![Gisma University of Applied Sciences logo]

Gisma University of Applied Sciences

Department of Computer and Data Sciences

Assessment Brief

M516 Business Project in Big Data & AI

Autumn 2025

Assessment Summary

Title:	Individual Project
Weighting:	70% Primary Task + 15% In-Class Progress Presentations + 15% Class Participation
Created on:	September 30, 2025
Deadline:	December 18, 2025 at 18:00 Berlin Time
Submission Method and Length:	This assignment must be submitted as a report (converted to a *.pdf file) in the corresponding submission folder to be found on Canvas. The report must contain a GitHub repository link, where you keep the well-documented implementation of the project. 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 report below 3000 words.

Assessment Details

Primary Task Topic

You are now an experienced data scientist at the company. For a chosen business problem, you are required to conduct comprehensive research and write a report that will be suitable for publication. To this end, you need to choose or collect a dataset and apply data science and engineering techniques. Your report will be submitted to the team lead of the company.

Your report should include a clear introduction that sets out the objectives and purpose of your research, a detailed description of methodology and analysis of your findings, and a conclusion that summarizes your main points and draws conclusions from your research. Take screenshots of code blocks and place them in the report when necessary. The report should also contain the URL link of a GitHub repository, where you keep the implementation of the project. The repository should be well designed, structured, and documented.

Note

All project files, including code and relevant documentation, should be uploaded to an online project repository (e.g., GitHub, GitLab). All files in the repository must be committed before the deadline. Any commits made after the deadline will be considered late submissions and will not be marked.

The project must also include:

- A short video (3 to 5 minutes max) where students record their screen, explaining the project and demonstrating the application from start to finish.

- The video should be uploaded to a preferred platform (e.g., YouTube, OneDrive).

- A link to the video must be included in the project documentation along with the project repository link.

Note on Project Report

The first page of the project report must include:

- **Student Information:** Full name, student ID, etc.
- **Project URLs:** Links to the project's GitHub repository and the recorded video demonstration.

The project report should provide a comprehensive overview of the entire development process and include the following sections:

- **Introduction:** Overview of the project, including objectives, problem domain, and significance of the chosen solution.
- **System Architecture:** Description of the overall architecture, including key components (e.g., backend, APIs, data principles).
- **Implementation:** Clear and detailed description of the approach, including a flow diagram to illustrate the process.
- **Challenges and Solutions:** Outline of technical or design challenges faced during development and the solutions implemented.
- **Results:** Demonstration of how the application functions, including screenshots or outputs.
- **Conclusion and Future Work:** A summary of the project outcomes and recommendations for future improvements or additional features.

You can define your topic of interest or alternatively choose one of the following topics, which will be discussed in the kick-off session.

Predicting Technological Shifts in Germany's Manufacturing: The Convergence of AI, IoT, and Green Technologies

Germany, known for its robust manufacturing sector, is at the cusp of a technological revolution. This project aims to use AI-driven data analytics to predict future shifts and trends in German manufacturing, focusing on the adoption and impact of technologies, such as Artificial Intelligence (AI), the Internet of Things (IoT), Green and Sustainable Technologies, and Advanced Robotics. By analyzing historical manufacturing data, global technology adoption rates, and Germany-specific economic indicators, the project will forecast how these technologies will reshape production lines, operational efficiencies, and overall manufacturing output. Additionally, the research will delve into the socio-economic implications, including potential job shifts, required workforce re-skilling, and the environmental impact of adopting green manufacturing processes.

Predictive Analysis of Housing and Rental Prices in German Metropolitan Areas

The housing market in cities like Berlin, Munich, and Frankfurt has seen significant fluctuations over the past few years. This project aims to use AI to analyze historical housing and rental price data, along with socio-economic indicators, to predict future trends. Factors to consider include urban development plans, population growth, transportation infrastructure changes, and broader economic indicators. The model will segment predictions based on neighborhoods or districts, offering a more granular view of the market dynamics.

Forecasting Rare Earth Element Prices: Integrating Geopolitical, Technological, and Economic Factors with AI

Rare earth elements, crucial for various high-tech applications and green technologies, have a volatile price trajectory influenced by geopolitical tensions, technological advancements, and global economic factors. This project aims to use AI-driven models to predict future prices of rare earth elements. By integrating data sources, such as mining outputs, global trade data, technological adoption rates (e.g., electric vehicles and wind turbines), and geopolitical events, the study will forecast both short-term and long-term price trends. Additionally, the research will explore scenarios like potential new mining discoveries or significant technological shifts to understand their impact on rare earth element prices.

Predicting Societal Sentiments Towards Generative AI and its Socio-Economic Implications in the Near Future

Generative AI, which includes technologies capable of producing content like text, images, and videos, is poised to have profound implications on various facets of socio-economic life. This project aims to use a combination of Natural Language Processing (NLP) and sentiment analysis techniques to gauge public sentiment toward generative AI from various data sources like news articles, social media, forums, and more. By tracking the evolution of sentiment over time and correlating it with major technological milestones and incidents, the study will forecast societal acceptance or resistance to generative AI. Additionally, the research will explore the potential socio-economic impacts, such as shifts in job markets (e.g., content creators), the emergence of new industries, ethical considerations, and potential regulations or policies.

Detecting and Mitigating Fake News

This project focuses on detecting and mitigating the proliferation of fake news. It will explore: (1) modeling fake news based on users' judgments to detect them with higher precision, (2) developing a model for teaching users how to perceive the truthfulness of news, (3) investigating the vulnerability of different groups of people to various kinds of news, and (4) studying whether humans trust AI more or less than other humans to detect fake news.

Human-AI Cooperation in Dynamic Networks

AI increasingly influences human life, and how humans trust and cooperate with AI is a significant area of interest. This project will explore how humans and AI cooperate in dynamic networks. Inspired by Adam Grant's claim that individuals who cooperate tend to be more successful, the study raises interesting questions: (1) Do humans trust AI? (2) Will individuals who can better cooperate with machines gain higher long-term benefits than those who cannot? The project will investigate these aspects, with implications for both work environments and broader societal contexts.

AI for Sustainability

AI has contributed to many applications in daily life, but it also holds the potential to support sustainable development goals. This project aims to design and implement an AI-based system that addresses sustainability challenges, such as environmental pollution detection, smart agriculture, and energy consumption reduction, with the overall goal of CO_2 reduction. An interesting example use case can be that the system analyzes customer behaviors to help them understand their carbon footprint by tracking activities like shopping, dining out, and food ordering. By analyzing purchase history, transportation choices, and energy usage, the AI can

calculate the carbon emissions associated with these activities and suggest sustainable alternatives, encouraging eco-friendly decisions that reduce overall carbon impact.

Fair Autonomous Vehicles

Autonomous vehicles (AVs), which rely on AI systems rather than human drivers, are revolutionizing transportation but raising ethical and fairness concerns in decision-making. Different AI systems and policies may lead AVs to behave differently in identical situations, particularly in scenarios that involve split-second decisions about the safety of various road users (e.g., cyclists, pedestrians, passengers, and other vehicles). These differences could result in disproportionate harm to specific groups, depending on how the AI prioritizes outcomes based on factors, such as speed, distance, or road conditions. This project will explore how to design AI systems for autonomous vehicles that ensure fairness across different populations and road users.

Social Media-Based Trend Prediction

Social media posts frequently capture real-world news, opinions, and emerging trends. This project aims to design and implement a system that continuously collects and analyzes social media data to predict future trends with greater accuracy. By leveraging advanced AI and machine learning techniques, the system will identify patterns and signals in posts to forecast a wide range of phenomena, from epidemic outbreaks and traffic congestion to fluctuations in stock prices and energy consumption. This proactive approach will enable timely responses and better decision-making across various sectors by providing insights into evolving trends and potential future developments.

Data Cleaning for Machine Learning

Data cleaning is a fundamental yet time-consuming step in the data science pipeline, traditionally performed as a standalone task to enhance data quality. This project seeks to innovate by developing an advanced data cleaning system that integrates directly with machine learning workflows. The system will dynamically select and apply the most effective cleaning techniques based on their impact on model performance, using performance-driven metrics to guide decision-making. By creating adaptive data cleaning pipelines, optimizing for scalability and efficiency, and providing customization options for users, the project aims to streamline data cleaning.

Explainable AI: Enhancing Transparency and Trust in Machine Learning Models

Explainable AI aims to address the black-box nature of many machine learning models, especially deep learning models. This project will focus on developing methods to make AI systems more transparent and interpretable by human users. By integrating techniques, such as feature attribution, model distillation, and counterfactual explanations, the project will explore how to provide clear, understandable reasoning for model predictions. The research will investigate how explanations affect user trust and decision-making across various industries, such as healthcare, finance, and autonomous systems.

Table Representation Learning for Enhanced Data Understanding

Tabular data is the most widely used data format across industries, yet it presents unique challenges for machine learning models. This project aims to develop novel approaches for learning effective table representations, which can be used to improve performance on downstream tasks,

such as classification, regression, and table completion. By exploring various techniques, such as relational embeddings, context-aware feature extraction, and transfer learning for tables, the project seeks to create a robust framework for machine learning models to better understand and operate on tabular data. The ultimate goal is to enhance the ability of models to generalize across diverse table formats and domains, enabling more accurate predictions and insights from tabular data.

Understanding Data Semantics and its Impact on Machine Learning Model Training

This project focuses on investigating how the semantics of data – the inherent meaning and context within the data – affect the performance and training of machine learning models. By studying how different data representations, labeling strategies, and feature encodings capture the underlying semantics of data, the project will explore their impact on model learning and generalization. The focus will be on how semantic mismatches between training and real-world data can lead to model performance degradation, and on proposing solutions to mitigate these issues.

AI for Real-Time Anomaly Detection in Industrial IoT Systems

With the increasing deployment of IoT devices in industrial environments, detecting anomalies in real-time is crucial for ensuring smooth and efficient operations. This project will leverage AI-driven methods, such as unsupervised learning and time-series analysis, to develop a robust anomaly detection system for IoT networks. By integrating various data sources, such as sensor readings, network traffic, and system logs, the system will be able to identify irregular patterns or potential faults. The project will explore how this system can be adapted across various industries, from manufacturing to energy management, to ensure reliability and minimize downtime.

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 progress presentations together contribute 30% of the total module grade for project-based modules.
	Students who actively participate in their scheduled synchronous classes, according to their designated mode of delivery, can earn up to 15% of their final module mark. The awarded percentage is proportional to their participation rate.
	Students who successfully deliver at least 3 in-class progress presentations during the term will earn an additional 15% of their final module mark. The awarded percentage is based on both the quantity and quality of the presentations. All progress presentations must take place during scheduled lectures and be completed no later than the last lecture.
	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%. Similarly, failing to deliver at least 3 progress presentations during class sessions 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 along with the video demonstration. (40%)
	• The structure of the report, quality of writing, and critical evaluation of codes and results in the text. (30%)

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.