| **Criteria** | **Ratings** | **Pts** | **Structure** |
| --- | --- | --- | --- |
| Clear statement of classification problem  **Bijo** | **2 pts**  **Full Marks**  Clear context given as to why this is a classification problem worth solving  **1 pts**  **Poor description**  Basic or little motivation for the classification problem being solved.  **0 pts**  **No Marks**  None or almost no motivation for the classification being solved. | / 2 pts | * **Having an early detection is critical to treat or prevent dementia / Alzheimer’s** * Common problem in the population ( 1 in 3 dies because of dementia, deaths greater than prostate + breast cancer ) * For governments, identifying people at risk (biomarkers) is important to budget for Medicare. * Identify 🡪 Detect 🡪 Deter |
| Graphical description of the data  **Jessica** | **3 pts**  **Full Marks**  Clear and appropriate visualisations to describe the dataset.  **2 pts**  **Basic description**  Basic visualizations used. Possibly lacking aesthetic formatting or detail.  **1 pts**  **CPoor visualization description**  Inappropriate visualisation techniques used.  **0 ptsm**  **No Marks**  None or almost no visualisations used | / 3 pts | * Class imbalance (Alzheimer’s Yes / No) * Correlation matrix (feature to feature, feature to response) * Histograms * Box Plots |
| Describe the data  **Hong** | **3 pts**  **Full Marks**  Clear Dataset description  **2 pts**  **Basic description**  Basic dataset description lacking some detail  **1 pts**  **Poor description**  Very little dataset description.  **0 pts**  **No Marks**  None or almost no description of the dataset | / 3 pts | * No. of observations * No. of features, call out response * No missing data * Table of all features names, data type, description |
| Clear description of potential challenges  **Bijo** | **2 pts**  **Full Marks**  Clear and insightful understanding of potential challenges and plans to overcome them.  **1 pts**  **Basic challenges or no plan to overcome.**  Either basic challenges or valid challenges but little idea or plan on how to solve them.  **0 pts**  **No Marks**  Dataset or its description too basic. No challenges foreseen. | / 2 pts | * Class imbalance * Feature selection because we have 33 features * Synthetic data * Limited no. of observations * No unseen real world data to validate the model? * Expensive to validate the model |
| Performance metrics  **Jessica** | **2 pts**  **Appropriate Performance metrics chosen**  Well motivated and appropriate performance metrics chosen to measure their performance that aligns with the context of their dataset and classification problem  **1 pts**  **Basic or not appropriate metrics**  Not well motivated and or slightly inappropriate performance metrics chosen.  **0 pts**  **No Marks**  Inappropriate metrics chosen for the dataset. | / 2 pts | **Model Performance (confusion matrix)**   * Recall – due to the medical context of our dataset * Accuracy (not as important) * Precision * Specificity * Sensitivity |
| Models chosen  **Bijo** | **2 pts**  **Appropriate Modelling techniques chosen**  Well motivated and appropriate modelling techniques chosen.  **1 pts**  **Crude modelling**  Basic modelling techniques chosen. Potentially not well motivated or not the best choice for the data.  **0 pts**  **No Marks**  Inappropriate modelling. | / 2 pts | * **Lasso** – feature selection given we have 33 features, and we have a binary response variable. Logistic regression with penalty can be the base model for comparison. * **Random Forest** – could help with the imbalance. Easy to understand / read. * **kNN** - |
| Plan or schedule  **Hong** | **2 pts**  **Well planned project**  Well thought out plan to tackle the problem including timeline and task division.  **1 pts**  **Basic plan**  Some kind of plan in place, potentially unclear on the timeline and/or task division.  **0 pts**  **No Marks**  No plan or schedule at all. | / 2 pts | * Gantt-chart |