Project instructions
2-INF-150: Machine Learning, Fall 2024
Lecturer: Tomáš Vinař

Project proposal deadline: 18.11.2024
Final report deadline: 7.1.2025

The task for the project is to:

- choose a problem to which it is possible to apply machine learning methods – it can be almost anything, partial work towards your thesis is welcome
- implement one or more methods to solve the problem, using methods that were taught in this course (or their appropriate extensions)
- evaluate these implementations on real or simulated data; using data assembled by you instead of "stock" data sets is welcome
- compare your implementation to a simple baseline or publicly available solution
- standard benchmark data sets or kaggle challenges are **not** a good problem for this project

The project proposal is a one page document which contains description of the problem, the data you plan to use for the training and testing, and the plan for evaluation (evaluation measures, baseline, etc.) We do not expect that you describe a concrete plan for solution of the problem. **The project proposal should be uploaded to google classroom as a PDF file,** you will receive our feedback and a few bonus points.

The project report is a 5-15 page document (the content is much more important than length). Report must contain:

- Title and name of the author
- Abstract (1-2 paragraphs summarizing goal, methods used, and results achieved)
- Introduction (problem description, how did you get the data, their brief characteristics including information such as size of the training and testing set, number of positive and negative examples, feature descriptions, etc.)
- Context (what Is the current state of the art, including references to how other people approached this problem, if relevant)
- Description and justification of methods that you used
- Brief description of techincal issues that you have encountered
- Experimental evaluation (tables, graphs, their interpretation and what conclusions do you draw from them)
- Conclusion (Was your project succesful? Why? If you started again from scratch, what would you do differently?)
- References

It is OK to fail. Well presented unsuccessful attempt can still get full grade if it contains a good analysis of why did it fail. In practice, development of successful models takes several failed iterations; we expect that you do one such iteration and if it fails, that you carefully analyze the reasons for the failure.

The project report should be submitted to google classroom as a single PDF file. Moreover, it should be accompanied by an electronic appendix (on github or elsewhere – please provide a link) which should contain:

- Source code of your own programs with appropriate comments
- Data sets or link to the data sets if they are large
- readme.txt file describing individual files and how to use them
- **the results in your report should be straightforwardly reproducible by using the files from this appendix**

Cheating. Do not copy from others or from the internet, acknowledge all sources, and follow guidelines regarding collaboration and use of AI tools posted on the course web page. Note that if you submit a solution violating these rules, you will receive -100% points which in the case of the project means automatically failing the course.