

Unit 6 - Evaluation of the Literature Review and Research Proposal

Submissions:

This evaluation critically reflects on my performance across two graded assignments within the module “Research Methods and Professional Practice”, part of my Master’s degree in Data Science. The first assignment involved the creation of a literature review on the use of machine learning to detect health insurance fraud, while the second required the development and delivery of a research proposal presentation on applying machine learning to sentiment analysis. These tasks provided an opportunity to develop core research competencies, including academic writing, time management, critical analysis, and the design and communication of research. Throughout both assignments, I encountered several challenges and areas for development, particularly in terms of structuring my arguments, referencing practices, and aligning content with deliverable constraints. However, I also made substantial progress between the two tasks by incorporating feedback and applying evidence-based strategies, such as enhanced time management (Alyami et al., 2021). Yet, some points for improvement remain such as project planning that considers both temporal and content-related dimensions (Graan and Rommel, 2024). This evaluation draws upon relevant academic literature to reflect on my learning process, identify key strengths and weaknesses, and highlight transferable lessons for future academic and professional work.

In the first graded assignment within this module, I created a literature review on the use of machine learning for predicting health insurance fraud. By reading existing literature on this topic extensively, I was able to classify the used methods and present different approaches within supervised, unsupervised, and ensemble

methods (du Preez, 2025; Bauder and Khoshgoftaar, 2018). While my approach had several strengths, it also had weaknesses. Strengths included the general structure. Classifying the used methods into the three named categories allowed me to give the literature a comprehensive structure, making it easier for the reader to follow the explanations (Bozkurt, Braun and Rossmann, 2022). Additionally, for the parts about supervised and unsupervised learning, I sourced many recent academic references which allowed me to critically evaluate all named approaches. The part about ensemble methods, however, could have been improved. There, the literature was a bit scarce and did not allow as deep an analysis as the two previous parts (Bauder and Khoshgoftaar, 2018; Joudaki et al., 2014). This lack of depth also resulted in mark reductions and negative feedback by the tutor. Here, I should have done a better job stressing the scarcity of literature on the topic, as recent reviews also highlight that ensemble methods, while promising, are less frequently studied in health insurance fraud detection compared to supervised approaches (du Preez, 2025). Additionally, I should have given myself more time to proof-read my work (Elsevier, n. d.). This would have allowed me to spot this discrepancy in the presentation of supervised and unsupervised learning methods on one hand and ensemble methods on the other hand (Joudaki et al., 2014). Furthermore, I did not use referencing in the conclusion, which resulted in an additional mark reduction as well as negative feedback. Here, I thought that referencing the same papers again to sum up findings I already discussed in detail was not necessary, but best practice suggests that referencing in conclusions is important to reinforce the evidence base of the summary statements (Divecha, Tullu and Karande, 2023). Overall, I took several findings from the evaluation of my approach within this assignment that I will include in my approach for future submissions. Firstly, I will ensure to use a high level of referencing throughout the whole paper, including the introduction and conclusion

wherever necessary as sentences are based on the work of others. Secondly, I need to structure my work better and manage my time to get finished slightly before the deadline to allow myself to proof-read my work in its entirety. This will then also help me fix minor errors before submission. Finally, I need to stress research gaps more comprehensively to allow the reader to follow my explanations.

Some of these findings were already included in my second graded submission in this module, a research proposal presentation. I there presented a research proposal for the usage of machine learning for sentiment analysis. Now, learning from mistakes made in the first graded assignment, I planned ahead of time and was able to adequately understand how much time I need for which part of the assignment, an approach supported by research on the positive impact of time management on academic performance (Alyami et al., 2021). I then step by step put together the research design for the proposed study. However, even though I was able to plan ahead, I lost track of the time limit for the presentation over time. The presentation was supposed to be 15 minutes long, with a tolerance of 10%. After I firstly recorded the full presentation, though, it was about 24 minutes long. The research I conducted was a little too detailed and now, only few days before submission, I had to cut out some parts to comply with the time limit. This process was tedious and should be avoided in the next task I work on, i.e., the master's thesis. Thus, a major takeaway from working on the research proposal presentation was not only to focus on how much time working on tasks will take but also on how much place in the deliverable it will take. This aligns with recent findings that effective project planning requires both temporal and content-based allocation (Graan and Rommel, 2024). This will then allow to plan how long each piece may become, informing the depth of research

necessary and helpful. Other than that, however, this second graded assignment worked way better than the first. Even after cutting the content to within the time limitation, I had enough time to proof-read the content and optimise my recordings. I only received two pieces of slightly negative feedback for this assignment. Firstly, in the future I need to improve the audio quality of recorded submissions as this was slightly fluctuating throughout the presentation. This is consistent with research showing that audio quality significantly impacts audience engagement and comprehension in educational presentations (Schiller et al., 2024; Gartner-Schmidt, 2023). Secondly, I should improve the slides used for the presentation. I created one slide per sub-topic of the presentation and this led to a rather static presentation, sometimes spending several minutes on the same slide. Thus, splitting up slides and switching to new slides more frequently could have been beneficial, as recommended by studies on multimedia learning and slide design (Liu, 2024; Bland, Guo and Dousay, 2024). These findings will be included in my approach for future presentations I hold at work, as in my remaining master's courses this will not be required anymore.

Overall, the progression between the first and second graded assignments in this module evidences my development as a reflective and strategic researcher. In the first assignment, strengths such as the structured classification of machine learning methods (Bozkurt, Braun and Rossmann, 2022) and critical engagement with literature on supervised and unsupervised learning were evident, yet the lack of depth regarding ensemble methods and insufficient referencing in the conclusion led to justified critical feedback (du Preez, 2025; Divecha, Tullu and Karande, 2023). These insights directly informed my improved approach in the second assignment.

There, I adopted more rigorous time management practices (Alyami et al., 2021). Remaining areas for improvement, such as slide design and audio quality, were identified and will inform future professional presentations (Liu, 2024; Schiller et al., 2024). Additionally, I will need to plan not just for time requirements but also for space requirements of each section in the future (Graan and Rommel, 2024). These experiences have enhanced my ability to evaluate academic sources critically, plan effectively, and present research findings in a coherent, engaging manner. These skills will be essential for my master's thesis and future professional roles in data science.

List of References:

Alyami, A., Abdulwahed, A., Azhar, A., Binsaddik, A. and Bafaraj, S. M. (2021) 'Impact of Time-Management on the Student's Academic Performance: A Cross-Sectional Study', *Creative Education* 12(3). Available at:

<https://www.scirp.org/journal/paperinformation?paperid=107573>

Bauder, R. and Khoshgoftaar, T. (2018) 'A Survey of Medicare Data Processing and Integration for Fraud Detection', *2018 IEEE International Conference on Information Reuse and Integration*, pp. 9-14. Available at:

<https://dl.acm.org/doi/10.1109/IRI.2018.00010>

Bland, T., Guo, M. and Dousay, T. A. (2024) 'Multimedia design for learner interest and achievement: a visual guide to pharmacology', *BMC Medical Education* 24(1).

Available at: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10845804/>

Bozkurt, Y., Braun, R. and Rossmann, A. (2022) 'The Application of Machine Learning in Literature Reviews: A Framework', *IADIS International Journal on Computer Science and Information Systems* 17(1), pp. 65-80. Available at:

https://www.researchgate.net/publication/363372594_THE_APPLICATION_OF_MACHINE_LEARNING_IN_LITERATURE_REVIEWS_A_FRAMEWORK

Divecha, C. A. Tullu, M. S. and Karande, S. (2023) 'The art of referencing: Well begun is half done!', *Journal of Postgraduate Medicine* 69(1), pp. 1-6. Available at:

<https://pmc.ncbi.nlm.nih.gov/articles/PMC9997612/>

du Preez, A., Bhattacharya, S., Beling, P. and Bowen, E. (2025) 'Fraud detection in healthcare claims using machine learning: A systematic review', *Artificial Intelligence in Medicine* 160. Available at:

<https://www.sciencedirect.com/science/article/pii/S0933365724003038>

Elsevier (n. d.) What is the main purpose of proofreading a paper? Available at:

<https://scientific-publishing.webshop.elsevier.com/manuscript-review/what-main-purpose-proofreading-paper/> (Accessed: 18 April 2025)

Gartner-Schmidt, J. (2023) 'The New Normal – Virtual and Hybrid Presentations: Developing Content, Designing Slides, and Delivery Guidelines', *Ear, Nose & Throat Journal* 101(9). Available at:

<https://journals.sagepub.com/doi/full/10.1177/01455613231158799>

Graan, A. and Rommel, C. (2024) 'Projects and Project Temporalities: Ethnographic Reflections on the Normative Power of the Project Form', *Social Anthropology/Anthropologie Sociale*, pp. 1-19. Available at:

<https://www.berghahnjournals.com/view/journals/saas/32/3/saas320302.xml>

Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., and Arab, M. (2014) 'Using data mining to detect health care fraud and abuse: a

review of literature', *Global Journal of Health Science* 7(1), pp. 194-202. Available at:
<https://pubmed.ncbi.nlm.nih.gov/25560347/>

Liu, D. (2024) 'The effects of segmentation on cognitive load, vocabulary learning and retention, and reading comprehension in a multimedia learning environment', *BMC Psychology* 12(1). Available at:
<https://pmc.ncbi.nlm.nih.gov/articles/PMC10759450/>

Schiller, I. S., Breuer, C., Aspöck, L., Ehret, J., Bönsch, A., Kuhlen, T. W., Fels, J., Schlittmeier, S. J. (2024) 'A lecturer's voice quality and its effect on memory, listening effort, and perception in a VR environment', *Scientific Reports* 14. Available at:
<https://www.nature.com/articles/s41598-024-63097-6>