1. 預處理技術（3 分）：根據您對資料集的分析，討論是否應考慮異常值檢測、歸一化和插補等預處理技術。描述如何使用交叉驗證來確定適當的技術。然後，基於提供的標記數據，整合此主要目標是開發一個分類器，能夠準確地將數據點分類為未見數據的兩個類別之一。教學團隊將使用第 9 週發布的測試數據來評估分類器的性能，只有教學團隊才能存取真實標籤。 V1.0 3 討論透過概述如何將這些技術應用於當前資料來優化模型的預測效能，自然地融入您的整體專案策略中。

2. 分類技術的應用（5分）：利用預處理的數據，概述應用課堂上學到的四種指定分類技術－決策樹、隨機森林、k近鄰和樸素貝葉斯的過程。描述透過交叉驗證進行必要的模型選擇和超參數調整。將建立和提供超參數搜尋合理範圍的需求納入您的討論中，詳細說明如何確定這些範圍及其與整體策略的相關性。此外，考慮不同模型的分類結果的集合以提高預測準確性。

3. 模型評估（3 分）：詳細說明使用交叉驗證評估模型的方法，特別說明所選評估指標如何適合項目的資料分佈。本節應根據資料集的特徵，對為什麼首選某些指標進行深思熟慮的分析。

4. 專案時間表（1分）：提供第二階段專案實施的詳細時間表。清楚地闡明每個部分如何對整體專案目標做出貢獻，確保時間表展示了精心規劃的方法。它應該反映一個簡潔且可行的計劃，顯示如何按計劃完成實施和測試階段，並在第二階段的截止日期前準備好提交所有可交付成果。

剩餘的 3 分將根據提案的呈現進行評分。該提案應該結構良好，以促進理解，並以簡潔和專業的方式呈現，包括根據提供的模板中的示例格式正確的參考書目。妨礙理解提案的格式、拼字、文法、參考書目、引用或標點符號錯誤將被扣分。

1. 專注於應用：在提案中，除非必要，否則避免詳細解釋每種技術的機製或如何計算每個指標。相反，應強調實際考慮因素，例如決定對資料應用哪些技術的標準。

2.技術結合：預處理技術的選擇應與其對分類結果的影響直接相關。使用交叉驗證來確定預處理和分類技術的最有效組合至關重要。由於協同效應，某些組合可能會比其他組合產生更好的性能。

3. 評估結果：當使用交叉驗證評估不同方法的有效性時，請考慮平均值和標準差來確定哪個結果更好。

4. 整合技術：考慮使用整合的潛在好處，例如透過多數投票組合隨機森林和 k-NN 的輸出。

5. 測試的一致性：確保在測試階段也使用與訓練階段相同的預處理技術。

使用生成式 AI 人工智慧 (AI)（例如 ChatGPT）提供了可以支援學生完成此評估任務的新興工具。學生只能在現有真實 V1.0 4 評估的修訂中適當使用 AI。這可能涉及糾正語法錯誤、改進句子結構、增強表達清晰度或進行其他相關修改。學生必須在每個實例中清楚地引用人工智慧的任何使用才能完成評估任務。在修訂現有真實評估之外，未能提及人工智慧的使用或任何其他使用人工智慧的方式可能會構成《學生行為準則》規定的學生不當行為。具體來說，如果您使用生成式人工智慧工具來幫助修改您的提案，您應該1. 致謝部分：在您的作品中包含一個專門的部分，您在其中承認人工智慧工具的使用，並提及哪些部分內容已被人工智慧修改。 2. 文件（不低於四頁限制）：追蹤 AI 互動以及修訂過程中產生的任何輸出。如果需要，該文件可以作為使用人工智慧的證據。學生可能需要透過面對面面試來展示對獨立於人工智慧工具的書面提交的詳細理解。

提案應遵循 Proposal\_Template.doc 的風格。提交的內容應在四頁以內，包括所有參考文獻和插圖（如果需要）。即使您使用講座投影片中的內容，也應在必要時正確提供參考文獻。可以使用未經同儕審查的網路資源，也應該正確引用。

提交 提案應提交

• PDF 或 Doc (Docx) 格式，不接受其他格式，且

• 透過 Blackboard -> 評估 -> 項目 -> 在截止日期前提交提案中提供的「提案提交」Turnitin 連結。

您可以在截止日期之前多次提交提案。僅標記最後提交的版本。自提交截止時間起每 24 小時將扣除最高分數 10% 的罰款，最多 7 天。 7 天后，您將收到 0 分（詳情請參閱 ECP）。即使您已經透過程式分析了數據，也不要在第一階段提交程式碼。

