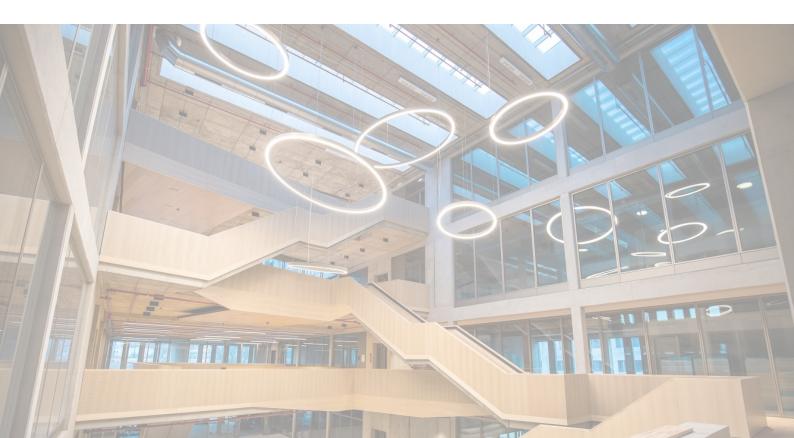# Gisma University of Applied Sciences

Department of Computer and Data Sciences

Assessment Brief

M508 Big Data Analytics

Spring 2025

Assessment Summary

Title:	Individual Final Project
Weighting:	70% Primary Task + 15% Online Assessments + 15% Class Participation
Created on:	April 22, 2025
Deadline:	July 3, 2025 at 18:00 Berlin Time
Submission Method and Length:	This assignment must be submitted as a notebook (converted to a *.html file) in the corresponding submission folder to be found on Canvas. Note that you do not need to sign and attach the Assessed Submission Form. Instead, you must read and accept the Declaration of Authorship provided in the submission folder on Canvas. Please keep the size of your notebook below 20,000 characters.

Assessment Details

Primary Task Topic

You are now an experienced data scientist at the company. For a chosen business problem, you are required to build an end-to-end NLP pipeline in a Jupyter Notebook. In fact, you should formulate your problem as an NLP task and build an NLP system to process textual data. Your designed and implemented pipeline will be submitted to the team lead data scientist of the company.

The notebook should contain the following information:

- A problem statement that elaborates the task. For example, what is the underlying business problem and why is it important? How solving this problem will benefit the company? How would you collect relevant data? How would you formulate this problem as an NLP task?

- A high-level system design that outlines the big picture of your NLP system. For example, what are the main components of your NLP system? How are they connected to each other? Why are they necessary?

- Detailed design, implementation, and analysis of each component of your NLP system, including a modeling phase where a machine learning model is trained for the defined task.

- An evaluation step to evaluate the overall system or each component individually using proper evaluation metrics.

- A final discussion on the overall pipeline. For example, what are the overall strengths and limitations of the proposed solution? What are the implications of the results for the business problem? What are your data-driven recommendations for solving the initial business problem?

Guidelines:	Mind the structure of your submission and its quality of writing. The texts and codes should be written in a clear and easy-to-follow manner.
	All the design decisions should be made in a principled and well-justified manner, either by explaining the intuition or by conducting empirical experiments.
	You can get inspired from any public resources (e.g., blogs, documentation, open-source projects). But the design and implementation of your project should be yours. Your submission should reflect your complete understanding of what you do. Otherwise, it could be a sign of academic misconduct.
	The use of generative AI technologies (such as ChatGPT) in your final assignments is not allowed unless the assessment guidelines explicitly clarify, under which terms, you are allowed to use these technologies. Any violation of this rule will result in an investigation of academic misconduct.
	[Applicable when the assignment is data-driven:] When you need to choose a dataset, choose a new dataset that was not used in the exercises. Mention the URL of your dataset in your submission, so we can find it on the web.
	[Applicable when the submission method requires a GitHub repository link:] When including the URL of a GitHub repository in your report, please ensure that no updates are made to the repository after the submission deadline. Any updates made after the deadline will be considered as continued work on the project and may result in the submission being marked as a failure.
	[Applicable when the assignment title is group work:] When the assignment type is group project, make sure all group members contribute equally and transparently. The size of the group must be 2 unless it is explicitly permitted by the tutor upfront. The group composition cannot be changed after week 7. Every group member must contribute to every task, including both technical and documentation. The contribution of all group members must be visible in both the report and the GitHub repository. In the report, a section must declare the contribution of each group member (who has done what). On the GitHub repository, the accounts of both group members must have contributed.
Purpose:	Designing and implementing such a project is one of your key responsibilities in your career. This assignment is designed to assess your ability in that regard. We are especially interested to see that you can apply various concepts that you have learned in the module in a systematic and principled way.
Links to Learning Outcomes:	The assignment relates to all the intended learning outcomes of the module.

Additional Components:	At Gisma University of Applied Sciences, in-class participation and engagement with asynchronous content contribute 30% of the total module grade.
	Students who actively participate in their scheduled synchronous classes, according to their designated mode of delivery, will earn up to 15% of their final module mark. The awarded percentage is proportional to their participation rate.
	Students who successfully engage with asynchronous materials and complete all required summative assessments will earn an additional 15% of their final module mark. Asynchronous tasks must be completed by the specific deadlines set by the tutors. All assigned tasks must be submitted by the final deadline associated with the principal assessment.
	Students who do not actively participate in synchronous sessions will still be allowed to submit their assessments. However, their final module mark will be reduced by up to 15%. Likewise, failing to engage with asynchronous materials and complete the required short summative assessments will result in a deduction of up to 15% from the final module mark.

Marking/Assessment Criteria for the Primary Task

Mark Weight (100%)	Fail (0 - 49%)	Sufficient (50 – 59%)	Satisfactory (60 – 74%)	Good (75-89%)	Very Good (90-100%)
Marking Criteria	Does not fulfil the requirements of the assessment.	Demonstrates acceptable knowledge and understanding of the subject-matter and achievement of learning outcomes at low to average level of performance.	Demonstrates substantial knowledge and understanding of the subject-matter and achievement of learning outcomes at average to above average performance levels.	Demonstrates a comprehensive knowledge and understanding of the subject-matter and achievement of learning outcomes at well above average levels of performance.	Demonstrates a comprehensive knowledge and understanding of the subject-matter and achievement of learning outcomes at high (highest) levels of performance.

Assessment Criteria:	• The correctness, completeness, and conciseness of runnable codes. (35%)
	• The structure of the report, quality of writing, and critical evaluation of codes and results in the text. (35%)
Notes about Marking	As part of our commitment to academic standards, assignments may be reviewed and marked by markers beyond the module tutor through our independent assessment process. This ensures consistency and fairness in grading.

General Tips

Answer the Question:	It may seem obvious, but make sure you are answering the question you have been set, not the question you would prefer to answer. If the brief has a number of tasks or parts, answer all of them. Parts that involve evaluation or analysis are usually longer and worth more marks than parts that ask for description or explanation. Keep the brief in front of you and check it regularly.
How to Use Assessment Criteria:	The assessment criteria document is not usually a guide to the structure of your assignment. Each section of the criteria is not a separate paragraph in your assignment, but qualities that you need to demonstrate throughout. Treat the assessment criteria as a checklist at the end not as a plan at the beginning. Also, the criteria document often tells you what to demonstrate (e.g., critical analysis) but not necessarily how to do it. For how to do it, look back at the skills and activities you have covered in the rest of the module. Above all, remember this is not a test of how much you know or how much you have read about the topic. It is a test of how well you can use your knowledge to answer the specific question set.
Planning and Preparation:	Make sure you attend the lectures, especially the first and the last one, where we will be 'unpacking' this assignment in greater detail.
Referencing:	Gisma University of Applied Sciences requires that students use Harvard Referencing.
Plagiarism and Cheating:	Your attention is drawn to the University's stated position on plagiarism. THE WORK OF OTHERS THAT IS INCLUDED IN THE ASSIGNMENT MUST BE ATTRIBUTED TO ITS SOURCE (a list of references and bibliography must be submitted). Please note that this is intended to be an individual piece of work. Ensure that you read through your work prior to submission. Action will be taken where a student is suspected of having cheated or engaged in any dishonest practice. Students are referred to the University regulations on plagiarism and other forms of academic misconduct. Students must not copy or collude with one another or present any information that they themselves have not generated. For more information on Plagiarism, please see the relevant section in your Programme Handbook.