## **Final paper** (individual):

Write a short (less than 10 pages) report of a machine learning application:

1) Formulate a research question
2) Describe the data
3) Describe the methods
4) Describe the analysis and results: Explain how you applied your methods and interpret the results
5) Give a conclusion
6) References

You are required to **use at least two methods from the topics for which you previously did not explicitly have an assignment** (presentation or short report) for. You are free to use more topics, including topics not covered during the course. Check with us if you doubt about feasibility:

- Methods and data should be appropriate for the problem you address
- Originality (both with respect to data as method) is appreciated

Give your paper a **title** and do not forget to add your **name and team number** on the first page

A tentative outline of your report could look like this:

1) **Introduction**: Introduce the research questions, why it is interesting and how you are going to answer the research questions.

2) **Data**: Briefly explain the data that you have, where they are from and why they are relevant for your problem. If appropriate, describe relevant data properties and general data preparation steps.

3) **Methods:** Provide a short description of the methods that you use as well as a motivation for choosing these methods.

4) **Analyses & Results**: Describe how you apply the methods and present the results.

5) **Conclusions**: Briefly summarize your findings and the corresponding conclusions. Relate the outcomes to the research questions.

6) **References**

The report should be in pdf format and must be accompanied by R-code (in a separate zip file) that can be used to reproduce the results. Make sure to clearly annotate your code (use clear comments) and/or give a clear description on what the code does. We will not grade the code, but we may look at it to help assess the analyses and results described in your paper.

**Do not use an Appendix!**

**AI & Plagiarism:**

Do not use ai to write your paper. We would like to assess your performance. Note that all papers will pass through a plagiarism and AI check that includes previous student papers as well as publications. It also scans for likely AI usage. Do not copy complete sentences/paragraphs (and be aware that AI may do just that). Be careful not to get too inspired by textbook explanations. Always explain in your own words. (So: Making small artificial changes in copied sentences/paragraphs (for example just changing a variable name or the word order) does not make it your own words.

**Grading:**

We grade the papers by considering the 6 "ingredients" listed above, as well as the presentation (e.g., writing, use of tables and graphs etc.) of your findings. Finally, in addition to proper selection, description and execution, we also evaluate ambition, originality and appropriateness of data, methods and analysis.

**Required methods:**

The table below indicates per team the methods (at least 2) to choose from for members of different team. Note that you are free to use more methods, including methods that we did not treat during the seminar.

| TEAM | METHODS | | | | | |
|---|---|---|---|---|---|---|
| **1** | K-Means | K-Nearest Neighbors | Conditional Inference Trees | Boosting | Support Vector Machines | Global Interpretation |
| **2** | Hierarchical Clustering | Naïve Bayes | Decision Trees | Random Forest | Neural Networks | Local Interpretation |
| **3** | K-Means | Naïve Bayes | Decision Trees | Boosting | Support Vector Machines | Global Interpretation |
| **4** | Hierarchical Clustering | K-Nearest Neighbors | Conditional Inference Trees | Random Forest | Neural Networks | Local Interpretation |
| **5** | K-Means | Naïve Bayes | Decision Trees | Random Forest | Neural Networks | Global Interpretation |
| **6** | Hierarchical Clustering | K-Nearest Neighbors | Conditional Inference Trees | Boosting | Support Vector Machines | Local Interpretation |