General Guidelines, Assignment 1

Contents

The Tasks

The first assignment consists of three tasks, of which you choose two. Each task is about applying machine-learning frameworks to answer a research question. You find the tasks' descriptions here: https://www.dropbox.com/sh/ffldgww49pon9h/AAB5HXRn16tzWojqf15i_jhma?dl=0

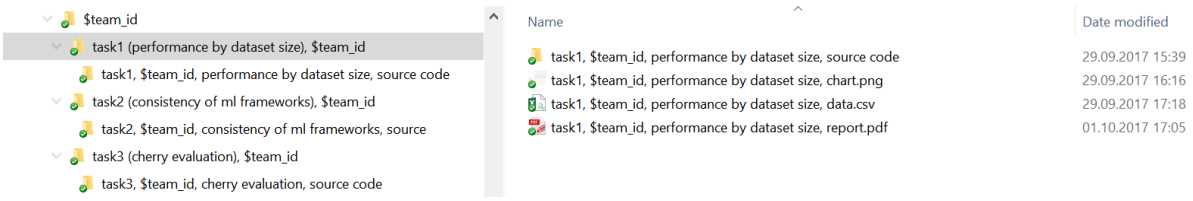
Read the descriptions carefully, and ask questions in the Blackboard forum.

Submission

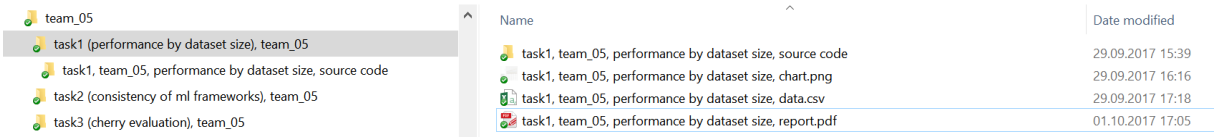
For each of the two tasks that you choose, you are required to submit

1. A CSV file containing your results
2. One or two charts that illustrate the results (PNG format)
3. A report that answers the research question (PDF)
4. The source code you wrote to accomplish the task

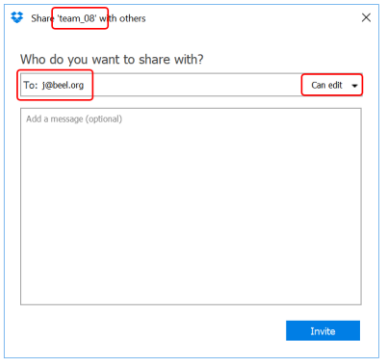
To submit the assignment, you create a folder named "$team_id" (e.g. "team_05"). This folder contains sub-folders for each task. Each of these sub-folders contains the four deliverables. The image below illustrates the structure and exact naming (you only submit the folders for the two tasks you picked, not all three). Do not submit any other files (we don't need copies of the guidelines, or your GitHub configuration files etc.).



Ideally, you download and use our template folder structure https://www.dropbox.com/sh/sa54zadx3ito5ae/AABP2jSsfsC5rDhlqmE97n9ka?dl=0 . This means, you copy the template files to your computer, remove the task you don't do, rename the remaining files and folders properly and replace the files with your own results.



To submit your assignment, you share the $team_id folder with us (please share with j@beel.org).



We strongly suggest you share the "$team_id" folder several days before the deadline with us, even if the folder should not contain any results yet. As confirmation, we will upload an "invitation received.txt" file into the shared folder. In case you don't receive that confirmation within 2 working days, you can contact us and we can fix the problem before the deadline.

Data

1. We want to ensure that when students use the same dataset, they really use the same dataset. Therefore, please only use datasets for your assignments that are stored in our Dropbox folder https://www.dropbox.com/sh/euppz607r6gsen2/AACcVFIxekZXYTEM5ZsMSczEa?dl=0 . Currently the folder contains only a few datasets. You are welcome to suggest new datasets (the lecture slides contain some links where you can find datasets). Actually, you have to suggest new datasets, because the existing ones will not be sufficient to solve all tasks. To suggest a new dataset, visit the Blackboard forum https://tcd.blackboard.com/webapps/discussionboard/do/forum?action=list_threads&course_id=_38354_1&forum_id=27479&nav=discussion_board&conf_id=_52338_1&content_id=_915577_1 and we will then add it to the Dropbox.
2. If the instructions say "use 2 (or more) datasets of your choice", do not use "The SUM dataset (without noise)" *and* "The SUM dataset (with noise)". You can use one of them, but not both.
3. Unless really necessary, you do not need to do any data cleaning, feature selection, feature transformation, etc. However, if you want to do it, feel free to do so (please describe what you did in your report under "additional information").

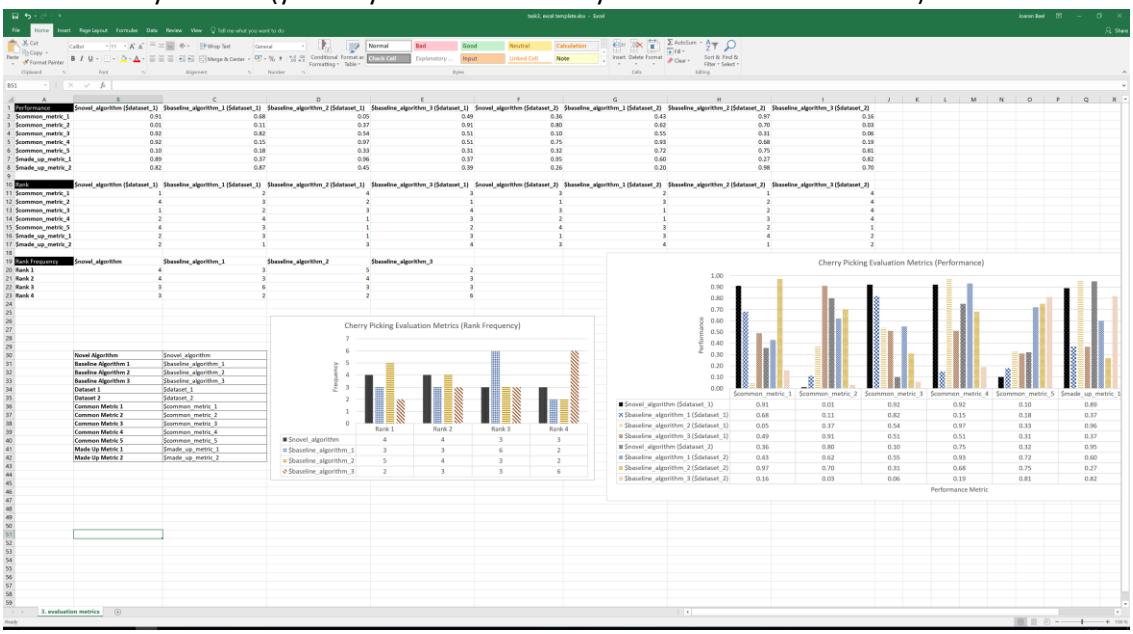
The CSV files

When creating CSV files, please consider the following

1. Create CSV files that contain exactly (!) the same rows and columns as in the templates https://www.dropbox.com/sh/sa54zadx3ito5ae/AABP2jSsfsC5rDhlqmE97n9ka?dl=0. Do not add or remove any columns, do not use different names, …
2. Use a semicolon (;) as separator
3. Do not use semicolons in any of the cells.

The Charts/Images

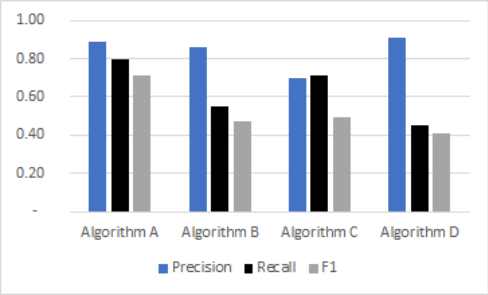
1. Store images in PNG format
2. The resolution must be „reasonably" high, i.e. all numbers and letters must be easily readable at 100% zoom.
3. If you have access to Microsoft Excel, you can create the charts easily. Just download the Excel templates we prepared https://www.dropbox.com/sh/waw4j78gjf5g3vu/AACHFcfg9RW8r1z9zrswQKu6a?dl=0, copy & paste the results from your CSV file into the Excel template, and then the charts are automatically created (you only need to find a way to save the charts as PNG).



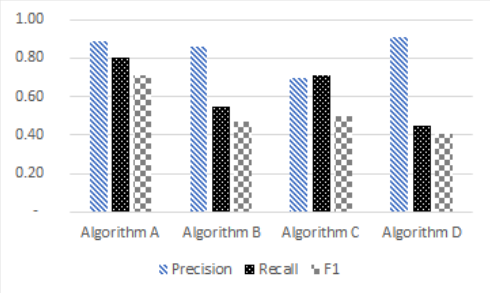
If you don't have access to Excel, you need to create the charts yourself. In that case, please consider the following.

1. Use different colours and patterns (for B/W printing)
   Keep in mind that we might want to print your results, any maybe in black and white. Therefore, when you use colours in your charts, do always use additionally different patterns for bar charts and lines (e.g. dotted and dashed lines).

a. Bad (imagine how this will look when you print it black-and-white – it will be barely distinguishable):

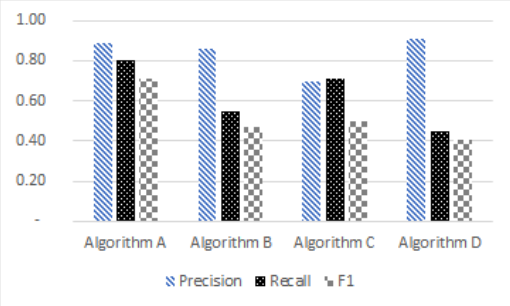


b. Good

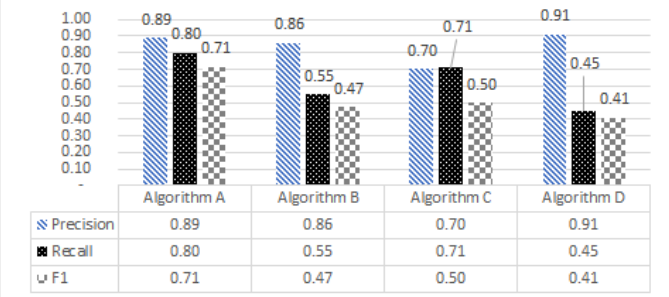


2. Provide specific numbers either as label, or as data table.

a. Bad (readers have to approximate what e.g. precision for Algorithm A was)

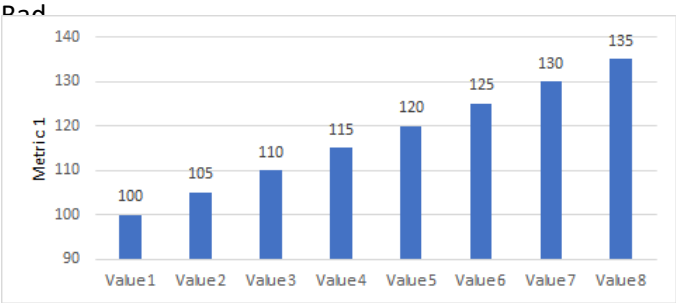


b. Good (though you don't need labels *and* a data table; pick one of the two options)

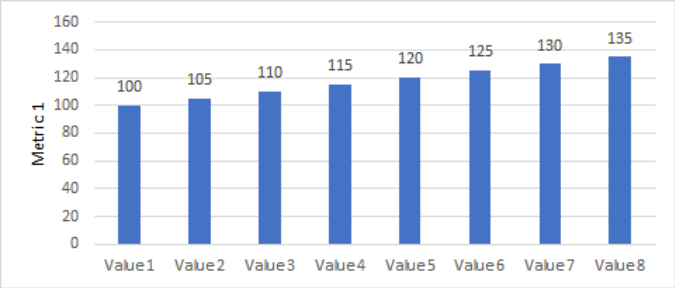


3. Start the y-axis with 0

a. Bad

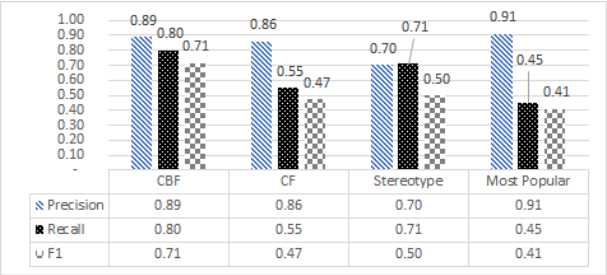


b. Good

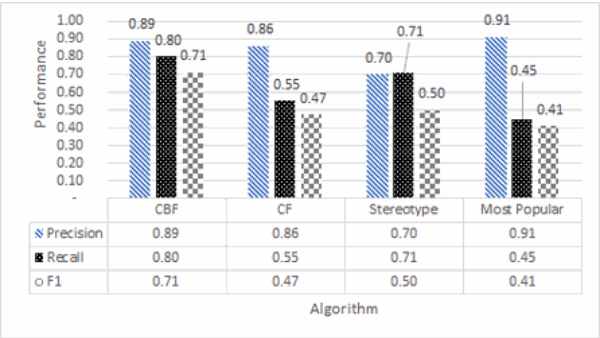


4. Always label the x and y axsis

a. Bad



b. Good



The Report

We prepared MS Word templates that you can use for writing the reports https://www.dropbox.com/sh/waw4j78gjf5g3vu/AACHFcfg9RW8r1z9zrswQKu6a?dl=0. However, you can also use your own document format as long as the content is as specified.

A good answer to a research question:

1. Describes the results objectively. For instance, “Our analysis showed that algorithm B performed better than algorithm A, regardless of the metric being used.” There should be no room for discussions here. If someone would say “I disagree”, then either you or the ‘someone’ would have done something severely wrong. It is also important that you are as

specific as possible. Do not write “As can be seen from the table, algorithm A performed better than algorithm B”. Instead, be specific and write e.g. “As can be seen from the table, algorithm A performed better than algorithm B with a precision of 0.67 vs. 0.82”.

2. Discusses the findings and draws (potentially subjective) conclusions. For instance, “Based on the results, it seems that algorithm A is more suitable for the classification of images than algorithm B”. Of course, ideally, people will agree with your conclusions, but it would be ok in some circumstances, if people would draw different conclusions or say “But…”
3. States its limitations. For instance, “However, it should be kept in mind that we only used low-resolution images. It would be interesting to see how the algorithms perform on high resolution images”.

The Source Code

When writing source code, please adhere to the developer guidelines https://www.dropbox.com/s/i6dh94mapquaf7g/Developer%20Guidelines%20%28Simple%29.pdf?dl=0 . In addition, consider the following

1. Comment your course code properly, so that we can easily understand what you did.
2. Submit only source code, no binaries, external libraries, etc.
3. It is particularly important that you follow the “single point of definition” and “nothing should be hardcoded” rules in the developer guidelines. For your assignment this mean that you should have a config file (e.g. config.ini; settings.json; config.py; …) that specifies the most important variables and hence allows us to run your scripts if necessary without us having to search for ages where in your source code you specified e.g. the path to the data. The config file should include in particular:
    a. The path and filename to the dataset(s) (e.g. "c:\dropbox\shared\datasets\the sum dataset\dataset.zip")
    b. The path to a temporary directory where the processed data is stored (e.g. "c:\temp\group assignment\tmp\"
4. You should write your scripts so that they directly work with the datasets provided in Dropbox. For instance, if the dataset is zipped, then your script should include the unzipping of the data (and the unzipped files should be stored in some temporary folder).

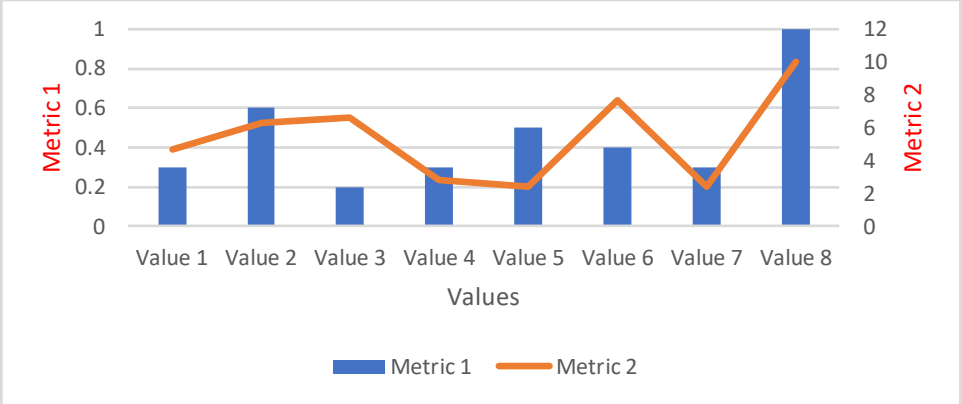
Frameworks

1. If you use different machine-learning frameworks for the same task, ensure, as far as possible, that you use the same parameters in the different frameworks (e.g. do not “normalize” data in one framework, but “standardize” it in another).
2. Do not use cloud-based machine-learning frameworks, but only frameworks you can locally install (ideally open-source, but commercial frameworks are also fine if you have a license).
3. If you want, you can also use deep-learning frameworks (e.g. TensorFlow).

Evaluation Metrics

1. If not stated otherwise, a “metric” always relates to a measure of effectiveness (precision, accuracy, RMSE, …) and not to some other measure such as runtime or storage requirements.
2. Often, you will be asked to calculate two metrics or more. In some cases, different metrics may not have the same scale. For instance, while most metrics are between 0 and 1, some might have higher values. In that case, please either scale all metrics to the same scale (0-1),

and make this clear in the naming (e.g. “RMSE (scaled)”), or use two axes (see image).



Misc

5. Whenever the instruction says that you should choose a few out of several options, then you should choose options being different to each other. For instance, if the instructions say “Use three algorithms of your choice”, then you shouldn’t choose linear regression, logistic regression, and stepwise regression. Instead, you should choose e.g. logistic regression, a decision tree, and a support vector machine algorithm (if the data allows the application of these algorithms). Of course, if you are required to use e.g. 8 metrics or algorithms then some overlap or similarity is ok.
6. Sometimes, it might be necessary to use more options than specified in the instructions. For instance, if the instructions say “The goal of this task is to find out if different evaluation metrics provide different assessments of an algorithm’s effectiveness. Calculate four evaluation metrics for algorithm A and B. Try to find metrics where once algorithm A, and once algorithm B performs best”. In that case, you might need to calculate e.g. 10 different metrics to find one that says algorithm A performs best, and one that says, algorithm B performs best. However, in your report, you present only results for the four most illustrative metrics. ← This is actually bad science. It’s called “cherry picking”, but for the sake of learning it’s ok.
7. Whenever you see a variable, indicated by a dollar sign $, then you need to replace that variable with the proper name. For instance, “$metric_1” would have to been replaced with e.g. “RMSE”, and “$team_id” would have to been replaced with your team id, e.g. “team_63”.

Marking

The first assignment is worth 10% of the overall marks for the module. The two tasks of the first assignment are weighted equally, i.e. each is worth 5% of the overall mark for the module. The marks for each task will primarily be based on the report, particularly

1. The preciseness of describing the results
2. The plausibility of your conclusions
3. The explanation of limitations

In addition, the source code quality is marked, i.e. the adherence to the “development guidelines”; the appropriateness of the algorithms, metrics, datasets etc. that you chose is marked. However, the weight is rather low, as this assignment is thought to make you familiar with the different frameworks, algorithms etc. and not to check how skilled you are in choosing and tuning the proper algorithms etc. The writing style is also marked.

Here is an illustration of the marking scheme.

<table>
<tr><td rowspan="6">Course Work [40%]</td><td rowspan="4">Assignment 1 [10%]</td><td rowspan="3">Task 1 [5%]</td><td rowspan="3">Report [2.5%]</td><td>Description of Results [1%]</td></tr>
<tr><td>Conclusions [1%]</td></tr>
<tr><td>Limitations [0.5%]</td></tr>
<tr><td>Source Code [1.25%]</td><td>Appropriateness Algorithms [0.5%]</td><td>Writing Style [0.75%]</td></tr>
<tr><td>Task 2 [5%]</td><td></td><td></td></tr>
<tr><td>Assignment 2 [10%]</td><td></td><td></td><td></td></tr>
<tr><td>Assignment 3 [20%]</td><td></td><td></td><td></td></tr>
<tr><td>Exam [60%]</td><td></td><td></td><td></td><td></td></tr>
</table>

We will probably not check the source code for all tasks for all teams but pick a few randomly. In that case, the marking is as illustrated below

<table>
<tr><td rowspan="6">Course Work [40%]</td><td rowspan="4">Assignment 1 [10%]</td><td rowspan="3">Task 1 [5%]</td><td rowspan="3">Report [3%]</td><td>Description of Results [1.25%</td></tr>
<tr><td>Conclusions [1.25%]</td></tr>
<tr><td>Limitations [0.5%]</td></tr>
<tr><td>Appropriateness Algorithms [1%]</td><td>Writing Style [1%]</td></tr>
<tr><td>Task 2 [5%]</td><td></td><td></td></tr>
<tr><td>Assignment 2 [10%]</td><td></td><td></td><td></td></tr>
<tr><td>Assignment 3 [20%]</td><td></td><td></td><td></td></tr>
<tr><td>Exam [60%]</td><td></td><td></td><td></td><td></td></tr>
</table>

If technical specifications are not met and cause us to spend more time for marking the assignment than necessary (e.g. due to improper file naming, or incorrect image formats, no config file, …), we will deduct a significant (!) amount of marks for that task.