ML Research Assignments 2018/19   
Instructions

These instructions are preliminary. They may change during the next couple of days.

Version 1.1, 28. Sept. 2018  
Joeran Beel and Douglas Leith

[Overview 2](#_Toc525909070)

[Deliverables 2](#_Toc525909071)

[Source Code [0%] 2](#_Toc525909072)

[Report 2](#_Toc525909073)

[Cover Sheet [0%] 2](#_Toc525909074)

[Content [85%] 3](#_Toc525909075)

[Abstract [0%, Optional] 3](#_Toc525909076)

[Individual Title and Authors [0%] 3](#_Toc525909077)

[Introduction [10%] 3](#_Toc525909078)

[Related Work [10%] 3](#_Toc525909079)

[Methodology [30%] 4](#_Toc525909080)

[Results & Discussion [30%] 4](#_Toc525909081)

[Limitations & Outlook [5%] 4](#_Toc525909082)

[Acknowledgements [0%, Optional] 4](#_Toc525909083)

[References [0%] 4](#_Toc525909084)

[Design / Layout / Writing [15%] 4](#_Toc525909085)

[Submission 5](#_Toc525909086)

[Questions? 5](#_Toc525909087)

# Overview

You need to complete 2 research assignments as part of the course work. Both assignments will cover a research question that you can pick yourself and that you need to answer through your own experiments with some machine learning libraries, programming languages and datasets that you can choose freely. In both assignments, you need to answer the same question.

The process is as follows

1. Find yourself a team (3 students); feel free to use the blackboard forum to ask other students to join you.
2. Read the project descriptions <https://1drv.ms/w/s!AtfAgPR4VDcEqI5tuSs9a5sEsCJ4Lw>
3. Feel free to propose your own project ideas in the Blackboard forum.
4. Submit your team and project preference <https://1drv.ms/x/s!AtfAgPR4VDcEqI4ScG8k0qVgIYezXQ> even if you have not found a team
5. Complete the project and submit the first report (i.e. assignment 1)
6. Receive feedback and marks for your first report
7. Improve the project, i.e. answer the same research question again but in more detail, incorporate the feedback you have received from us, and submit the second report (i.e. assignment 2).

All relevant files (templates, project descriptions, schedule, …) are stored here: <https://1drv.ms/f/s!AtfAgPR4VDcEqI4RPjF1MM8wxoe8cA> (read only)

# Deliverables

## Source Code [0%][[1]](#footnote-1)

We want you to use a source code repository such as GitHub throughout your work. Name your repository “ML1819—task-$taskID--team-$teamID”, e.g. *ML1819--task-201--team-05*.

Push your work regularly to the repository. The repository must demonstrate your continuous development activities.

You have two options for giving us access to the repository

1. Make the repository public
2. Use GitHub, make the repository private, and give admin rights to the GitHub user “Joeran”

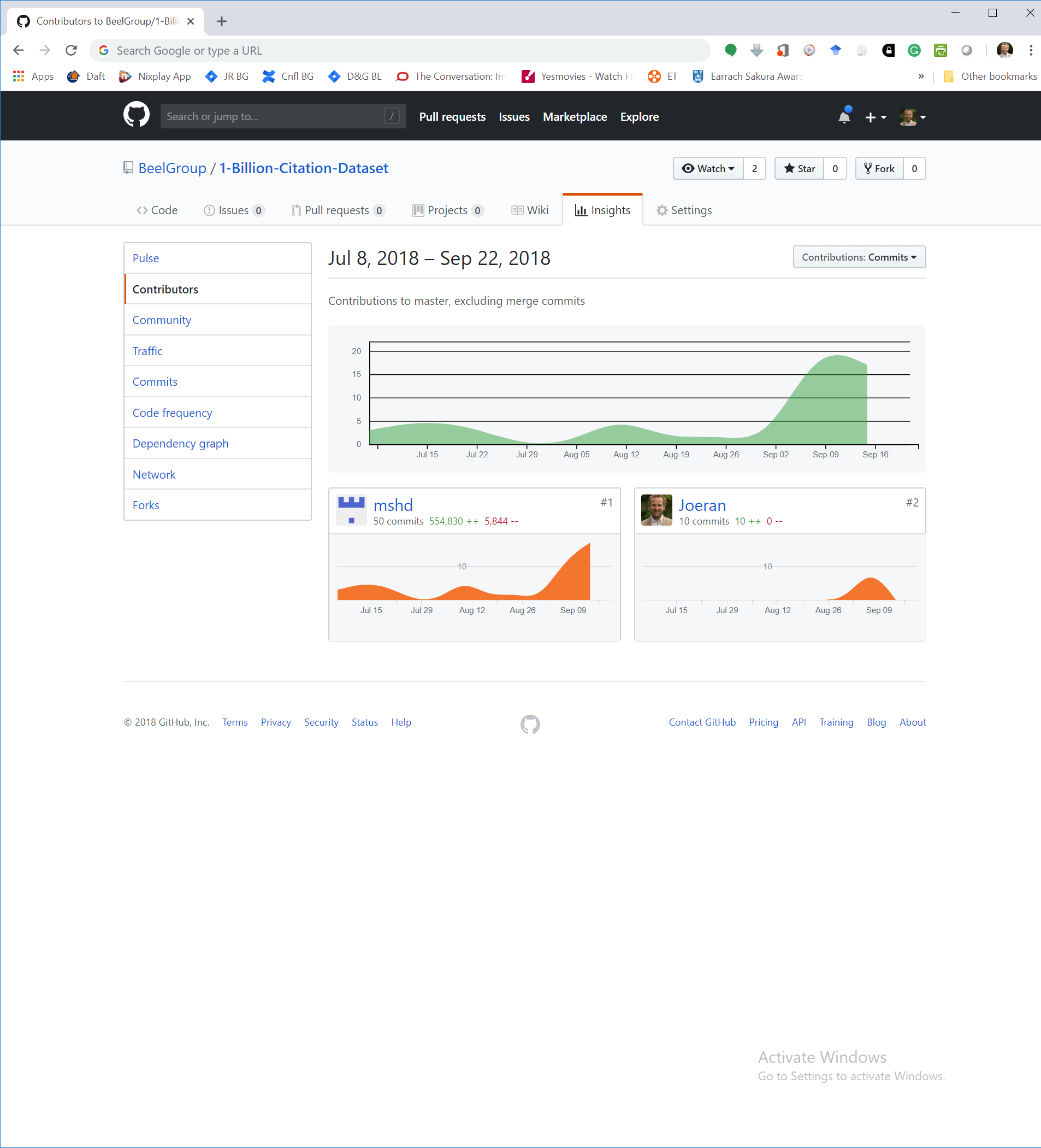
## Report

1,000-word limit (hard) for the first research assignment, 1,500-word limit (hard) for the second research assignment. Cover sheet, title, author names, tables and figures, references, acknowledgement, and appendix do not count towards the word limit.

### Cover Sheet [0%][[2]](#footnote-2)

If needed, the cover sheet can expand over multiple pages, though we would prefer a single page. Include the following information.

* 1. Title (“ML1819 Research Assignment 1” for the first assignment and “ML1819 Research Assignment 2” for the second assignment).
  2. Team ID
  3. Task ID and Title
  4. All Student Names and IDs
  5. Explanation how each student contributed to the work
  6. Word Count
  7. Exceptional circumstances (if you feel there were some exceptional circumstances preventing you from completing the project as it should have been done, please outline them here)
  8. URL to source code repository, e.g. <https://github.com/BeelGroup/1-Billion-Citation-Dataset>
  9. URL to source code repository activity e.g. <https://github.com/BeelGroup/1-Billion-Citation-Dataset/graphs/contributors>
  10. Screenshot showing the commit activity of all students, e.g.



### Content [85%]

#### Abstract [0%, Optional]

If you like, add a short abstract to your report. The abstract will not be marked but it may be sensible to have if you plan to publish your report e.g. on arXiv or as a blog post.

#### Individual Title and Authors [0%][[3]](#footnote-3)

Add your own title (it should differ from the title in our project description) and your names (including affiliation and email) to the first page of the report (see illustration below).

#### Introduction [10%]

Introduce a) the background of your work, and clearly state b) the research problem and c) your research question/goal.

#### Related Work [10%]

Explain and discuss critically what other researchers did to answer the research question and what their results were. Do not spend too much time in finding related work but focus on describing properly the work you found. We will not mark you down when some important related work is missing (as long as you describe some related work).

#### Methodology [30%]

Briefly describe the dataset used, including the format of the data, and how the data was processed. Explain how and why you (pre-)processed the data to make it suitable for your analysis. Describe the machine learning algorithms selected and how you went about selecting appropriate values for the algorithm parameters. Present plots justifying your choices and discuss your decisions. Given the limited time you have, we do not expect a perfectly tuned system. Rather it is the critical discussion here that is important, and this should cover the major issues affecting your choices plus the level of uncertainty that your analysis indicates for the parameter choices. Also, explain and justify how you evaluate your work (e.g. chosen metrics, how training and test data was split …).

#### Results & Discussion [30%]

Present, explain and discuss the results that you obtained. Include tables and figures where appropriate. Finally, answer clearly the research question. The answer must be based only on your own experiments and results. Discuss your results also under consideration of the related work. For instance, are your results confirming the results of related work or contradicting it?

#### Limitations & Outlook [5%]

Discuss the limitations of your work, and what steps you would undertake next if you were to continue the project.

#### Acknowledgements [0%, Optional]

If you feel your assignment is particularly good and has some interesting results, you are welcome to publish it as a Blog post or pre-print or submit it even to a real conference. In that case, we would appreciate if you add the following acknowledgement or footnote and a citation.

***Acknowledgement***

*This analysis was conducted as part of the 2018/19 Machine Learning module CS7CS4/CS4404 at Trinity College Dublin [1].*

***Bibliography***

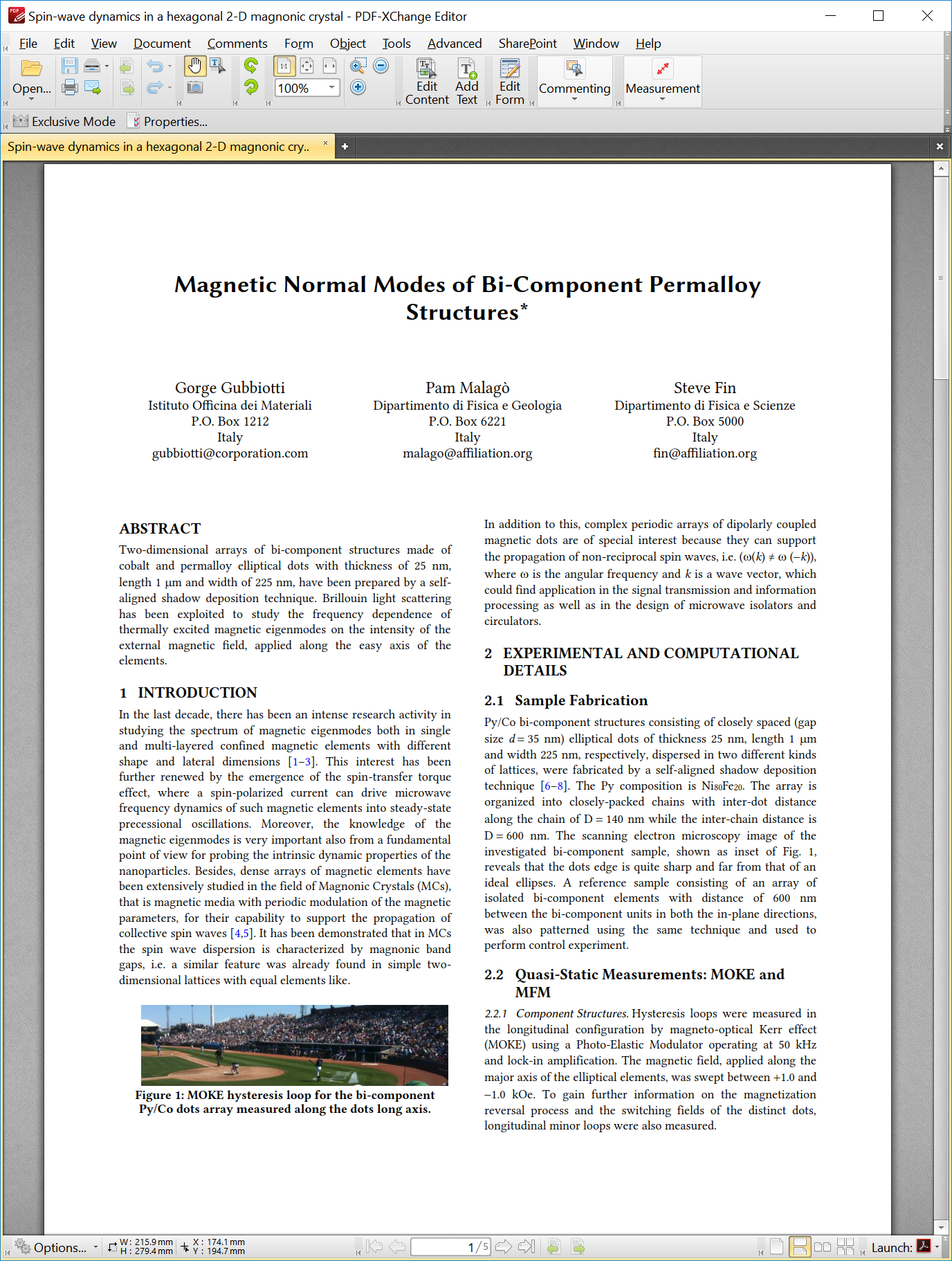
*[1] Joeran Beel and Douglas Leith. Machine Learning (CS7CS4/CS4404). Trinity College Dublin, School of Computer Science and Statistics. 2018.*

#### References [0%]

Around 5-10 including at least 2 or 3 journal or research articles. Other references may include blog posts etc. No Wikipedia.

### Design / Layout / Writing [15%]

For your report, use the ACM SIG Proceedings template.[[4]](#footnote-4) You will not be marked down for minor differences (e.g. if references to figures are not blue) but overall your report should look like the illustration below. You may also use the Word template that we provide, which has some slight modifications. The cover sheet may be in whatever layout you like.



Also, follow our writing guidelines <https://www.scss.tcd.ie/joeran.beel/students_corner/academic-writing/>.

As a general advice: Spend some time to read through your assignment sentence by sentence and remove irrelevant words.

# Submission

* Store your report as PDF
* Name the PDF file “ML1819--task-$taskID--team-$teamID.pdf”, e.g. *ML1819--task-201--team-05.pdf*.
* Send a gDrive/Dropbox/Onedrive/Web/… link to the PDF via email to Joeran.beel@scss.tcd.ie. Do not attach the PDF directly!
* The subject of the email should be
  + “[ML1819] Research Assignment 1 task-$taskID--team-$teamID” for the first assignment
  + “[ML1819] Research Assignment 2 task-$taskID--team-$teamID” for the second assignment

# Questions?

Please use the forum.

1. If you do not submit the source code and data, or submit it incorrectly or without demonstrating continuous activity, up to 100% of your marks will be deducted. [↑](#footnote-ref-1)
2. If the required information is not or inadequately provided, up to 2 marks will be deducted for each missing/incorrect item on the cover page. If the word count is manipulated, up to 10 marks will be deducted. [↑](#footnote-ref-2)
3. If not provided, marks will be deducted. [↑](#footnote-ref-3)
4. <https://www.acm.org/publications/proceedings-template> and <https://1drv.ms/f/s!AtfAgPR4VDcEqI4RPjF1MM8wxoe8cA> [↑](#footnote-ref-4)