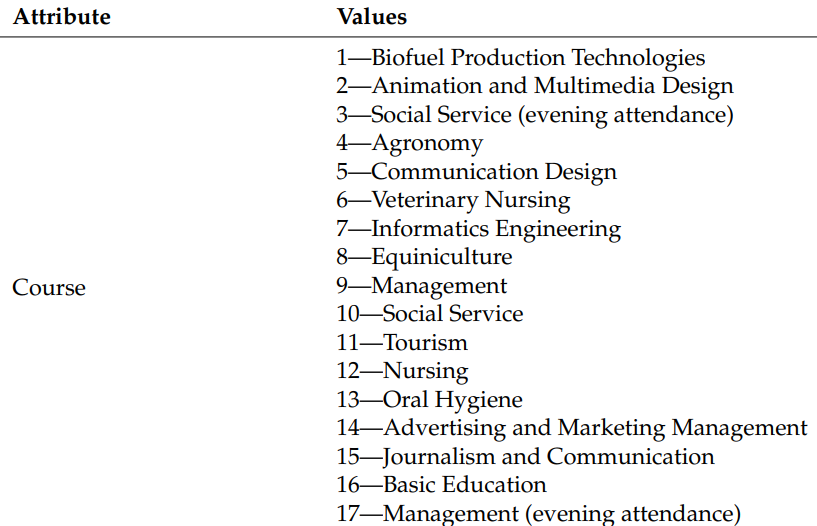


Figure 5: Values for attribute Course

Text

Description automatically generated

Text

Description automatically generated

Figure 6: Values for attribute Previous Qualification

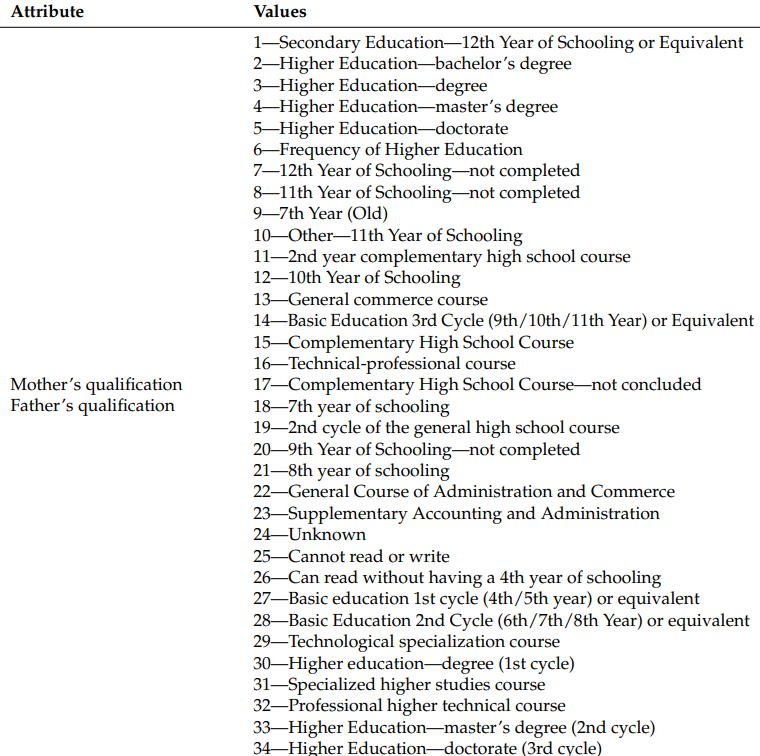


Figure 7: Values for attribute qualification of parents

Shape

Description automatically generated with medium confidence

Figure 8: Values for attribute attendance

Text

Description automatically generated

Figure 9: Values for Miscellaneous attributes

A picture containing text

Description automatically generated

Text

Description automatically generated

Figure 10: Values for attribute occupation of parents

**4.Data Analysis:**

* The data is skewed as most of the students enroll in the age group 20-30, so we have more records with age 20-30

**5.Data Pre-processing and evaluating plots:**

* Converting the numerical categories into nominal categories

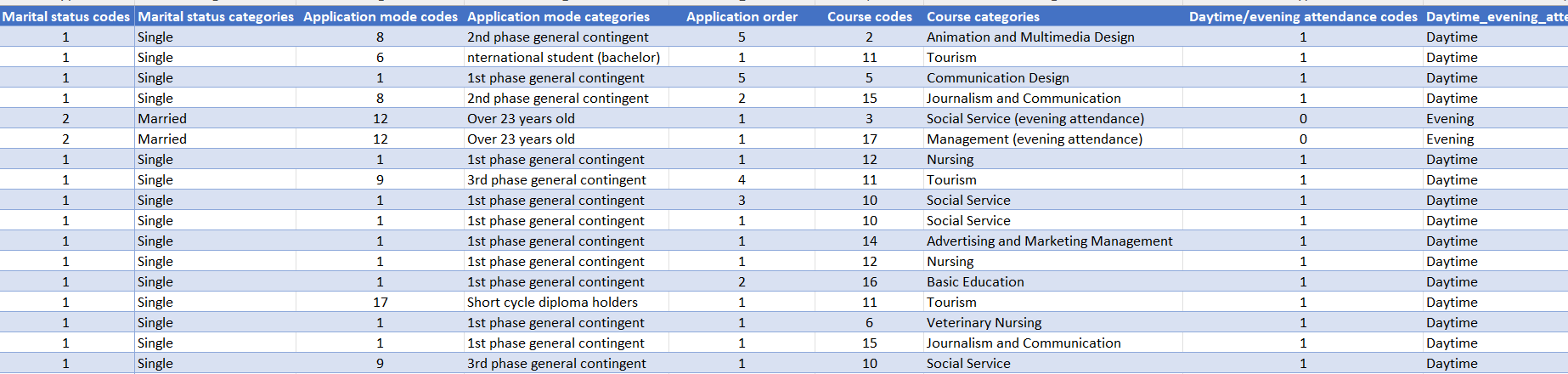


Figure 11: Conversion of numerical categories to nominal categories

* Made use of plots like Histogram and Boxplot to understand distribution of variables like age, GDP, inflation rate, gender, course categories.

**6. Model Evaluation using ROC:**

* The area under the Curve (AUC) is 0.94 showing that the logistic regression model built is performing equally well on training data as well as test data set.

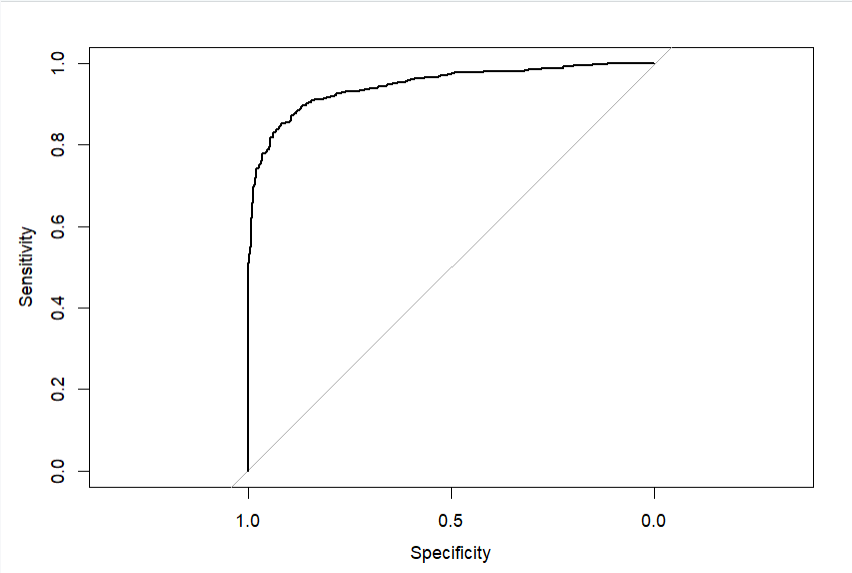


Figure 12: ROC Curve

**7. Results:**

From the summary of Logistic Regression Model, we can say that “Course Codes”, “Debtor Codes”, “Tuition fees up to date codes”, “Gender codes”, “Scholarship holder codes” , “Curricular units 1st sem (credited)” , “Curricular units 1st sem (enrolled)”, “Curricular units 1st sem (approved)” have a major impact on predicting the drop out probability.

**8.Conclusion:**

The odds of students dropping out are highly dependent on the course they are enrolled in, if they have any debts, age, gender and curricular units and grades in 1st semester.

Based on analysis, we can provide suggestions to academic institutions to reduce the dropout rate among students. Since dropout rate depends on curricular units, institutions can provide higher flexibility in course options. Students with debt may receive additional student assistance. Many students in the age group above 30 have obligations outside school such as looking after families and employment. To avoid drop out of these students, institutions can avail themselves of the facility of online classes.

**9.References:**

<https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention>

<https://cran.r-project.org/>

<https://utdallas.primo.exlibrisgroup.com/permalink/01UT_DALLAS/2hgl0t/alma9927850104601421>