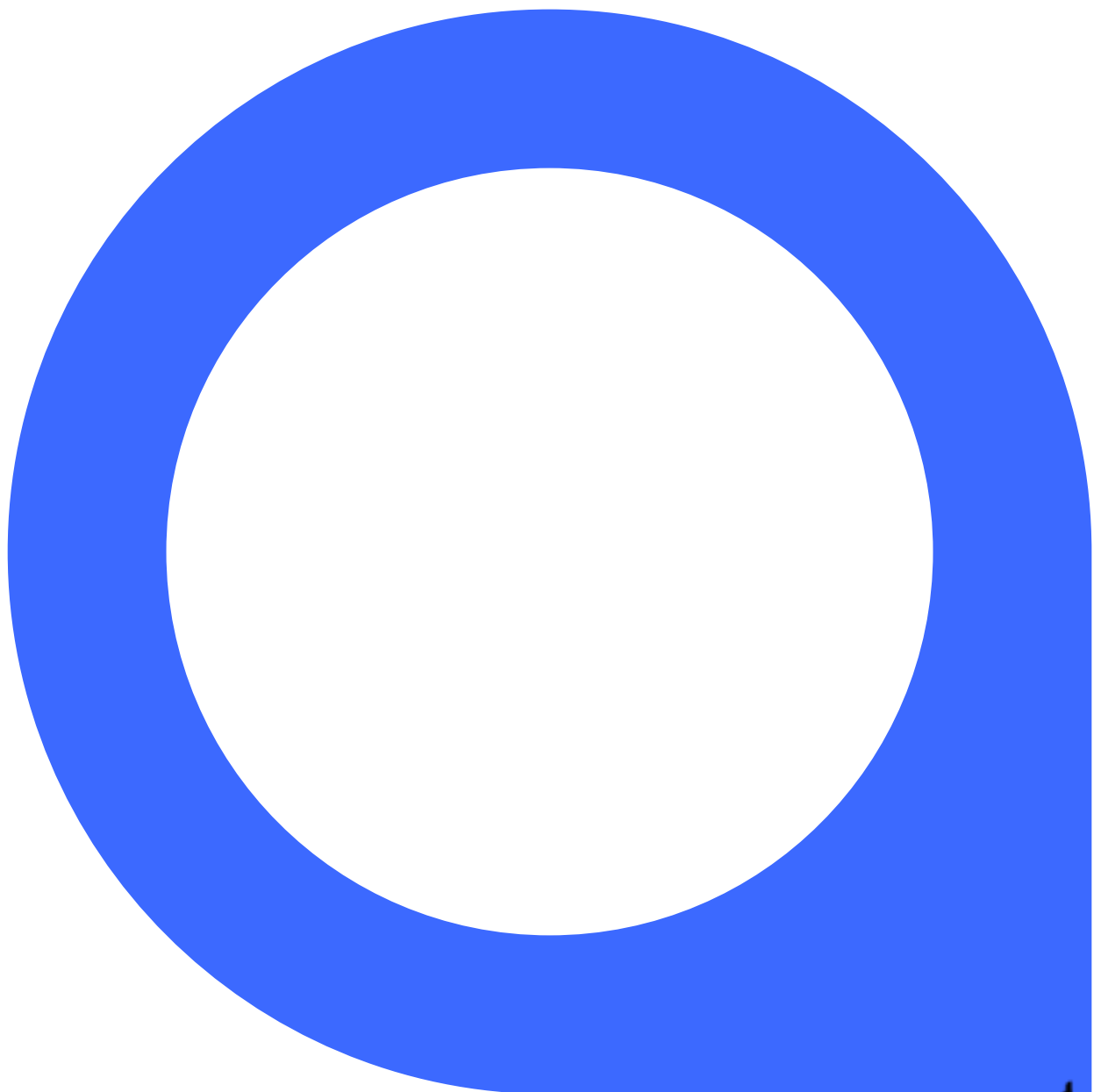


Data Science Applications

Assignment Semester 1 2024 – Marking Rubric





Marking Rubric

The following rubric outlines the specific standards required for each question in the assignment.

Criteria in bold at the top of each rubric box provide a holistic view on what differentiates answers under each rubric grade. Other criteria in each rubric box provide more specific guidance about components of an answer that are usually required to meet each holistic criterion.

Criteria designated as 'hurdle' in the left most column are those that must be met to gain a certain grade (usually either a grade of 2 or 3).

Criteria designated as 'differentiator' in the left most column are those that help to distinguish between assignment responses at each rubric grade.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
1a	5%	Demonstrates a very strong understanding of the dataset.	Demonstrates a good understanding of the dataset.	Demonstrates some understanding of the dataset.	Demonstrates a limited understanding of the dataset.	Does not demonstrate an understanding of the dataset.
Hurdle (for a 4)		Splits the data into training, validation, and test sets at an appropriate stage in the data exploration to avoid leakage in the model to be built in Question 3.		Does not split the data into training, validation, and test sets, or does not split the data at an appropriate stage in the data exploration to avoid leakage in the model to be built in Question 3.		
Differentiator		Examines the dataset by applying appropriate data exploration techniques and visualisations that are tailored to each variable.	Examines the dataset by applying appropriate data exploration techniques and visualisations.	Summarises the dataset by applying appropriate data exploration techniques.	Summarises the dataset but some of the data exploration techniques are not appropriate.	Summarises the dataset but most or all the data exploration techniques applied are not appropriate.
Differentiator		Describes key characteristics of the dataset, with a strong link to the problem context.	Describes key characteristics of the dataset, with some link to the problem context.	Outlines key characteristics of the dataset, with some link to the problem context.	Outlines characteristics of the dataset, but these are not key characteristics or there is no link to the problem context.	Does not outline the characteristics of the dataset.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
1b	5%	Demonstrates a very strong ability to clean data.	Demonstrates a good ability to clean data.	Demonstrates some ability to clean data.	Demonstrates a limited ability to clean data.	Does not demonstrate an ability to clean data.
Hurdle (for a 3)		Accurately applies at least five appropriate cleaning steps.			Applies some cleaning steps, although it is not clear whether these steps are appropriate for the problem context and/or there are errors in the steps.	
Differentiator		Explains why each cleaning step has been undertaken for the purpose of preparing the data for the analysis that will follow.	Explains why most cleaning steps have been undertaken for the purpose of preparing the data for the analysis that will follow.	Outlines each cleaning step undertaken with some reference to preparing the data for the analysis that will follow.	Outlines each cleaning step undertaken but does not link this outline to the analysis that will follow.	Does not outline each cleaning step that has been undertaken.
Differentiator		Correctly interprets the outcomes of appropriate check(s) for each cleaning step in a thorough and insightful way.	Correctly interprets the outcomes of appropriate check(s) for each cleaning step.	Correctly interprets the outcomes of appropriate check(s) for most cleaning steps.	Applies at least one appropriate check for most cleaning steps but these are usually not interpreted or not interpreted correctly.	Demonstrates only one or two cursory checks have been performed.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
1c	5%	Demonstrates a very strong ability to apply text vectorisation.	Demonstrates a good ability to apply text vectorisation.	Demonstrates some ability to apply text vectorisation.	Demonstrates a limited ability to apply text vectorisation.	Does not demonstrate an ability to apply text vectorisation.
Hurdle (for a 2)		Accurately vectorises the 'Movie Quote' feature in the training, validation, and test datasets.				Vectorises the 'text' feature with some errors in the application of the vectorisation method.
Differentiator		Justifies the chosen vectorisation method in relation to preparing the data for the analysis that will follow.	Describes the vectorisation method used with strong reference to preparing the data for the analysis that will follow.	Outlines the vectorisation method used with some reference to preparing the data for the analysis that will follow.	Outlines the vectorisation method used but does not link this outline to the analysis that will follow.	Does not outline the vectorisation method used.
Differentiator		Correctly interprets the outcomes of at least two types of appropriate checks on the vectorisation completed in a thorough and insightful way.	Correctly interprets the outcomes of one appropriate check on the vectorisation completed in a thorough and insightful way.	Correctly interprets the outcomes of one appropriate check on the vectorisation completed.	Applies at least one appropriate check on the vectorisation completed but does not interpret its outcome or does not interpret its outcome correctly.	Applies at least one check on the vectorisation completed but the check is not appropriate, and the outcome of the check is not interpreted.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
1d	5%	Demonstrates a very strong ability to construct a response variable that is suitable for the given problem.	Demonstrates a good ability to construct a response variable that is suitable for the given problem.	Demonstrates some ability to construct a response variable that is suitable for the given problem.	Demonstrates a limited ability to construct a response variable that is suitable for the given problem.	Does not demonstrate an ability to construct a response variable that is suitable for the given problem.
Hurdle (for a 3)		Constructs a suitable response variable to use in the model.			Constructs a response variable but it is not suitable for use in the model.	
Differentiator		Justifies the response variable constructed with strong reference to the problem context and previous data exploration conducted.	Justifies the response variable constructed with reference to the problem context and previous data exploration conducted.	Describes the response variable constructed with reference to the problem context and previous data exploration conducted.	Describes the response variable constructed but does not reference the problem context or previous data exploration conducted.	Does not describe the response variable constructed.
Differentiator		Correctly interprets the outcomes of at least two types of appropriate checks on the response variable constructed in a thorough and insightful way.	Correctly interprets the outcomes of one appropriate check on the response variable constructed in a thorough and insightful way.	Correctly interprets the outcomes of one appropriate check on the response variable constructed.	Applies at least one appropriate check on the response variable constructed but does not interpret its outcome or does not interpret its outcome correctly.	Applies at least one check on the response variable constructed but the check is not appropriate, and the outcome of the check is not interpreted.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
2a	5%	Demonstrates a very strong ability to apply a clustering algorithm.	Demonstrates a good ability to apply a clustering algorithm.	Demonstrates some ability to apply a clustering algorithm.	Demonstrates a limited ability to apply a clustering algorithm.	Does not demonstrate an ability to apply a clustering algorithm.
Hurdle (for a 2)		Accurately applies a clustering algorithm to the dataset.				Applies a clustering algorithm to the dataset with errors in the way the algorithm has been applied.
Hurdle (for a 3)		Evaluates the clustering outputs using internal validation, including an interpretation of the evaluation output.			Does not evaluate the clustering outputs or does not interpret the evaluation output.	
Differentiator		Justifies the type of clustering algorithm used and the selection of hyperparameters such as the distance or linkage method and number of clusters.	Discusses the type of clustering algorithm used and the selection of hyperparameters such as the distance or linkage method and/or number of clusters, and the selections seem reasonable.	States the type of clustering algorithm used and the selection of hyperparameters such as the distance or linkage method and/or number of clusters, and the selections seem reasonable.	States the type of clustering algorithm used and/or the selection of hyperparameters such as the distance or linkage method and/or number of clusters, but the selections do not seem reasonable.	Does not select hyperparameters such as the distance or linkage method and number of clusters.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
2b	15%	Demonstrates a very strong ability to evaluate a clustering algorithm.	Demonstrates a good ability to evaluate a clustering algorithm.	Demonstrates some ability to evaluate a clustering algorithm.	Demonstrates a limited ability to evaluate a clustering algorithm.	Does not demonstrate an ability to evaluate a clustering algorithm.
Hurdle (for a 2)		Evaluates the clustering output using manual validation, including a correct interpretation of the output.				Does not evaluate the clustering output, does not interpret the evaluation output, or does not interpret the evaluation output correctly.
Differentiator		Describes the key characteristics of each cluster with a strong link to the problem context.	Describes the key characteristics of each cluster with some link to the problem context.	Outlines the key characteristics of each cluster with some link to the problem context.	Outlines some characteristics of the clusters but these are not key characteristics or there is no link to the problem context.	Outlines some characteristics of movie quotes but these are not based on the output of the clustering algorithm.
Differentiator		Communicates the description in a way that is highly suitable for executives of the movie production company.	Communicates the description in a way that is suitable for executives of the movie production company.	Communicates the outline in a way that is mostly suitable for executives of the movie production company.	Communicates the outline in a way that is mostly not suitable for executives of the movie production company.	Communicates the outline in a way that is not suitable for executives of the movie production company.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
2c	5%	Demonstrates a very strong ability to apply clustering outcomes to a business problem.	Demonstrates a good ability to apply clustering outcomes to a business problem.	Demonstrates some ability to apply clustering outcomes to a business problem.	Demonstrates a limited ability to apply clustering outcomes to a business problem.	Does not demonstrate an ability to apply clustering outcomes to a business problem.
Differentiator		Suggests at least three specific and reasonable ways the company can use the clustering outcomes from 2a and 2b.	Describes at least three specific and reasonable ways the company can use the clustering outcomes from 2a and 2b.	Outlines at least three specific and reasonable ways the company can use the clustering outcomes from 2a and 2b.	Outlines at least three specific and reasonable ways the company could use clustering outcomes, but these are hypothetical and not related to the actual findings from 2a and 2b.	Outlines one or two ways the company could use clustering outcomes, but these are either not specific and reasonable or are hypothetical and not related to the actual findings from 2a and 2b.
Differentiator		Communicates in a way that is highly suitable for the movie production company.	Communicates in a way that is suitable for the movie production company.	Communicates in a way that is mostly suitable for the movie production company.	Communicates in a way that is mostly not suitable for the movie production company.	Communicates in a way that is not suitable for the movie production company.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3a	5%	Demonstrates a very strong understanding of how to address class imbalance.	Demonstrates a good understanding of how to address class imbalance.	Demonstrates some understanding of how to address class imbalance.	Demonstrates a limited understanding of how to address class imbalance.	Does not demonstrate an understanding of how to address class imbalance.
Differentiator		Suggests at least two specific and reasonable strategies that will be undertaken to address class imbalance in the response variable.	Describes at least two specific and reasonable strategies that will be undertaken to address class imbalance in the response variable.	Describes one specific and reasonable strategy that will be undertaken to address class imbalance in the response variable.	Outlines one or two reasonable strategies that will be undertaken to address class imbalance in the response variable, but these strategies are not specific.	Outlines one or two strategies that will be undertaken but these strategies will not address class imbalance in the response variable.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3b	20%	Demonstrates a very strong ability to construct a classifier.	Demonstrates a good ability to construct a classifier.	Demonstrates some ability to construct a classifier.	Demonstrates a limited ability to construct a classifier.	Does not demonstrate an ability to construct a classifier.
Hurdle (for a 2)		Correctly constructs a neural network to classify quotes as 'problematic' or 'not problematic'.				Constructs a neural network but does not classify quotes as 'problematic' or 'not problematic'.
Hurdle (for a 2)		Experiments with different model architectures, features, and hyperparameters.				Does not experiment with different model architectures, features, or hyperparameters.
Hurdle (for a 3)		Applies one or more appropriate strategies to address class imbalance in the response variable.			Does not apply appropriate strategies to address class imbalance in the response variable.	
Hurdle (for a 3)		Correctly interprets appropriate check(s) on the final classifier's outcomes.			Applies at least one check on the final classifier's outcomes but the check is not appropriate or is not interpreted.	
Differentiator		Improves predictions between the initial and final iterations, using a range of appropriate metrics.	Attempts to improve predictions between the initial and final iterations, using a range of appropriate metrics.	Takes an iterative approach, but more appropriate metrics could have been used.	Does not take an iterative approach but takes steps to prevent the classifier from over or under fitting to the training data.	Does not take steps to prevent the classifier from over or under fitting to the training data.
Differentiator		The code is structured and very easy to read.	The code is structured and easy to read.	The code is structured and mostly easy to read.	The code lacks structure and is mostly difficult to read.	The code lacks structure and is difficult to read.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3c	5%	Demonstrates a very strong ability to calculate relevant measures of success for a classifier.	Demonstrates a good ability to calculate relevant measures of success for a classifier.	Demonstrates some ability to calculate relevant measures of success for a classifier.	Demonstrates a limited ability to calculate relevant measures of success for a classifier.	Does not demonstrate an ability to calculate relevant measures of success for a classifier.
Hurdle (for a 2)		Calculates how good the final classification model's predictions are using appropriate measures of success.				Calculates how good the final classification model's predictions are but the measure(s) of success chosen are not appropriate for this problem.
Hurdle (for a 4)		Compares the final classification model's outcomes to those under a suitable (simple) benchmark model.		Does not compare the final classification model's outcomes to those under a suitable (simple) benchmark model.		
Differentiator		Justifies the measures used given the movie producer's objective.	Discusses the measures used in relation to the movie producer's objective.	Outlines the measures used in relation to the movie producer's objective.	Outlines the measures used but does not reference the movie producer's objective.	Does not outline the measures used.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
4a	5%	Demonstrates a very strong ability to evaluate a classifier and explain the findings to a non-technical audience.	Demonstrates a good ability to evaluate a classifier and explain the findings to a non-technical audience.	Demonstrates some ability to evaluate a classifier and explain the findings to a non-technical audience.	Demonstrates a limited ability to evaluate a classifier and explain the findings to a non-technical audience.	Does not demonstrate an ability to evaluate a classifier and explain the findings to a non-technical audience.
Hurdle (for a 3)		Correctly assesses the success of the model in meeting the movie producer's objective.			Does not correctly assess the success of the model in meeting the movie producer's objective.	
Differentiator		Summarises the evaluation findings with a strong and insightful link to the movie producer's objective.	Summarises the evaluation findings with a clear and direct link to the movie producer's objective.	Summarises the evaluation findings with some link to the movie producer's objective.	Summarises the evaluation findings but the summary lacks a link to the movie producer's objective.	Outlines some evaluation findings.
Differentiator		Communicates clearly and concisely in a way that is highly suitable for executives of the company.	Communicates clearly and concisely in a way that is suitable for executives of the company.	Communicates clearly and concisely in a way that is mostly suitable for executives of the company.	Communicates in a way that is mostly not suitable for executives of the company.	Communicates in a way that is not suitable for executives of the company.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
4b	10%	Demonstrates a very strong understanding of the limitations of the modelling and actions to overcome them.	Demonstrates a good understanding of the limitations of the modelling and actions to overcome them.	Demonstrates some understanding of the limitations of the modelling and actions to overcome them.	Demonstrates a limited understanding of the limitations of the modelling and actions to overcome them.	Does not demonstrate an understanding of the limitations of the modelling and actions to overcome them.
Differentiator		Discusses at least four key limitations of the modelling that are directly relevant to the problem context, and actions to overcome the limitations.	Discusses three key limitations of the modelling that are directly relevant to the problem context, and actions to overcome the limitations.	Outlines at least three key limitations of the modelling that are directly relevant to the problem context, and actions to overcome the limitations.	Outlines at least three material limitations of the modelling but these are not directly relevant to the problem context. Outlines actions to overcome the limitations.	Outlines at least three limitations of the modelling but these are not material, are not relevant to the problem context, or actions to overcome the limitations are not provided.
Differentiator		Communicates clearly and concisely in a way that is highly suitable for executives of the company.	Communicates clearly and concisely in a way that is suitable for executives of the company.	Communicates clearly and concisely in a way that is mostly suitable for executives of the company.	Communicates in a way that is mostly not suitable for executives of the company.	Communicates in a way that is not suitable for executives of the company.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
4c	10%	Demonstrates a very strong ability to communicate technical concepts to a non-technical audience.	Demonstrates a good ability to communicate technical concepts to a non-technical audience.	Demonstrates some ability to communicate technical concepts to a non-technical audience.	Demonstrates a limited ability to communicate technical concepts to a non-technical audience.	Does not demonstrate an ability to communicate technical concepts to a non-technical audience.
Hurdle (for a 1)		Prepares an explanation of how the classifier works in a video format.				
Hurdle (for a 3)		Demonstrates an understanding of how their classifier works.			Does not demonstrate an understanding of how their classifier works.	
Differentiator		The video has a clear start, middle, and end. There are clear transitions between all sections.	The video has a clear start, middle, and end. There are clear transitions between most sections.	The video has some structure. There are transitions between some sections.	The video has some structure. There are no transitions between sections.	The video lacks structure.
Differentiator		Communicates clearly and concisely in a way that is highly suitable for script writers.	Communicates clearly and concisely in a way that is suitable for script writers.	Communicates clearly and concisely in a way that is mostly suitable for script writers.	Communicates in a way that is mostly not suitable for script writers.	Communicates in a way that is not suitable for script writers.
Total	100%					

Students are advised that a mark of zero (0) will be allocated for any question (or sub-question) where there is no attempt made or the marker finds the attempt is completely unsatisfactory.