2023S2 QBUS6850 Group Assignment – Marking Guidelines

Task A (15 marks)

Part	Criteria	Marks
1	Missing values • Attempt to identify missing values and apply treatment. • Provides sound justification of treatment.	
2	Exploration data analysis • Univariate analysis, e.g., skewness, distribution, etc. • Bivariate analysis, e.g., visualisations to show relationships between variables, correlation analysis, discriminative features, etc.	

Task B (15 marks)

Part	Criteria	Marks
1	Data pre-processing and feature engineering • Pre-process "review_text" with necessary steps. • Generate text features using appropriate feature engineering techniques. • Provide justification for the chosen method/setting.	
2	Benchmark model implementation • Build a random forest or gradient boosting model using the extracted text features. • Describe the details of model, e.g., the choice of hyperparameters. • Appropriate hyperparameter tuning if applies.	
3	Benchmark model validation • Validate the model with proper evaluation strategy and evaluation metric with sufficient justification. • Identify the strengths and weaknesses of your model. • Provide sufficient details of your analysis.	

Task C (20 marks)

Part	Criteria	Marks
1	Vanilla RNN model • Build a vanilla RNN model using "review_text". • Full report of model specifications, e.g., activation functions, input/hidden/output layer dimensions, etc. • Detailed explanations of model training with appropriate hyperparameters.	
2	Hyperparameter tuning • Tune at least three hyper-parameters that most affect model performance.	

	• Describe the tuning process and analyse the impact of each hyper-parameter on model performance.	
3	RNN model validation and comparison • Evaluate the performance of the RNN model and provide sufficient details of the evaluation process. • Compare the performance of the RNN model against the model from Task B. • Explain and analyse the results of model comparison.	

Task D (10 marks)

Part	Criteria	Marks
1	Pre-trained word embeddings incorporation • Incorporate pre-trained word embedding into your RNN model from Task C. • Provide sufficient explanations for your approach.	
2	Model validation and comparison • Evaluate the performance of this model and provide sufficient details of the evaluation process. • Compare the performance of this model with the model from Task C. • Explain and analyse the results of model comparison.	

Task E (25 marks)

Part	Criteria	Marks
1	Improvement attempts • Depth of exploration and justification based on the evidence from the data. • Provide sufficiently detailed explanations for each attempt. • Summarise what has worked and what has not.	
2	Model validation • Suitable validation process for your new models. • Appropriate hyperparameter tuning. • Comprehensive report on the new models.	
3	Model Comparison: • Principled comparison with the previous models with technical correction and detailed analysis.	

Task F (5 marks)

Part	Criteria	Marks
1	Final challenge results: • Correct name and format for the final challenge result file.	

		• Successful application of your best model to the challenge dataset provided (e.g., correct number of predictions)	

Presentation (10 marks)

Part	Criteria	Marks
1	Well-structured report with clear presentation of text, figures, tables, and formula (if applicable), free of spelling and grammar errors, etc.	
2	Well-documented code with necessary comments.	

Bonus Mark (maximum of 6 marks)

Criteria	Marks
Only awarded to the top 3 groups in the group competition (6 marks for the first place, 3 marks for the second place, and 1 mark for the third place).	