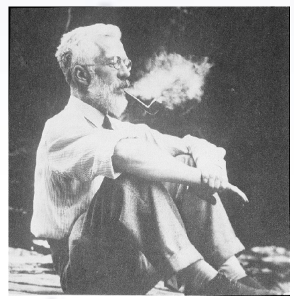
Applied Machine Learning

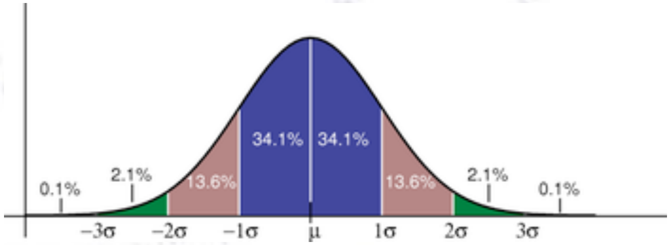
Comments on Grading








Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense - Machine Learning is a sharpening of it!"

Final Project Scores

The grading input data

There were 22 projects in total: 14+8 on the 15th and 16th.

We were 6 teachers, who all gave 5 grades [0,10] on the points of evaluation:

- Complexity of problem and depth of solution (incl. appendix)
- Choice of methods and arguments behind
- ML performance and own evaluation of it
- Clarity of presentation
- Implementation, technical details, optimisation, etc. (your appendix)

Thus each project got 30 scores from the teachers. We decided to weight teachers equally, and the five points of evaluation as: [0.3, 0.15, 0.25, 0.15, 0.15]

In addition, each project got on average ~60 scores from fellow students.

The following are to show you the cross checks that we've gone through to even out differences, and evaluate as accurately as possible.

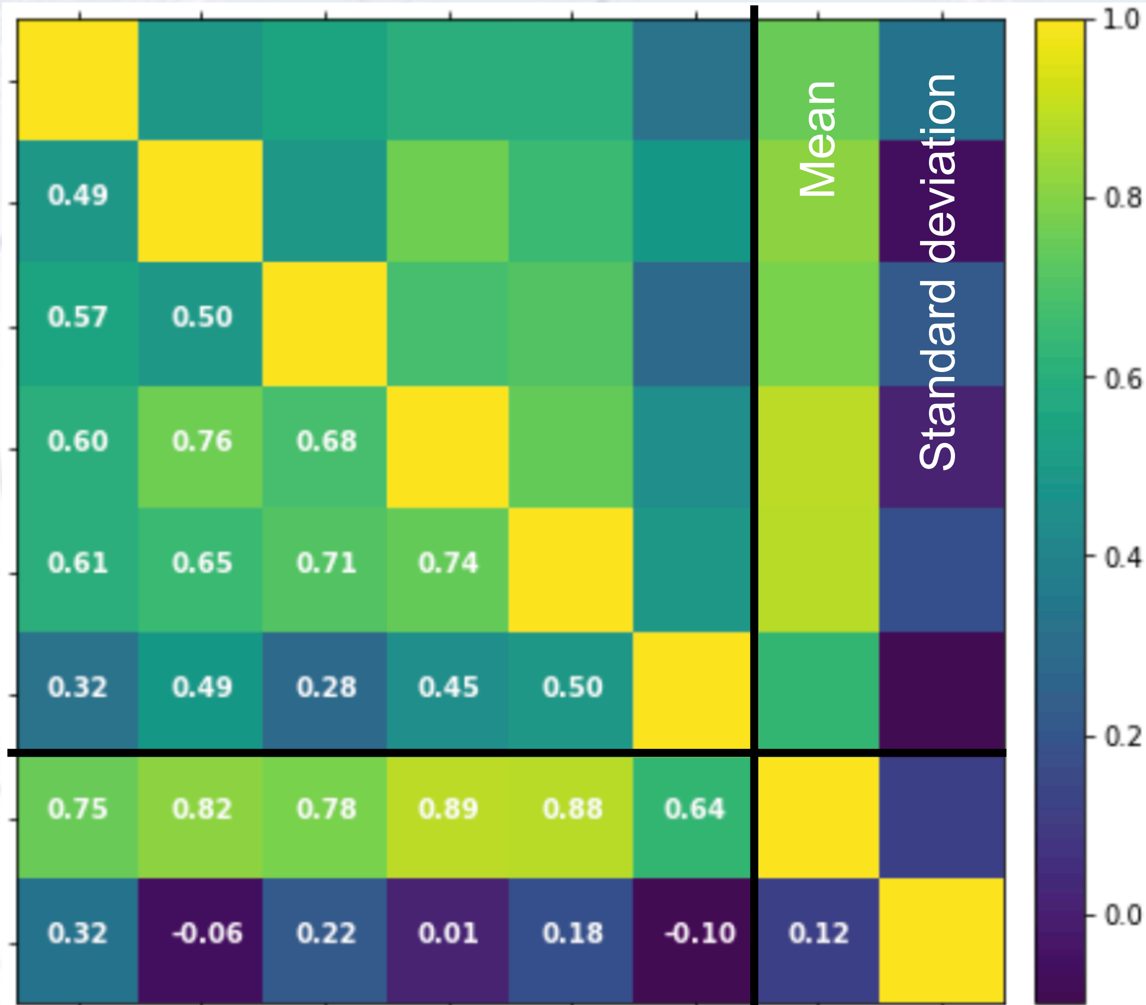
Checks between teachers

The typical std. between the six teacher averages was about 0.57 (largest: 1.02).

It is a difficult task to evaluate! That is why we take averages.

We don't agree perfectly among ourselves, but only one combination of persons have no (Spearman) correlations!

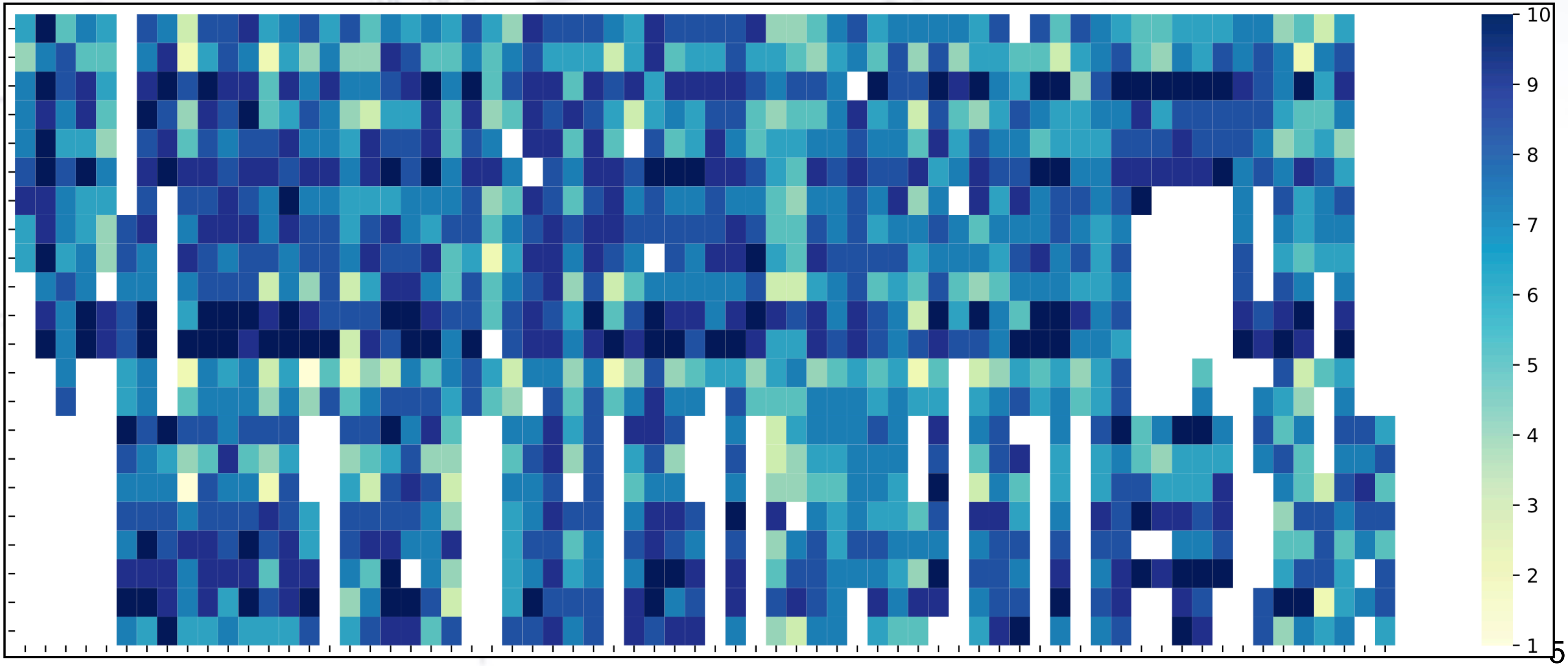
The mean and std. are included in the last two columns/rows, respectively.



Student gradings

One evaluation point was how good YOU were at evaluating others ML work.

Below are your grades to all the projects (white: Absent).

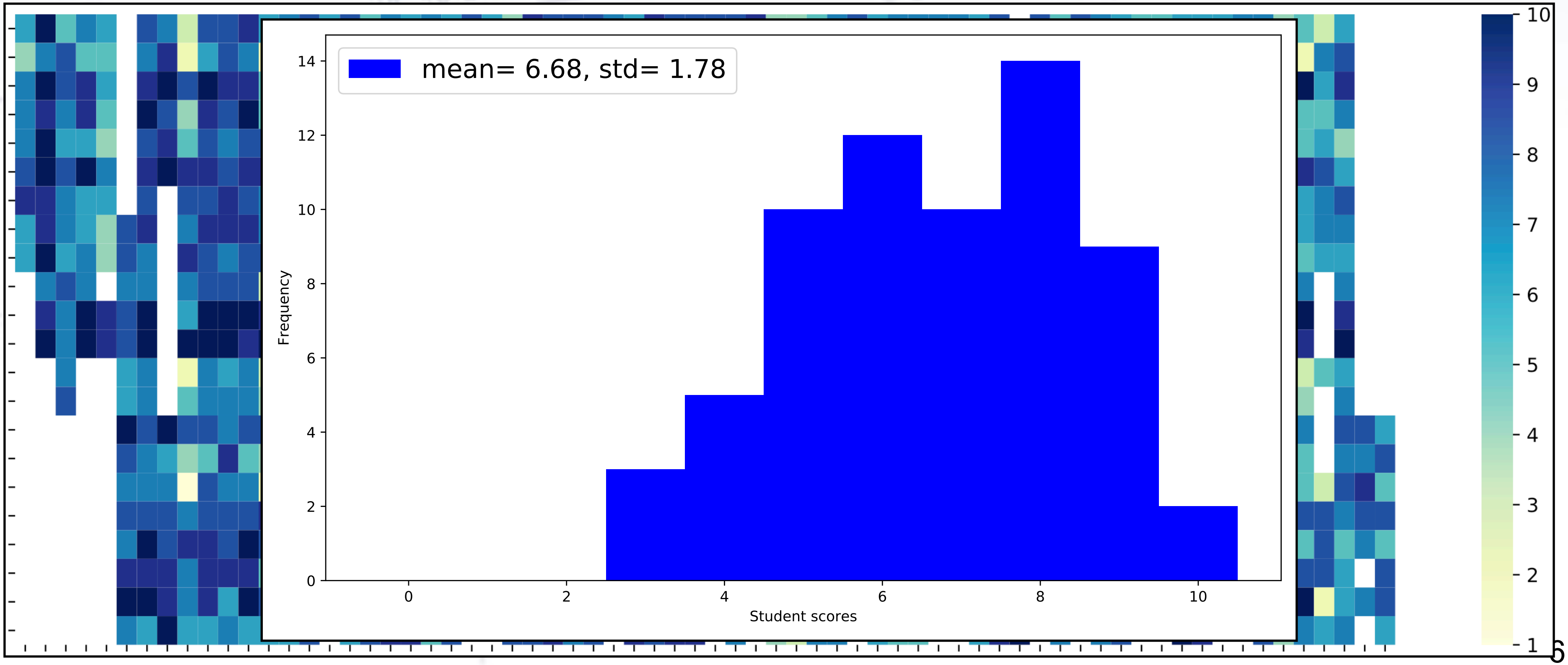


Student gradings

One evaluation point was how good YOU were at evaluating others ML work.

Below are your grades to all the projects (white: Absent).

Students (also) don’t agree - in fact less! Below is a distribution of scores.



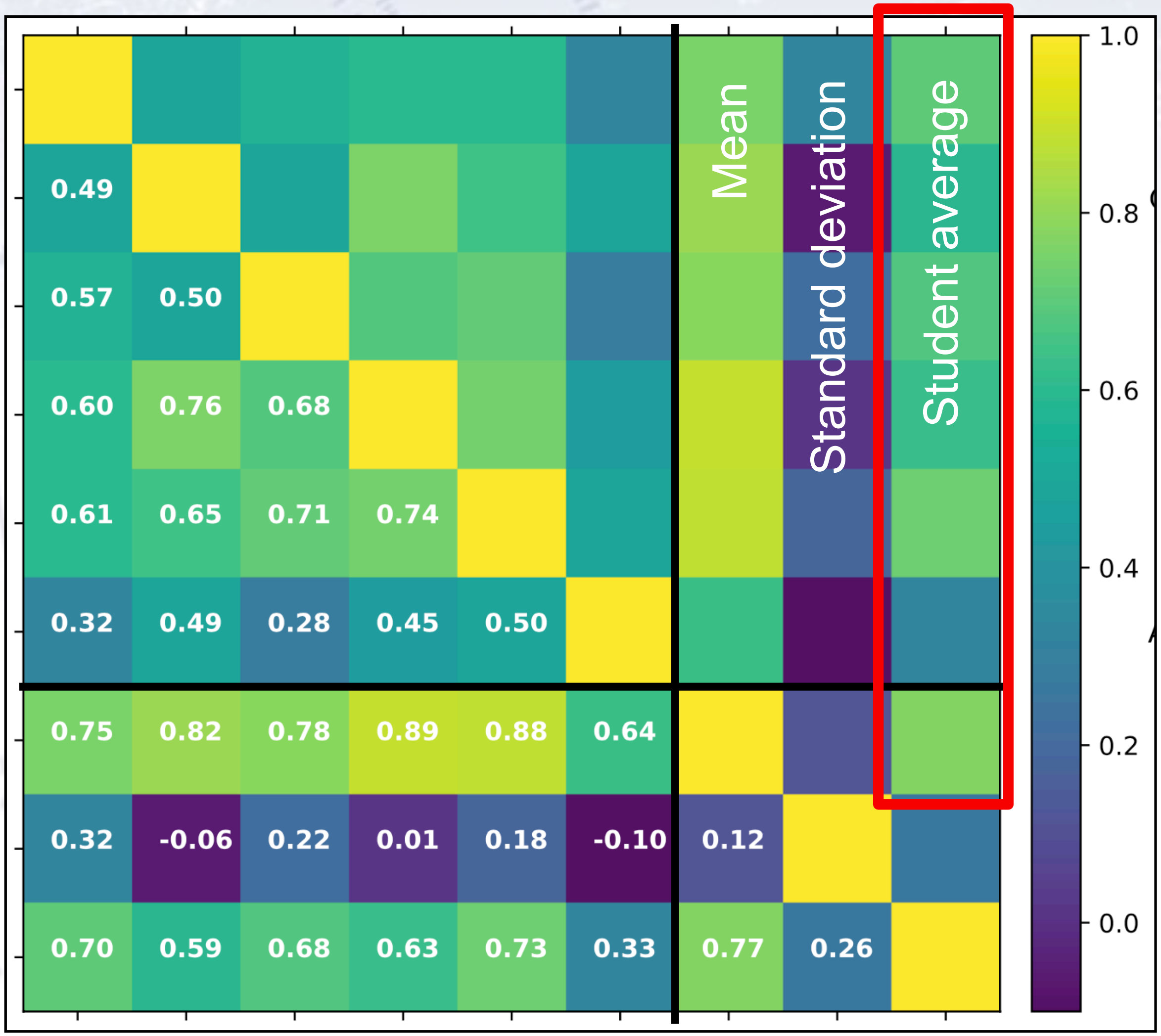
Teachers vs. student average

Including the average of the student evaluations, we can compare….

I'm happy to see, that teachers to a large extend agree with students.

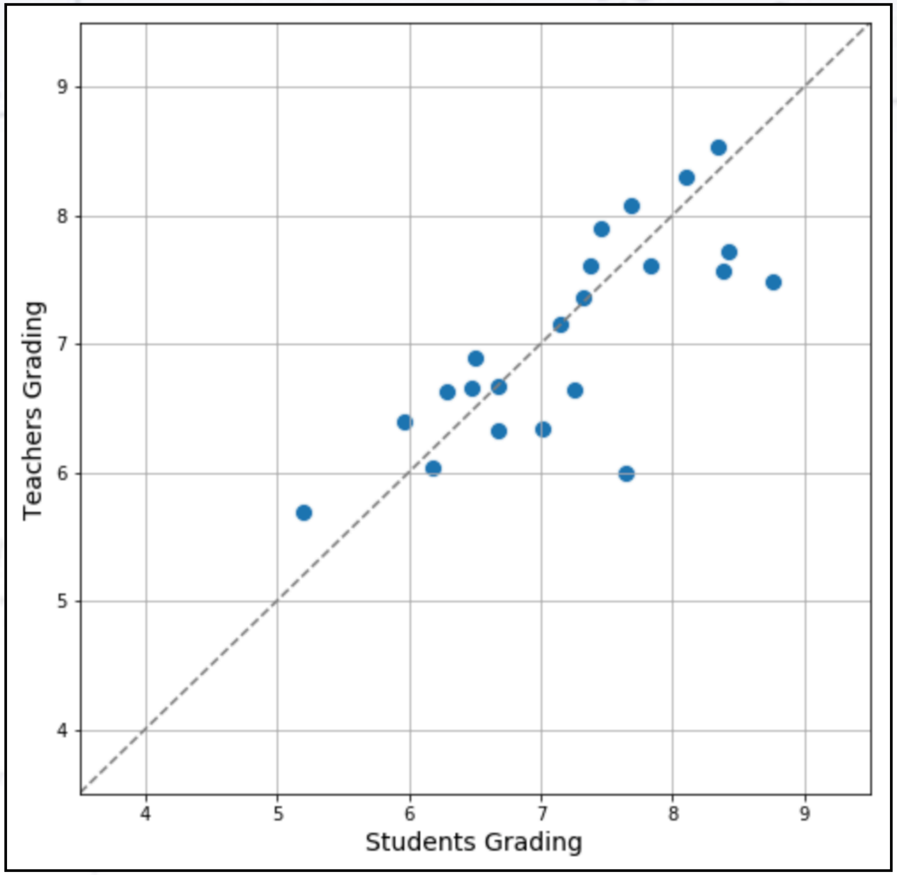
The correlation between the teacher average ("Mean:") and the student average is very high.

No single teacher do not correlate significantly (and positively) with the student average.



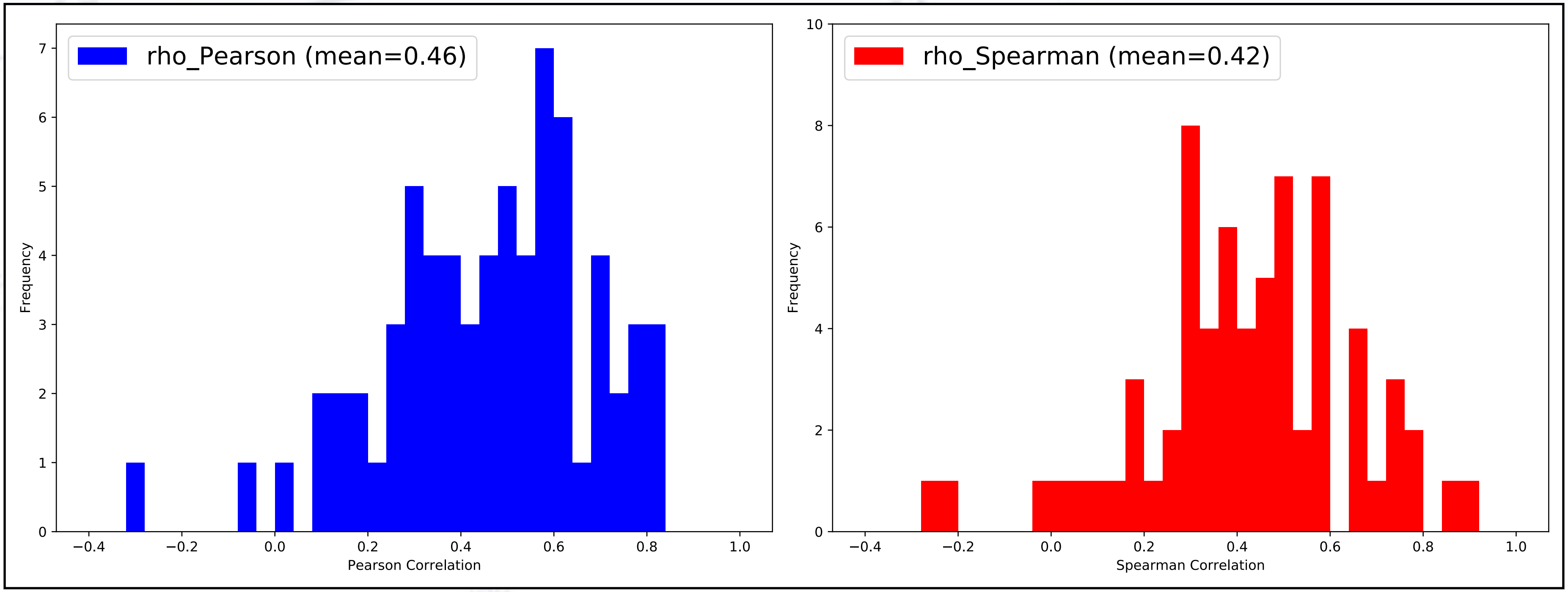
Teachers vs. student average

The correlation can also be seen for the single projects. Generally, teachers graded a bit higher than students, and the correlation is very clear (77%). If one project is not considered, the correlation is very high (85%).



Student evaluations

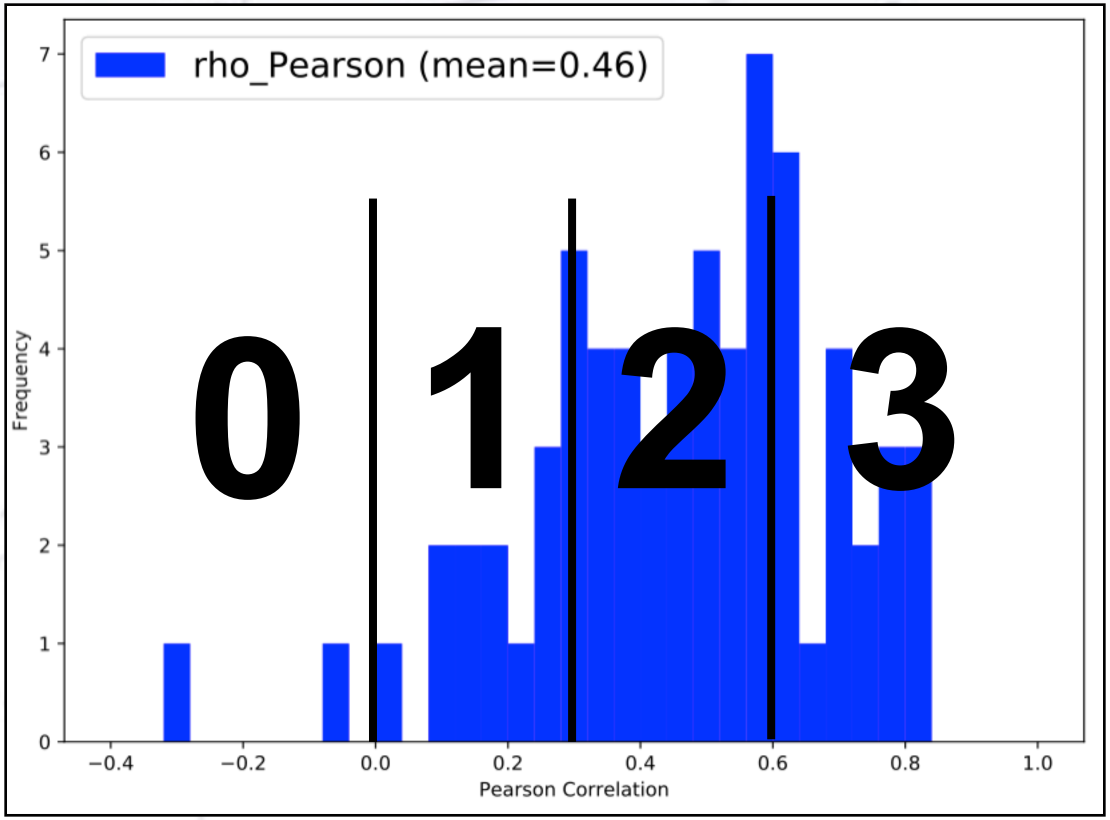
The student evaluations were scored by considering the correlation with the teachers average.



Student evaluations

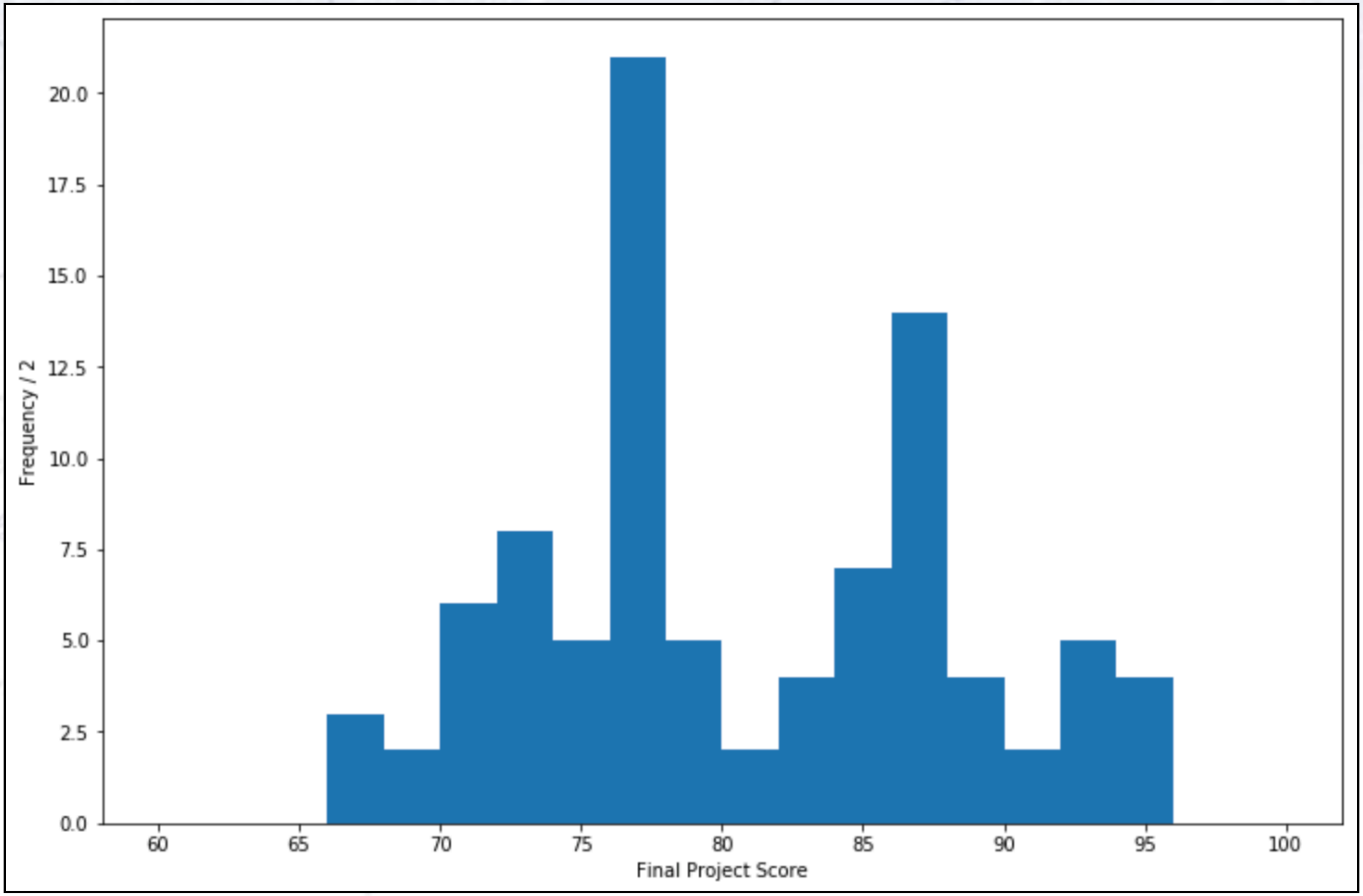
The student evaluations were scored by considering the correlation with the teachers average.

Almost all (Pearson) correlations were positive, and generally around 0.5.
We gave scores [0,3] as illustrated below.



Final project scores

The scores obtained by taking the teachers (calibrated) average multiplied by 10 and then added 10. The final distribution of final project scores is shown below.

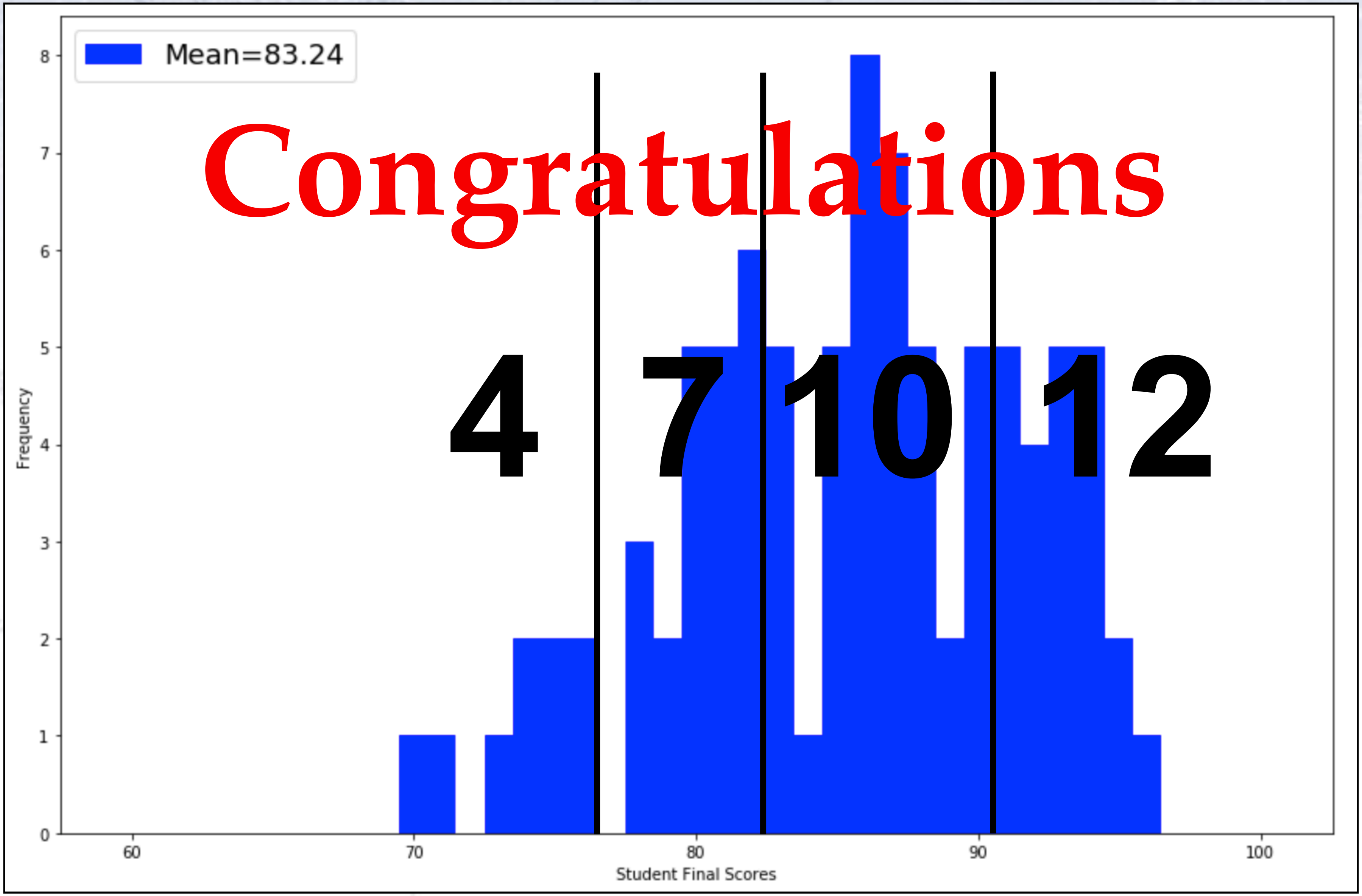


Final Course Scores

Final course scores

The scores from the initial ("small") project and the final project (and the ML scoring) was put together as prescribed, and produced the following scores/grades:

Course scores and grades



Bonus Slides

Summary of experiences 2021

During the final presentations, a summary of experiences were:

- Start with a quick-and-dirty method (BDT!) and get it to work. Then refine it.
- Computing power is important.
  - ✓ Parallel computing is good. GPUs are great.
- Pre-processing is very important. So is data inspection.
  - ✓ Use e.g. quantile transformation to make distributions "nice".
  - ✓ Check if data is unbalanced, sparse, or otherwise needs re-weighting.
- HyperParameter (HP) optimisation is cumbersome.
  - ✓ Specifying HPs is nice: Both for reproducibility and as a help to others!
- Diversity in ML "phase space" is immense and overwhelming.
  - ✓ Adam, CNN, RNN, DNN, pDNN, ??NN, One-Hot encoding, etc.
  - ✓ Manage to navigate in this jungle and find any good solutions.

Generally, everybody felt, that they could actually get ML to work and solve problems with it. We hope that this was your impression too.