Note: For all sections below, please ensure you stick to the scope of the assignment and follow the instructions. Extra work outside the scope of the assignment will not result in additional points. This is important so we can accurately assess your skills and understanding related to the topics and learning outcomes we are evaluating. It also ensures equity across all students, as some have more experience in this topic while others are new. Please focus on the specified tasks to demonstrate your understanding and skills in relation to the learning outcomes. Some suggestions for best practice approaches can be found in the Assignment 1 feedback document.

Criteria		Ratings						Total Points
		Full marks	**HD**	**D**	**C**	**P**	**F**	
Part 1: Load data, visualise it and identify erroneous values [30%]	Data is correctly loaded from the specified CSV file and properly formatted. Code is clear, logical and follows best practice.	5	5	4	3	2	0 – 1	**/30 Pts**
	The required number of visualisations is adhered to, and carefully chosen to be clear, appropriate, and effectively communicate key information about the dataset. Visualisation elements are appropriate and understandable (e.g., labels, axes). Code is clear, logical and follows best practice.	10	10	8	6	5	0 – 4	
	Erroneous values in the dataset are appropriately identified/justified. Appropriate methods are used to handle erroneous values, in line with best practice approaches. Code	15	15	12	9	5	0 – 3	

	is clear, logical and follows best practice.							
Part 2: Prepare data and models, and make baseline measurements [20%]	Data is split appropriately to maintain the proportion of classes in each set, and appropriate pre-processing steps are used following best practice approaches (i.e. pre-processing methods are used appropriately for features and implemented in the appropriate part of the code etc.). Code is clear, logical and follows best practice to avoid data leakage.	6	6	5	4	3	0 -2	**/20 Pts**
	Both baseline models are appropriately implemented (i.e. coded). Code is clear, logical and follows best practice.	6	6	5	4	3	0 -2	
	Baseline performance is appropriately evaluated using all proposed metrics and confusion matrices. Presented baseline performance values are logical. Code is clear, logical and follows best practice.	8	8	6	5	4	0 – 3	

Part 3: Model Optimisation [40%]	Performance metric selected is appropriate and thought process/rationale behind this selection is clearly explained. Response is within word limit and well written.	5	5	4	3	2	0 -1	**/40 Pts**
	Hyper-parameter optimisation performed correctly on all three models where needed using appropriate and best practice methods (e.g. ensuring no data leakage). All results displayed, including confusion matrices, are logical. The best model is selected logically, with sound rationale. Code is clear, logical and follows best practice.	25	25	20	15	12	0 - 11	
	Probability as described is correctly calculated, with logical values. Code is clear, logical and follows best practice.	10	10	8	6	5	0 – 4	
Part 4: Decision Boundaries [10%]	T-score is appropriately calculated and correctly interpreted in relation to features. Code is	4	4	3.5	2.5	2	0 – 1	**/10 Pts**

	clear, logical and follows best practice.							
	Decision boundaries appropriately and effectively displayed in relation to features. Code is clear, logical and follows best practice.	4	4	3.5	2.5	2	0 – 1	
	The interpretation and evaluation of the model's performance are logical and clearly relate to the client's criteria. Thought process and rationale behind this response are well described/written. Response is within the word limit.	2	2	1.75	1.5	1	0 – 0.9	**/10 Pts**
							Total Points:	**/100 pts**