|  |
| --- |
| cid:image003.jpg@01D44128.3F325EF0  **Machine Learning** |
| **Lab assignment 1: Classification** |
| **Important information** |
| * Write a report analyzing the results obtained after completing the assignment below. **The maximum length is 8 pages.** * The report should contain text and figures that illustrate the analysis performed. It is **very important** to interpret the results obtained. Do not include figures without a comment about it. Remember to include a conclusion section at the end. * Upload your report and the R code. |
| **Statement** |
| **Datasets**:  This assignment will analyze the **cardioMod4.csv** datasetwhich contains the following information:  Attributes information   * **age**: Age: of the patient in years. * **gender**: Gender * **height**: Height (cm) * **weight**: Weight (kg) * **ap\_hi**: Systolic blood pressure. * **ap\_lo**: Diastolic blood pressure. * **cholesterol**: Cholesterol: 1: normal, 2: above normal, 3: well above normal * **gluc**: Glucose: 1: normal, 2: above normal, 3: well above normal * **smoke**: Whether the patient smokes or not * **alco**: Alcohol intake * **active**: Physical activity   Output   * **cardio**: Presence of cardiovascular disease.   **Assignment**  The objective of this practice is to compare different classification algorithms with a real dataset.  Load the dataset **cardioMod4.csv** and:   * Perform an exploratory analysis of the data (Type of variables, Missing values, outliers, and relations between variables). * Identify all classification models seen in theory to predict the **cardio** variable. * Perform a comparative analysis of the performance of the models and predictors. The comparison should be summarized in the form of a table including different performance measures on cross-validation, training and validation.     Hints:   * Identifying a model implies selecting the most significant variables and the optimal structure of the model. |
| cid:image002.png@01D44127.A8A3BCE0 |

