**Statistical Learning Theory – Data Science – 4th Term – 11/23/2023**  
**Report Development**  
To compose the second grade of the semester (P2), you must prepare a report based on the study of student alcoholism.

The file **student-lpor.csv** contains data from a research study conducted by Fabio Pagnotta and Hossain Mohammad Amran, Università Di Camerino – UNICAM. This research was carried out with high school students and includes numerous social and gender-related variables, as well as grades in the Portuguese Language subject. The file **student-lpor.docx** contains a description of this dataset.

Based on the data from the files, you are to prepare a report presenting **descriptive statistics** (measures of central tendency, dispersion, and shape) and **inferential statistics** (measures of association and classification) for two or more related variables, along with the appropriate **graphical representations**.

In addition, you must conduct a study to introduce a **new variable that classifies students as alcoholic or not**. From there, identify **which variables are predictors** of this new response variable.

You do **not** need to present code routines, but the text **must mention the tool used** for the analysis.

**Student Alcoholism**  
**Context:**  
The data was obtained from a survey conducted with high school students, providing extensive social and gender information, as well as grades in the Portuguese Language subject.

**Source Information:**  
P. Cortez and A. Silva. Using Data Mining to Predict the Performance of High School Students. In A. Brito and J. Teixeira (Eds.), Proceedings of 5th Future Business Technology Conference (FUBUTEC 2008), pp. 5-12, Porto, Portugal, April 2008, EUROSIS, ISBN 978-9077381-39-7.  
Fabio Pagnotta, Hossain Mohammad Amran. Università Di Camerino - UNICAM.

**Labels:**  
school – student’s school (binary: GP - Gabriel Pereira or MS - Mousinho da Silveira)  
sex – student’s gender (binary: 'F' - female or 'M' - male)  
age – student’s age (numeric: 15 to 22)  
address – type of home address (binary: 'U' - urban or 'R' - rural)  
famsize – family size (binary: 'LE3' - less than or equal to 3 or 'GT3' - greater than 3)  
Pstatus – parents’ cohabitation status (binary: 'T' - living together or 'A' - apart)  
Medu – mother's education level (0 - none, 1 - primary education 1, 2 - primary education 2, 3 - secondary education, or 4 - higher education)  
Fedu – father's education level (0 - none, 1 - primary education 1, 2 - primary education 2, 3 - secondary education, or 4 - higher education)  
Mjob – mother's job (nominal: teacher, health, services, at\_home, or other)  
Fjob – father's job (nominal: teacher, health, services, at\_home, or other)  
reason – reason for choosing this school (nominal: home - close to home, reputation - school’s reputation, course - course preference, or other)  
guardian – student’s guardian (nominal: mother, father, or other)  
traveltime – home-to-school travel time (intervals: 1 - <15 min, 2 - 15–30 min, 3 - 30 min–1 hour, or 4 - >1 hour)  
studytime – weekly study time (intervals: 1 - <2 hrs, 2 - 2–5 hrs, 3 - 5–10 hrs, or 4 - >10 hrs)  
schoolsup – extra educational support (binary: yes or no)  
famsup – family educational support (binary: yes or no)  
paid – private lessons (binary: yes or no)  
activities – participates in extracurricular activities (binary: yes or no)  
nursery – attended preschool (binary: yes or no)  
higher – wants to pursue higher education (binary: yes or no)  
internet – Internet access at home (binary: yes or no)  
romantic – in a romantic relationship (binary: yes or no)  
famrel – quality of family relationships (categorical: 1 - very bad to 5 - excellent)  
freetime – free time after school (categorical: 1 - very low to 5 - very high)  
goout – time spent with friends (categorical: 1 - very low to 5 - very high)  
Dalc – weekday alcohol consumption (categorical: 1 - very low to 5 - very high)  
Walc – weekend alcohol consumption (categorical: 1 - very low to 5 - very high)  
health – current health status (categorical: 1 - very bad to 5 - very good)  
absences – number of school absences (numeric: 0 to 93)

These grades are related to the subject under study:  
G1 – first semester grade (numeric: 0 to 20)  
G2 – second semester grade (numeric: 0 to 20)