ASSIGNMENT 5 (15 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**: January 8th, 2024 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.

Consider the mushroom data that you can find in the DATASET folder.

- Variable Information
  1. type: edible=e, poisonous=p
  2. cap-shape: bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
  3. cap-surface: fibrous=f,grooves=g,scaly=y,smooth=s
  4. cap-color: brown=n,buff=b,cinnamon=c,gray=g,green=r, pink=p,purple=u,red=e,white=w,yellow=y
  5. bruises?: bruises=t,no=f
  6. odor: almond=a,anise=l,creosote=c,fishy=y,foul=f, musty=m,none=n,pungent=p,spicy=s
  7. gill-attachment: attached=a,descending=d,free=f,notched=n
  8. gill-spacing: close=c,crowded=w,distant=d
  9. gill-size: broad=b,narrow=n
  10. E gill-color: black=k, brown=n, buff=b, chocolate=h, gray=g, green=r, orange=o, pink=p, purple=u, red=e, white=w, yellow=y
  11. stalk-shape: enlarging=e,tapering=t
  12. stalk-root: bulbous=b,club=c,cup=u,equal=e, rhizomorphs=z,rooted=r,missing=?
  13. stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s
  14. stalk-color-above-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y
  15. stalk-color-below-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y
  16. veil-type: partial=p,universal=u
  17. veil-color: brown=n,orange=o,white=w,yellow=y
  18. ring-number: none=n,one=o,two=t
  19. ring-type: cobwebby=c,evanescent=e,flaring=f,large=l, none=n,pendant=p,sheathing=s,zone=z
  20. spore-print-color: black=k,brown=n,buff=b,chocolate=h,green=r, orange=o,purple=u,white=w,yellow=y
  21. population: abundant=a,clustered=c,numerous=n, scattered=s,several=v,solitary=y
  22. habitat: grasses=g,leaves=l,meadows=m,paths=p, urban=u,waste=w,woods=d

The dataset contains 8124 descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family. Each species is identified as definitely edible, definitely poisonous (or of unknown edibility and not recommended). Mine the mushroom data and find association rules that can be used to identify the edibility of a mushroom.

Consider the following **research question**: What are the characteristics of edible mushrooms? And of poisonous ones?

To answer the questions consider different parameters related with: length of rules, minimum support, minimum confidence. Find only rules relevant to edibility. Are there any redundant rules? Can you remove them?

The document you turn in should explain data preprocessing (if any), application of mining models (and use of different parameters), comparison of model's results and a conclusion. The conclusion summarises all the work done and what you have learned. A conclusion closes the assignment stating how you have achieved the proposed objectives. Comment on anything interesting about the assignment completed.