ASSIGNMENT 3 (20 POINTS)

• This assignment will be solved individually or in groups of two students. You must upload your solution (**html and Rmd files**) at Aula Digital, be sure to include your full name at the top of your solution. NO NAME, NO GRADE!

• **Due date**: December 7th, 2023 at 23.55h.

• **NO late assignments will be allowed.**

To complete this assignment create a .Rmd file with your answers. Once completed, generate an .html document and upload both documents at Aula Digital in the space provided. Make sure your name is visible at the top of the document and show your R code.

This assignment uses the data frame `Cirrhosis` dataset that you can find in the DATASET folder.

During 1974 to 1984, 424 PBC patients referred to the Mayo Clinic qualified for the randomized placebo-controlled trial testing the drug D-penicillamine. Of these, the initial 312 patients took part in the trial and have mostly comprehensive data. The remaining 112 patients didn't join the clinical trial but agreed to record basic metrics and undergo survival tracking. Six of these patients were soon untraceable after their diagnosis, leaving data for 106 of these individuals in addition to the 312 who were part of the randomized trial.

- Variable Information
    1. ID: unique identifier
    2. N_Days: number of days between registration and the earlier of death, transplantation, or study analysis time in July 1986
    3. Status: status of the patient C (censored), CL (censored due to liver tx), or D (death)
    4. Drug: type of drug D-penicillamine or placebo
    5. Age: age in [days]
    6. Sex: M (male) or F (female)
    7. Ascites: presence of ascites N (No) or Y (Yes)
    8. Hepatomegaly: presence of hepatomegaly N (No) or Y (Yes)
    9. Spiders: presence of spiders N (No) or Y (Yes)
    10. Edema: presence of edema N (no edema and no diuretic therapy for edema), S (edema present without diuretics, or edema resolved by diuretics), or Y (edema despite diuretic therapy)
    11. Bilirubin: serum bilirubin in [mg/dl]
    12. Cholesterol: serum cholesterol in [mg/dl]
    13. Albumin: albumin in [gm/dl]
    14. Copper: urine copper in [ug/day]
    15. Alk_Phos: alkaline phosphatase in [U/liter]
    16. SGOT: SGOT in [U/ml]
    17. Triglycerides: triglicerides in [mg/dl]
    18. Platelets: platelets per cubic [ml/1000]
    19. Prothrombin: prothrombin time in seconds [s]
    20. Stage: histologic stage of disease (1, 2, 3, or 4)
- Class Labels: status of the patient 0 = D (death), 1 = C (censored), 2 = CL (censored due to liver transplantation)

Consider the following **research question**: Utilize the data provided for predicting survival state of patients with liver cirrhosis.

To answer this question you will consider different classification models such as perceptrons, decision trees, nearest neighbours, ...
For each method, use different structures and/or parameters, different training/testing sets, compare the different models obtained and comment your results. Which model is the "best"?
Be sure to include a conclusion that summarises all the work done and what you have learned. The purpose of a conclusion is to restate the central idea and supporting arguments of the work done. A conclusion closes the assignment stating how you have achieved the proposed objectives.
