DAT390

Data science seminar

2   Reviewing the state of the art

2.5   Literature review and first presentations

Simula master thesis fair

simula

Sign up for the Simula master thesis fair to learn more about your opportunities with **Simula Research Laboratory**.

Date: Monday, 14th October 2024
Time: 16:30 – 18:00
Location: Downtown Oslo (Kristian Augusts gate 23)

About the event: Food and refreshments will be served and you will get the opportunity to talk with our researchers to learn more about your opportunities in **Simula**. Check out Simula's pre-defined master's projects here. You are also welcome to bring project ideas of your own.

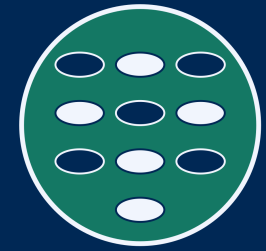**Relevant disciplines:** Informatics, data science, mathematics, and physics.

Registration deadline: 8th October 2024. Sign up here.
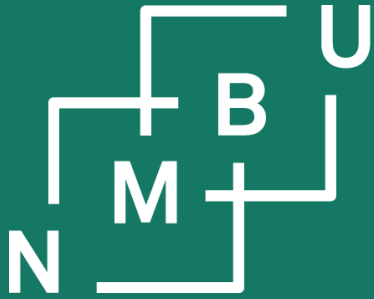
2 Literature review

2.5 Writing up the review
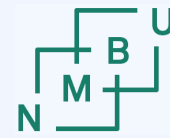
Overview over week 38 submissions

Statistics over the main advisors for master thesis of students from DAT390.
(In some cases, it was unclear who is the main advisor. Also, not all answered.)

Within data science:

Fadi Al Machot	2
Martin T. Horsch	1
Jonas Kusch	1
Hans E. Plesser	1 – 3
Alexander Stasik	0
Oliver Tomic	0 – 2
Kristin Tøndel	1 – 2
Habib Ullah	2 – 3
Eirik Valseth	1
Guang Yang	0

Outside data science:

From REALTEK

Cecilia Futsæther (fysikk)	0 – 2
Kristian Liland (maskin)	2
Abbas Roozbahani (bygg)	1
Leonardo Rydin Gorjão (fysikk)	1 – 3
M. Salman Siddiqui (maskin)	2

From other faculties

Daumantas Bloznelis (HH)	1
Sahameh Shafiee (BIOVIT)	1

How not to write the review …

(▮▮▮▮▮ et al., 2018) is relevant, because it addresses the crucial issue of defending ▮▮▮▮ against ▮▮▮▮▮▮▮▮▮, which is vital for ensuring the reliability of AI models used in ▮▮▮▮▮▮▮. The approach leverages model robustness without the need for retraining or modifications, making it practical for immediate application. This aligns with the ▮▮▮▮ requirement for AI models in the context of ▮▮▮▮▮▮.

(▮▮▮▮▮ et al., 2018) is important, because it introduces a methodology to evaluate the limitations of ▮▮▮▮ methods. It emphasizes the pitfalls of solely relying on ▮▮▮▮ ▮▮▮▮▮ for explanations, a critical consideration in ▮▮▮▮▮ where precise insights are vital. Additionally, it highlights the inadequacy of certain methods for tasks sensitive to data or model interactions, aligning with the challenges in this domain.

(▮▮▮▮▮ et al., 2021) is relevant, because it assesses AI system ▮▮▮▮▮▮▮ ▮ detection, highlighting concerns about reliance on ▮▮▮▮▮▮. This issue extends to broader ▮▮▮▮▮▮ contexts, emphasizing the need for ▮▮▮▮ The article underscores the necessity of ▮▮▮▮▮ for trustworthy ▮▮▮▮ models, aligning with the importance of ▮▮▮▮ in ▮▮▮ research data analysis.

(▮▮▮▮, 2023) is crucial for my research because it improves AI models, which are essential for reliable ▮▮▮▮▮ analysis. Its impressive performance in ▮▮▮▮▮ ▮▮▮▮ aligns perfectly with my focus on ▮▮▮▮▮ for ▮▮▮▮▮. Furthermore,

2 Referansevurdering

[1]Jeg valgte denne kilden på grunn av dens relevans for ███████████████ ███████████████ Resultatene indikerer betydningen av ███████████████ ███████████

[2]Denne kilden presenterer forskning som benytter klassifikasjon og regresjonstrær for ██████████████ ved hjelp av et datasett fra en studie om ██████ ████████████████████ Studien antyder at ████████████████████ █████████████████████████████

[3]Denne kilden utforsker hvordan ulike modeller kan ████████████ █████████████████████, samtidig som de skiller mellom ulike ██████ █████████████████ anses som en vesentlig faktor for å forstå ████████████████

[4]Denne kilden gjennomgår studier fra perioden 2016 til 2020 som benytter seg av maskinlæring for å identifisere nøkkelvariabler i forbindelse ████████ ████████████ Videre utforsker studien ulike typer ████████████ og antyder at ████████████ kan være mer tilpasset for ███████████, mens andre er bedre egnet for ████████████
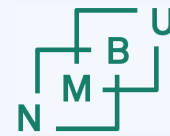
… how to write it instead

We asked ourselves this question at last year's seminar, and came up with this:

What questions can we ask ourselves when reviewing the state of the art?

- Why should the reader believe my story?
- What new does this reference bring to the table?
- What impact does the previous work have on overcoming challenges from my work?
- Is my method actually novel, or was it already there in the literature?

These are all much better questions than "why is this reference included?"
If you find yourself asking that, look at the above for a better guidance.

… how to write it instead: (1) Focus on the fact

Last week we saw the following examples:

"In general, a slow moving neutron has a much higher probability of interacting with a nucleus than a fast moving neutron [27]."

"Previous studies of metal probes, a pea root and an expanding tube in soil suggested that soil density might decrease exponentially from a root's surface (Dexter & Tanner, 1972; Greacen *et al.*, 1968)."

This is the most common way of approaching the literature in scientific writing. The author wants to say something that is part of the research background.

Therefore, they state the fact, and cite the reference as support for that fact.

… how to write it instead: (2) Coverage and gaps

In *Introduction* sections, it is common to mention very briefly what has been done and what the gaps are (so readers know how your own work is novel).

Examples from Al Machot *et al.*, doi:10.3390/jimaging8060171, **2022**:

"Therefore, different learning paradigms with limited labeled data have been presented in the literature, namely semi-supervised learning [4], life-long learning [6], and active learning [7]."

"However, zero-shot learning is still a challenging research field since we need to predict unseen test categories that are never used when training the models [21–23]. For example, most ZSL methods like Deep Embedding Model (DEM) [24 –26] discover direct embeddings from global features to the semantic space. However, the methods cannot capture the appearance relationships between different local regions in this way."

Criteria for a good literature review

As mentioned by Kristin last week:

1) Are you covering the key aspects of the state of the art?

2) Do you analyse the references from your own point of view?

3) Are the references to high-quality material, and are they cited correctly?

Expect feedback on:
- Clarity of the topic, domain, and problem
- Attribution of developments or findings to the right sources
- Coverage of the domain
- Writing style
- Overall, including formal aspects (*e.g.*, number of references cited)

Peer reviewing of the reports: How and why

You will receive detailed feedback from Kristin or me, depending on whose section you are in. In addition, groups for peer review have been formed.

- Four eyes see more than two: You receive another independent assessment.
- Looking at others' writing closely can give you ideas for your own.

Look up your partners on Canvas and send them your material upon submission.

Suggestion – provide feedback on:
- Clarity of the topic, domain, and problem
- Attribution of developments or findings to the right sources
- Coverage of the domain
- Writing style
- Anything else that would come to your mind

Proposed structure for the literature review

You are in principle completely free here with respect to the structure.
It is understood that this is a working document, the first step to your report.

Still some suggestions:

- An **abstract** is not needed. The abstract of your report or master thesis would more strongly focus on your work, rathern than the state of the art.

- Write the **Introduction** section already now. This is something that can be done to 90% based on reviewing the literature. Then you already have it.

- Your **Theory and Background** section (the main element at this stage) ideally consists of two or three subsections, to guide the reader.

- Give the reader a takeaway message in a **Discussion and Conclusion**. This will be very different from the Discussion and the Conclusion in the report. But it is probable that you will still be able to reuse this material.

Presentation plan and scheduling

We will be scheduling 15 minutes per presentation:

- About one minute on average to get set up
- **Ten minutes** for the **presentation**
- **Two minutes** for **feedback** from two pre-selected observers
- **Two minutes** for **discussions** and acknowledgment of the feedback

The presentations in weeks 45 to 48, with a focus on your own methodology and how you go beyond the state of the art will follow the same scheme.

The two presentations and 2 × 2 feedback statements (≥3 will be accepted) are the only elements in DAT390 that are mandatory, other than the final report. Please communicate any dates where you know you will be unavailable (by 27.9.).
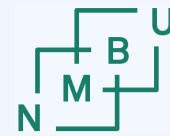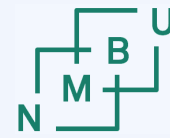
Feedback to presentations

Best learning occurs when there is a good balance between positive and negative feedback. But it is hard to give negative feedback. Here, do state both:

1) What about the presentation was strongest, or most clear and convincing?
2) What about the presentation was weakest, or least clear and convincing?

See it as a gift that you are making to other students (formulate it friendly).

Good method for constructing such feedback: Prepare a list of criteria: What is it that I expect from an ideal presentation on the state of the art (or the own me-thodology)? Follow the presentation and give a score on each item on the list.

Then, you necessarily find a strongest and a weakest point that you can mention.
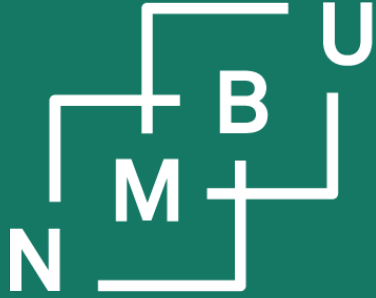
Response to the feedback

Best learning occurs when there is a good balance between positive and negative feedback. But it is hard to give negative feedback. Here, do state both:

1) What about the presentation was strongest, or most clear and convincing?
2) What about the presentation was weakest, or least clear and convincing?

It is useful to receive such feedback. Now, you don't need to respond at all. It can make sense to acknowledge the feedback and let it stand as it is.
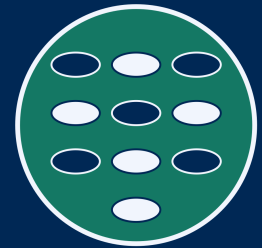
Don't enter into an argument with others about how they were wrong.
You are obviously the person in the room who knows your subject best.

You would easily win the argument, but that would discourage other students from giving important critical feedback.

DAT390
Data science seminar

2   **Reviewing the state of the art**

2.5   **Literature review and first presentations**