Actuaries Institute Australia

DATA ANALYTICS APPLICATIONS

ASSIGNMENT MARKING RUBRIC SEMESTER 2 2021

Education Program
ACTUARIES INSTITUTE

Education that Counts

MARKING RUBRIC

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

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
1 Domain knowledge	10 marks	Explains at least five key characteristics of the television industry and explains how these are relevant to this analysis. Outlines the source(s) of the information acquired. Clearly demonstrates a strong ability to acquire domain knowledge about an industry.	Explains three or four key characteristics of the television industry and explains how these are relevant to this analysis. Outlines the source(s) of the information acquired. Demonstrates an ability to acquire domain knowledge about an industry.	Identifies at least three key characteristics of the television industry and states how these are relevant to this analysis. Lists the source(s) of the information acquired. Demonstrates an ability to acquire domain knowledge about an industry.	Identifies at least three key characteristics of the television industry but does not state how these are relevant to this analysis. Lists some of the source(s) of the information acquired. Demonstrates a limited ability to acquire domain knowledge about an industry.	Does not identify at least three relevant key characteristics of the television industry. Does not list any of the source(s) of the information acquired. Does not demonstrate an ability to acquire domain knowledge about an industry.
2 Twitter rules	5 marks	Explains how the analysis will comply with at least three of Twitter's security, privacy, and other data use rules that are relevant to this analysis.	Discusses how the analysis will comply with at least three of Twitter's security, privacy, and other data use rules that are relevant to this analysis.	Discusses how the analysis will comply with two of Twitter's security, privacy, and other data use rules that seem relevant to this analysis.	States some of Twitter's security, privacy, and other data use rules that might be relevant to this analysis but does not discuss how the analysis will comply with these rules.	Does not identify any Twitter rules that are relevant to the analysis.

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3a i Extract and tokenise	3 marks	Successfully imports and tokenises the tweet text. The code is well structured and easy to read. Demonstrates that a range of data checks have been performed and the quality of the data is suitable for the analysis.	Successfully imports and tokenises the tweet text. The code is well structured and easy to read. Demonstrates that data checks have been performed and the quality of the data is suitable for the analysis.	Successfully imports and tokenises the tweet text. The code has structure and is mostly easy to read. Demonstrates that data checks have been performed.	Successfully imports and tokenises the tweet text. The code lacks structure or is difficult to read. Does not demonstrate that any data checks have been performed.	Does not successfully import and/or tokenise the tweet text. The code lacks structure and is difficult to read.
3a ii Clean	3 marks	Accurately applies at least four reasonable data cleaning techniques. The code is well structured and easy to read. Explains each of these steps and why they have been undertaken. Demonstrates that a range of checks have been performed on the output from data cleaning and demonstrates that the results are reasonable.	Accurately applies three reasonable data cleaning techniques. The code is well structured and easy to read. Discusses each of these steps. Demonstrates that checks have been performed on the output from data cleaning and states that the results are reasonable.	Accurately applies one or two reasonable data cleaning techniques. The code has structure and is mostly easy to read. Demonstrates that some checks have been performed on the output from data cleaning.	Applies some data cleaning techniques with a few errors in the way these techniques have been applied. The code lacks structure or is difficult to read. Does not demonstrate that any checks have been performed on the output from data cleaning.	Does not apply data cleaning techniques or demonstrates no or limited understanding of this NLP step. The code lacks structure and is difficult to read.

Education that Counts

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3a iii Stem and/or lemmatise	3 marks	Accurately applies stemming and/or lemmatisation. The code is well structured and easy to read. Explains each of these steps and why they have been undertaken. Demonstrates that a range of checks have been performed on the output from each of these steps and demonstrates that the results are reasonable.	Accurately applies stemming and/or lemmatisation. The code is well structured and easy to read. Discusses each of these steps. Demonstrates that checks have been performed on the output from each of these steps and states that the results are reasonable.	Accurately applies stemming and/or lemmatisation. The code has structure and is mostly easy to read. Demonstrates that some checks have been performed on the output from these steps.	Applies stemming and/or lemmatisation with a few errors in the way these steps have been undertaken. The code lacks structure or is difficult to read. Does not demonstrate that any checks have been performed on the output from these steps.	Does not apply stemming or lemmatisation or demonstrates no or limited understanding of this NLP step. The code lacks structure and is difficult to read.
3 a iv Vectorise	3 marks	Accurately calculates vectorised features that represent the tweet_text feature. The code is well structured and easy to read. Explains this step and why it has been undertaken. Demonstrates that a range of checks have been performed on the output of vectorisation and demonstrates that the results are reasonable.	Accurately calculates vectorised features that represent the tweet_text feature. The code is well structured and easy to read. Discusses this step. Demonstrates that checks have been performed on the output of vectorisation and states that the results are reasonable.	Accurately calculates vectorised features that represent the tweet_text feature. The code has structure and is mostly easy to read. Demonstrates that some checks have been performed on the output of vectorisation.	Calculates vectorised features with a few errors in the way this step has been applied. The code lacks structure or is difficult to read. Does not demonstrate that any checks have been performed on the output of vectorisation.	Does not calculate vectorised features that represent the tweet_text feature or demonstrates no or limited understanding of this NLP step. The code lacks structure and is difficult to read.

Education that Counts

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3b i Advantages and disadvantages of PCA	3 marks	Explains at least two advantages and two disadvantages of applying PCA that are relevant to the problem context. Explains the relevance of these points to the problem context.	Explains one advantage and one disadvantage of applying PCA that are relevant to the problem context. Explains the relevance of these points to the problem context.	Discusses one advantage and one disadvantage of applying PCA that are relevant to the problem context.	States one advantage and one disadvantage of applying PCA that are relevant to the problem context.	Does not state one advantage and one disadvantage of applying PCA that are relevant to the problem context.
3b ii Apply PCA	2 marks	Accurately applies principal components analysis to reduce the dimension of the vectorised features. The code is well structured and easy to read. Explains this step and why it has been undertaken. Justifies the reduced dimension size that has been selected. Demonstrates that a range of checks have been performed on the output of PCA and demonstrates that the results are reasonable.	Accurately applies principal components analysis to reduce the dimension of the vectorised features. The code is well structured and easy to read. Discusses this step. Discusses the reduced dimension size that has been selected. Demonstrates that checks have been performed on the output of PCA and states that the results are reasonable.	Accurately applies principal components analysis to reduce the dimension of the vectorised features. The code has structure and is mostly easy to read. States the reduced dimension size that has been selected. Demonstrates that some checks have been performed on the output of PCA.	Applies principal components analysis with a few errors in the way the principal components have been calculated. The code lacks structure or is difficult to read. Does not state the reduced dimension size that has been selected. Does not demonstrate that any checks have been performed on the output of PCA.	Does not apply principal components analysis to reduce the dimension of the vectorised features or demonstrates no or limited understanding of PCA. The code lacks structure and is difficult to read. Does not state the reduced dimension size that has been selected.

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3c k-means clustering	3 marks	Accurately applies k-means clustering to group the tweets into a small number of distinct clusters. The code is well structured and easy to read. Explains this step and why it has been undertaken. Justifies the selection for number of clusters.	Accurately applies k-means clustering to group the tweets into a small number of distinct clusters. The code is well structured and easy to read. Discusses this step. Discusses the selection for number of clusters and the selection seems reasonable.	Accurately applies k-means clustering to group the tweets into a small number of distinct clusters. The code has structure and is mostly easy to read. States the selection for number of clusters.	Applies k-means clustering with a few errors in the way the algorithm has been applied. The code lacks structure or is difficult to read. Does not state how the number of clusters was selected.	Does not apply k-means clustering to group the tweets into a small number of distinct clusters or demonstrates no or limited understanding of k-means clustering. The code lacks structure and is difficult to read.
3d Evaluation of clustering	5 marks	Applies internal and manual validation to evaluate the output from the clustering algorithm. Uses visualisation tools to demonstrate that the clustering output is reasonable. Domain knowledge gathered in Question 1 is incorporated into the evaluation. The evaluation demonstrates a strong understanding of the clustering output and the industry context.	Applies internal and manual validation to evaluate the output from the clustering algorithm. States that the results are reasonable. The evaluation demonstrates a good understanding of the clustering output and/or the industry context.	Applies internal or manual validation to evaluate the output from the clustering algorithm. Some consideration of the problem context was incorporated into the evaluation. The evaluation shows an understanding of the clustering output and/or the industry context.	Does not correctly apply internal or manual validation to evaluate the output from the clustering algorithm. No consideration of the problem context was incorporated into the evaluation. The evaluation does not show an understanding of the clustering output or the industry context.	Does not apply internal or manual validation to evaluate the output from the clustering algorithm. No consideration of the problem context was incorporated into the evaluation. The evaluation does not show an understanding of the clustering output or the industry context.

Education that Counts

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
4a Response variable	10 marks	Calculates an appropriate response variable to use in the model. Justifies the response variable selected. Demonstrates that a range of checks have been performed on the response variable calculated and demonstrates that the results are reasonable.	Calculates an appropriate response variable to use in the model. Discusses the response variable selected. Demonstrates that checks have been performed on the response variable calculated and states that the results are reasonable.	Calculates an appropriate response variable to use in the model. States the response variable selected. Demonstrates that some checks have been performed on the response variable calculated.	Calculates a response variable to use in the model. Does not state what response variable has been used or it is not clear whether the response variable is appropriate. Does not demonstrate that any checks have been performed on the response variable calculated.	Does not select an appropriate response variable to use in the model.

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
4b Neural network	10 marks	Correctly constructs a neural network to predict whether share prices will increase, decrease, or remain unchanged following a tweet. Several types of model architecture and hyperparameter selections are tried, along with different sets of features, before deciding on the best design for the network. Takes an iterative approach to build the neural network where evaluation outcomes are used to improve predictions after each iteration. Justifies the final model architecture and parameter selections. Demonstrates that a range of checks have been performed on the final neural network constructed and demonstrates that the results are reasonable.	Correctly constructs a neural network to predict whether share prices will increase, decrease, or remain unchanged following a tweet Explains the model architecture, hyperparameter, and feature selections. Demonstrates that checks have been performed on the neural network constructed and states that the results are reasonable.	Correctly constructs a neural network to predict whether share prices will increase, decrease, or remain unchanged following a tweet States the model architecture, hyperparameter, and feature selections. Demonstrates that some checks have been performed on the neural network constructed.	Constructs a neural network with some errors in the code that negatively impacts the model's predictions. Does not state the model architecture, hyperparameters, or features used in the model. Does not demonstrate that any checks have been performed on the neural network constructed.	Does not construct a neural network to predict whether share prices will increase, decrease, or remain unchanged following a tweet. .

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
4c Tree-based model	10 marks	Correctly constructs a tree-based model to predict whether share prices will increase, decrease, or remain unchanged following a tweet. Several hyperparameter and feature selections are tried before deciding on the best tree-based model. Takes an iterative approach to build the model where evaluation outcomes are used to improve predictions after each iteration. Justifies the final hyperparameter and feature selections. Demonstrates that a range of checks have been performed on the final model constructed and demonstrates that the results are reasonable.	Correctly constructs a tree-based model to predict whether share prices will increase, decrease, or remain unchanged following a tweet Explains the hyperparameter and feature selections. Demonstrates that checks have been performed on the model constructed and states that the results are reasonable.	Correctly constructs a tree-based model to predict whether share prices will increase, decrease, or remain unchanged following a tweet States the hyperparameter and feature selections. Demonstrates that some checks have been performed on the model constructed.	Constructs a tree-based model with some errors in the code that negatively impacts the model's predictions. Does not state the hyperparameters or features used in the model. Does not demonstrate that any checks have been performed on the model constructed.	Does not construct a tree-based model to predict whether share prices will increase, decrease, or remain unchanged following a tweet.

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
4d Model evaluation	10 marks	Evaluates how good the neural network and tree-based models' predictions are using at least three appropriate measures of success. Justifies the measures used given the problem context. Demonstrates, via comments in the notebook, an excellent understanding of how predictive each model is.	Evaluates how good the neural network and tree-based models' predictions are using two appropriate measures of success. Explains the choice of measures given the problem context. Demonstrates, via comments in the notebook, a good understanding of how predictive each model is.	Evaluates how good the neural network and tree-based models' predictions are using one appropriate measure of success. Demonstrates, via comments in the notebook, some understanding of how predictive each model is.	Calculates how good the neural network and tree-based models' predictions are using one appropriate measure of success. Does not demonstrate an understanding of how predictive each model is.	Does not select an appropriate measure to evaluate how good the neural network and tree-based models' predictions are or does not appear to understand how to interpret any of the calculated measures.

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
5a Key risks	5 marks	Explains at least three key risks that are relevant in this context. Demonstrates a very strong understanding of these risks.	Explains two key risks that are relevant in this context. Demonstrates a good understanding of these risks.	Identifies two risks that seem relevant in this context. Demonstrates some understanding of these risks.	Identifies at least one risk that may be relevant in this context. Demonstrates a weak understanding of this risk.	Does not identify at least one risk that is relevant in this context. Demonstrates no understanding of the risks in this context.
5b Implementation considerations	5 marks	Explains at least three key implementation considerations that are relevant in this context. Demonstrates a very strong understanding of these implementation considerations.	Explains two key implementation considerations that are relevant in this context. Demonstrates a good understanding of these implementation considerations.	Identifies two implementation considerations that seem relevant in this context. Demonstrates some understanding of these implementation considerations.	Identifies at least one implementation consideration that may be relevant in this context. Demonstrates a weak understanding of this implementation consideration.	Does not identify at least one implementation consideration that is relevant in this context. Demonstrates no understanding of the implementation considerations in this context.
6	10 marks	The executive summary is presented as a video summary. The summary has good structure and presents all points clearly and concisely. Recommendations made are supported by the summarised findings. The student clearly has an excellent grasp of the concepts being assessed.	The executive summary is presented as a video summary. The summary has good structure and presents most points clearly and concisely. Recommendations made are supported by the summarised findings. The student has a good grasp of the concepts being assessed.	The executive summary is presented as a video summary. The summary has structure and presents some points clearly and concisely. Recommendations made have some link to the summarised findings. The student has a grasp of the concepts, but the language used could convey the information more clearly in some cases.	The executive summary is not presented as a video summary or the summary lacks structure or does not present points clearly or concisely. Recommendations made are not supported by the summarised findings. It is not clear whether the student really understands the concepts being assessed.	No executive summary is provided or the student does not appear to understand the concepts being assessed.
Total	**100 marks**					

Education that Counts