Report title
Data Analysis and Machine Learning
Group 2

Otto Bervoets (S3417522)
Nick Boer (S5427894)
Jeroen Laarman (S3504778)
Student names + s-numbers

January 10, 2025

Abstract

This document is a template for the report and contains detailed intructions on the content of the report.

1 Introduction

1. What is the problem you want to analyze? State it in clear terms.
2. Background/relevance: why is it a problem? Where does it come from?
3. Type of problem, e.g., classification, regression

The report has the following structure. After having introduced the data set in Section 2 we describe in Section 3 the methods to analyze the data and validate the analysis. Etc.

2 Data description

1. Why this data, how does it help to answer your question?
2. Describe the data. Where did you get it from? Include dates, and e.g. the like. Ensure that if people want to repeat your work, they can find your dataset.
3. Just mention simple filters you applied
4. Exlain difficult filters, if this was necessary.
5. Include in any case some samples of the data so that people can see how things are organized.
6. Use graphs, and tables to provide overview, statistical insights such as size of the data.

3 Methods and validation

1. What method are you going to use? Where does it come from: ref to the literature, paper, book(s)?

2. Why is that method(s) suitable for you problem? To what extent, and what are its limitations?

3. How are you going to validate your analysis? e.g., cross validation? Why this method?

4. You are not bound to using the methods we discuss in the course. If you like to use deep learning, for instance, then that's fine too. However, in case you want to use a method that is not part of the course, mail the instructors and ask whether we agree. Make sure to prevent surprises at the end.

4 Analysis and insights

1. What can we learn/see from the data?
2. How confident are you about your claims?
3. Can you come up with an estimate of the expected gain (monetary, or other KPIs) by using your methods?
4. Include graphs (mostly) and tables (sometimes). Some people tend to skip tables. Include a short description in the caption of the figures. You should know that people often start with glancing over the figures, and skip most (or all) of the text. Hence, figures and captions are your main initial selling points.

5 Conclusion and summary

1. Main findings, quality/reliability of findings,
2. Evaluation of chosen methods, and extensions.

6 Hours per student

Names	Student A	Student B	Student C	Student D
Intro	0	0	0	0
Data	30	0	0	0
Methods	10	30	0	0
Analysis	10	0	10	0
Conclusion	0	0	10	5
Total	50	30	20	5

A Appendix

1. Include here the core parts of your code, and explain how it works.
2. Don't include long tables with data, because we will just skip them.

Python Code

```
a = 10
b = 5
print(a*b)
```

R Code

```
a <- 5
b <- 10
b*a
```

B Feedback on report

You can use this page to organize your feedback to the other group, so that the other group can improve their work.

Guidelines:

1. Be respectful.
2. When reading, ensure you really understand each and every step.
3. If you particularly (dis)like certain aspects of the other's report, use that to improve your own report. Try to figure out why you (dis)like that aspect.
4. Give a grade at the end: bad, average, good, excellent ($< 4, 6, 8, 10$).

B.1 Problem and method

1. Do you understand the problem, motivation, chosen method?
2. Include advice on what to improve, and why. Better method perhaps?

B.2 Analysis

1. Do you understand the problem, motivation, chosen method?
2. Include advice on what to improve, and why. Better method perhaps?
3. Ideas on to improve the type statistical analysis, quality of analysis.

B.3 Style

1. Quality of writing?
2. Presentation: use of graphs and tables.