Data Mining, Machine Learning, and Deep Learning

Final Project Guidelines

Dataset

- For the data analysis project, you can **choose** a custom dataset of your **own choice**.
- Chosen data must be suitable to answer all the questions.
  - We discourage you to use the data sets that have been used during the coursework.
    - You can choose a dataset used during the course work, but it must have some strong reasons.

- Follow the below rules while choosing the data.
  - # of data points: min. few thousands.
  - # of data columns: a good number of columns.
  - Missing values: Must not have too many missing/NA values.

Computing Resources

- Use Ucloud service : https://cloud.sdu.dk
- Microsoft Azure: https://azure.microsoft.com/en-us/free/students/
- To get a Windows server for your exam/project, fill the form (link) to access to the server hosted at Data Science cluster. If you did not get the server in two working days then email to Kiran Kocherla <kkk.digi@cbs.dk>.
- You can also Google Colab (Google Colab).

About Project Report Template

- Title
- Author
- Abstract One sentence each for each question below:
  - What was the topic?
  - What was the problem formulation?
  - What was the research question?
  - What were the concepts?
  - What was the dataset and what were the main data analytics methods and tools?
  - What were the most important results in terms of meaningful facts, actionable insights, and valuable outcomes?
  - What are the conclusions and recommendations?
- Keywords: at least five

About Project Report Template

- Introduction
- Motivation
- Research Question(s)
- Related Work
  - Academic Articles on Similar Research Questions (Hints: Use Google Scholar)
- Conceptual Framework
  - Concepts of relevant to research problem data analytics methods and techniques.
  - Problem statement with problem modelling (if relevant)
- Methodology
  - Dataset Description
  - Data Analysis Process with suitable Diagram (if any)
  - Data Pre-Processing: Methods/Tools (if relevant)
  - Data Filtering, Transformation and Combination (if relevant)
  - Data Analytics: Modeling, Methods and Tools
  - Model complexity analysis (such as running time) compared to baseline model.

About Project Report Template

- Results
  - Meaningful Facts (Describe the result with explanation of the reason(s))
  - Actionable Insights (do some Interpretations/predictions from the result)
  - Valuable Outcomes (Applicability of the result) (Hints--How this information can be used to make some recommendations)
- Discussion
  - Answers to the Research Question(s)
  - Implications for Research / Learning Reflections
  - Limitations of the dataset/work (if any)
- Conclusion & Future Work
- References
- Appendices