Gisma University of Applied Sciences

Assessment Brief

M606 Machine Learning

Spring 2024

Assessment Summary

Title:	Individual Final Project
Weighting:	70% Primary Task + 15% Online Assessments + 15% Class Participation
Created on:	April 16, 2024
Deadline:	July 4, 2024 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 by submitting your assignment, you automatically sign the Assessed Submission Form and take responsibility for your submission. Please keep the size of your notebook below 20,000 characters.

Assessment Details

Primary Task Topic

You are the newly appointed data scientist of the company. For a chosen dataset, you are required to build an end-to-end machine learning pipeline in a Jupyter Notebook. 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 a machine learning task?
- A data exploration discussion on the characteristics of the given dataset. For example, does the dataset suffer from any data quality issues? Does the dataset require any sampling or balancing techniques? Which evaluation metrics fit this dataset?
- Data preprocessing and feature engineering steps to prepare the dataset for learning.
- A model training step to select the best machine learning algorithm and tune its hyperparameters.
- A model assessment step to evaluate the final performance of the best trained model on an unseen test set.
- 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? What are the most informative features of your model? Is your model explainable?

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. 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. 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.
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:	GISMA University rewards in-class participation, and engagement with asynchronous content, at a rate of 30% per module. Students participating $\geq$ 80% (factoring on possible extenuating circumstances) of their synchronous classes as per their due mode of delivery will gain 15% towards their final module mark. Students successfully engaging with asynchronous material on the gamification/microlearning path and completing all summative assessments in the asynchronous environment will equally gain 15% towards their final module mark. Designated asynchronous tasks should be completed by the deadlines specified by the tutors. Do note that all tasks must be completed by the deadline applicable for the principal assessment task. The above also entails that students falling below 80% of participation, although they will still be allowed to submit, will have their final mark capped at 85/100. Equally, if they fail to engage with the asynchronous material and complete the short summative assessments included in specific checkpoints during each term (usually 4), their module mark, irrespective of their engagement and participation in synchronous delivery, will drop by a maximum rate of 15%.

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%)

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.