

Data Science Applications

Semester 2 2025 Assignment Feedback





Question 1

Common features of strong answers

1a

- Clearly defined cleaning steps at the top of the code and the reason they were undertaken, with reference to the business context.
- Clear checks for each step beyond simply sample output of the dataset.
- Correct application of both embeddings and TF-IDF.
- Insightful interpretation of vectorisation outputs.
- Thoughtful consideration of the importance of vectorisation checks, such as discussing cosine similarity distribution or using pairwise comparisons to examine highly similar and highly different reports.

1b

- Prioritisation of the most important input features to use.
- Clear justification for why dimension reduction was needed and which components to retain.
- Clear justification for the chosen clustering algorithm.
- Thoughtful selection and justification of hyperparameters.
- Thorough examination of clustering outputs using internal validation.
- Internal validation metrics were clearly defined and explained to justify cluster selection.
- Combined business context with internal validation outcomes to balance statistical metrics with business needs and interpretability.

1c

- Correct application of the random forest algorithm to the cluster analysis.
- Accurate calculation of variable importance.
- Clear connection between the keywords and the business context of healthcare advice.
- Insightful analyses to support the feature importance narrative (e.g. frequency or proportion of each keyword in each cluster compared to others, statistical tests to determine if keyword usage differs significantly between clusters).

1d

- Thorough examination of clustering outputs using multiple manual validation techniques.
- Clear and insightful interpretation of validation results.
- Good balance of quantitative and qualitative analysis in manual validation.
- Identification of meaningful patterns or trends across clusters.



- Strong connection between cluster characteristics and the business context of healthcare advice.
- Synthesising information across multiple clusters to form a coherent narrative.
- Provided intuitive labels for each cluster with strong links to business context.
- Digging into the manual text descriptions (simple spot checks of text were important, rather than just viewing one-way summaries).
- Highlighted limitations of cluster analysis.

1e

- Clear and concise summary of key findings from the cluster analysis.
- Striking a good balance between depth of analysis required and the 500-word limit.
- Logical flow of ideas, connecting clustering results to business implications.
- Language and explanations tailored for a non-technical management audience.
- Appropriate level of detail included for an executive team.
- Suggested realistic and actionable insights directly addressing the information content of the commentary.
- Highlighted risks and limitations of clustering.

Common features of weak answers

1a

- Incorrect application of vectorisation methods.
- Lack of interpretation of vectorisation outputs.
- No connection to the business context.
- No checks or checks were trivial, such as checking the size of the vectorisation output without any meaningful analysis.

1b

- Use of too many input features (e.g. text vectors with hundreds of dimensions).
- Made feature selections without justifying the variation of the data that was explained.
- Did not use subsets of both embedding and TF-IDF features.
- Lack of justification for algorithm choice/misunderstanding purpose of different clustering algorithms.
- Arbitrary (or no/default) selection of hyperparameters.
- Superficial or incorrect interpretation of internal validation results.
- No connection to the business context of analysing healthcare advice or only trivial references.



1c

- Misunderstanding the purpose of discriminator modelling in this context.
- Errors in applying the random forest algorithm (e.g. using bag-of-words features rather than TF-IDF).
- Incorrect calculation of variable importance, or confusing variable importance with other metrics.
- Listing keywords without relating them to clusters or the business context.
- No attempt to validate or interpret the results.
- Determined top 10 keywords per cluster rather than a single set of top 10 keywords distinguishing all clusters.
- Descriptions did not match with code outputs.

1d

- Superficial examination of clustering outputs.
- Brief or no description of cluster characteristics.
- Confusing internal and manual validation.
- Lack of or incorrect interpretation of validation results.
- No connection to the business context of healthcare.
- Describing individual comments rather than cluster-level characteristics.
- Overinterpreting minor differences between clusters.
- Failing to recognise the limitations of the clustering approach.

1e

- Overly technical description of clustering results without clear business relevance.
- Misaligning cluster characteristics with unrelated management issues.
- Use of jargon or complex statistical terms without explanation.
- Technical description of the clustering methodology rather than business insights.
- Unrealistic or unfeasible actions or recommendations.
- Did not use intuitive labels from Question 1d.
- Did not outline value of pathology commentary.
- Unclear what the findings mean for business ("so what" missing).

Question 2

Common features of strong answers



2a

- Clear explanation of cleaning steps.
- Justifying data cleaning decisions in the context of the business problem.
- Appropriate data splitting (with strong justification of choices) to avoid leakage.
- Identifying appropriate cleaning steps for different types of data.

2b

- Clear definition of the entity and timestamp basis.
- Strong justification for the chosen unit of analysis, linked to the acute diagnosis task.
- Consideration of both regular and event-driven timestamp options.
- Balancing granularity of analysis with practical business considerations.
- Consideration of how the unit of analysis might affect model performance and interpretability.
- Reflection on potential limitations or trade-offs of the chosen approach.

2c

- Clear and justified definition of acute diagnosis in the coming 12 months.
- Discussion of how the constructed variable aligns with business objectives.
- Appropriate handling of potential class imbalance in the response variable.
- Appropriate checks on the constructed variable (e.g. distribution, class balance).
- Insightful interpretation of checks, linking back to the business problem and modelling to be performed.
- Considered different aspects of the business context (e.g. commercial, patient views, medical professionals' views).

2d

- Suggesting four diverse and relevant evaluation metrics.
- Clearly justifying each metric in the context of acute diagnosis prediction.
- Explaining how different metrics capture different aspects of model performance.
- Strong linkage between chosen metrics and Betahelp's business objectives (e.g. early detection and efficient resource use).
- Explaining how the metrics complement each other.
- Consideration of the ability of stakeholders to understand and interpret the metrics.
- Explained why some commonly used metrics were not selected.

2e

- Thoughtful feature selection and engineering.



- Interprets the impact of different features on model predictions.
- Comprehensive experimentation with different architectures and hyperparameters.
- Systematic approach to model improvement with justification for each iteration.
- Appropriate use of regularisation techniques to prevent overfitting.
- Clear and correct interpretation of model performance.
- Well-structured, readable code with clear comments.

2f

- Accurate calculation of suggested evaluation metrics.
- Clear comparison to suitable benchmark models with justification for why benchmarks were selected.
- Translated technical metrics into business-relevant insights using plain, business-friendly language (e.g. '1 in 3 alerts is accurate', '9 out of 10 acute cases identified').
- Linked metrics directly to Betahelf's operational and strategic goals (proactive outreach, workload balance, clinician trust, patient safety, resource allocation).
- Explained the implications of false positives versus false negatives in the healthcare context.
- Management communication was clearly delineated and focused on implications of results rather than numerical descriptions.
- Professional tone and concise structure that read like an executive briefing, not a technical report.
- Good presentation and visualisation of results.

2g

- Accurate calculation and insightful interpretation of feature importance, partial dependence, and SHAP values with clear visualisation.
- Clearly explained key risk drivers (age, past diagnoses, lifestyle factors) in clinically intuitive language.
- Strong connection between model behaviour and business implications for diagnosis prediction and patient care.
- Proper assessment of model fairness using appropriate metrics, demonstrating awareness of demographic differences in model performance and how that should be monitored.
- Management communication section was clear and written using non-technical language with focus on implications rather than just describing results.
- Discussion of unexpected model behaviours and how insights can inform business strategies.

Common features of weak answers



2a

- Limited or inappropriate cleaning steps.
- Failure to split the data or inappropriate splitting or justification for the splitting decision.
- Split data but then applied cleaning in a way that effectively made the split redundant (e.g. causing data leakage).
- Uniform approach to all variables without consideration of their nature.
- No connection to the business context of healthcare.
- Did not identify and handle data privacy issues, especially date of birth.
- Spent excessive time attempting to clean issues that were unlikely to have much impact on the model.
- Code, especially that which appeared to be produced by AI, was not as concise as it could have been so was harder to follow.

2b

- Vague or unsuitable definition of the entity or timestamps.
- Proposing a unit of analysis that doesn't align with the available data.
- Lack of justification for the chosen unit of analysis.
- Consideration of only one type of timestamp without explanation.
- Failure to recognise potential issues with the proposed approach.
- Mainly described how or what they were proposing without justifying why it was suitable for the business context of predicting acute diagnoses.

2c

- Lack of justification for the chosen response variable.
- No consideration of time frame or its impact.
- Absence of checks on the constructed variable.
- Failure to interpret checks or link them to the business context.

2d

- Listing the four most common or simplest metrics without regard to the context.
- Lack of diversity in the suggested metrics (e.g. all threshold-dependent).
- Lack of description of metrics proposed.
- No justification for the choice of metrics.
- Overly technical or mathematical explanations without clear business relevance.
- Failure to consider the business context of acute diagnosis prediction.
- Misunderstanding what the metrics measure.



- Misunderstanding the implications of class imbalance on metric choice.
- Unclear on how to balance sensitivity (safety) versus precision (efficiency) in a healthcare context.

2e

- Limited or no feature engineering (e.g. created no features other than summarisation of visits, prescription, or anomaly counts and one-hot encoding of high cardinality features).
- No consideration of overfitting or underfitting (e.g. insufficient epochs or no early stopping).
- Lack of experimentation or justification for model choices.
- Failure to interpret model performance or incorrect interpretation of model metrics.
- Poorly structured, difficult-to-read code.
- Code that appeared to be AI-generated with blind copy and paste without explanation of how it related to the context of the problem.
- Interpreted this as a purely technical question with little commentary justifying decisions.

2f

- Inaccurate calculation of metrics or misunderstanding what the metrics measure.
- Lack of comparison to benchmark models or insufficient effort in designing suitable benchmarks.
- Failed to comment on overfitting issues or investigate why models performed as they did.
- Overly technical language without translating metrics to business implications (e.g. listed metric values without explaining what they mean for patient safety or resource allocation).
- Confused statistical significance with business significance.
- No clearly delineated management communication section.
- Too numeric and descriptive, lacking actionable implications or clear recommendations.
- Struggled to explain trade-offs and threshold selection in practical, non-technical language.
- No clear conclusion or recommendation for Betahelf management.

2g

- Inaccurate calculation or superficial interpretation of model behaviour metrics.
- Misinterpreting partial dependence plots, confusing correlation with causation, or overlooking feature interactions.
- Overly technical language (e.g. listing raw variable names or SHAP values) with no high-level summary for decision-making.
- No clear management communication section.



- Did not use proper fairness metrics (e.g. only examined score distributions or feature usage rather than model performance across different groups) or did not attempt fairness assessment at all.
- Misinterpreted fairness findings by describing bias but not explaining its impact or mitigation strategies.
- Failure to connect model behaviour to the business problem of acute diagnosis prediction.

Question 3

Common features of strong answers

3a

- Used all five components of a strong prompt effectively (context, input data, questions or commands, output requirements, constraints or limitations).
- Clear instructions for the LLM to produce output in text format.
- Appropriate use of examples or few-shot learning in the prompt (multi-shot examples).
- Handled potential variability in lab result content, especially long lab results which may exceed the maximum allowed prompt length.
- Successful application to five randomly selected lab results.
- Ensured consistent formatting across different types of responses.
- Output that accurately provided pathology commentary and advice.
- Used markdown formatting to help structure the prompt.

3b

- Clear and concise summary of key findings from the critique of the LLM outputs.
- Actionable insights directly addressing the suitability of AI-generated commentary and advice.
- Language and explanations tailored for a non-technical management audience.
- Logical flow of ideas, connecting LLM results to business implications.
- Strong linkage of each strength and weakness to specific healthcare stakeholders.

Common features of weak answers

3a

- Poorly constructed prompt that missed one or more required components.
- No examples or context provided in the prompt.
- Incorrect application of the prompt or selection of random reports.
- Output that did not match the specified requirements.



- Prompts that were not automatically generated and reusable (e.g. contained only one patient's details).
- Used only the human pathologist's report without additional prompt structure.
- Did not query the LLM programmatically (e.g. just pasted results rather than showing live code).
- Obtained empty results from LLM and did not try enough different samples to obtain valid outputs.

3b

Common features of weak answers

- Overly technical description of LLM outputs without clear business relevance.
- Failure to address the suitability of AI-generated pathology commentary.
- Use of AI or NLP jargon without explanation.
- Lack of actionable insights or recommendations based on the LLM outputs.
- Generic comments that could apply to any AI project.
- Mentioned stakeholder needs but did not link them to specific strengths and weaknesses of the output.

Question 4

Common features of strong answers

- Clear structure with logical flow, effective transitions, and well-managed time within the five-minute limit.
- Demonstrates strong understanding of key findings from all three questions with in-depth consideration.
- Clear linkage between data science techniques and answering business questions (e.g. clustering of reports resulted in topics A, B, C which were not found in structured data, demonstrating that humans add value).
- Communicates clearly and concisely using language suitable for management team, avoiding technical jargon.
- Prioritises most crucial insights with actionable recommendations.
- Strong connection of findings to Betahelf's business challenges.
- A conclusion that ties together main insights and offers forward-looking recommendations.

Common features of weak answers



- Incomplete coverage of findings or overemphasising one part of the analysis at the expense of others.
- Lack of structure, coherent flow, or clear transitions between sections.
- Overly technical language or unexplained jargon unsuitable for management audience.
- Failure to relate findings to Betahelf's specific business issues with lack of actionable insights.
- Appeared to be reading word for word from a script.
- Included extra sections that were not necessary so could have used time more effectively.



Sample assignment graded as 'Significantly above pass level'

A sample assignment is provided as an example of one that was graded as 'Significantly above pass level'. Students should use this example, along with the assignment rubric, to help them self-assess their own assignment attempts.

This assignment was marked as Significantly above pass level. Markers made the following comments about individual questions in the assignment.

Assignment question	Marker comments
1	<p>The student demonstrated strong technical skills in applying text vectorisation and clustering techniques, with more than five cleaning steps, multiple vectorisation checks, and five manual validation analyses with intuitive cluster labels. The student provided sufficient justification for choices and considered business context throughout, analysed keyword relationships through proportions and partial dependence plots, and communicated findings using appropriate language with actionable insights.</p>
2	<p>The student demonstrated strong technical and business acumen throughout the classification modelling task. Although not required, they conducted EDA to guide cleaning, appropriately removed personally identifiable information, split data to avoid leakage, and provided detailed justification for selecting patient as the entity and regular monthly intervals as timestamps. The student performed exhaustive, thoughtful and intuitive feature engineering across all tables with well-structured code that was easy to follow, took an iterative approach to fine-tuning the neural network with detailed explanations, and selected evaluation metrics with strong business linkage.</p> <p>The student communicated model performance and trade-offs clearly in business-relevant, executive-level language with appropriate diagnostics and comparison to two reasonable benchmarks and identified and visualised key model features. However, the response would have benefited from deeper analysis of why the neural network did not outperform simpler models, and better assessment of model fairness.</p>



Assignment question	Marker comments
3	The student's LLM prompt included context with an explicitly set role and input data formatted appropriately for the LLM to understand. Further specification of constraints could have helped to ensure consistent outputs across different models or applications. The critique demonstrated all necessary components with strengths and weaknesses broken down by different stakeholders and insights that were well written.
4	The student delivered a very clear presentation that considered all three questions well, with good structure and effective transitions between topics, using language suitable for the management audience.