Write a summary of all your results (textual summary) by answering the following questions in detail while mentioning values of the performance metrics and graphs if feasible.
How does the classification performance compare across the 7 DR/FS methods?
How does the regression performance compare across the 7 DR/FS methods?
What is the effect of increasing or decreasing the total number of your desired features on the classification and regression performance?

LDA:

Using LDA with 5 features, classification performance was really well for the first dataset with all algorithms providing accuracy, precision and recall of 100, and AUC and F-Score of 1. Regression performance of first dataset was also good with R-Squared and Adjusted R-Squared error just under 1, and RMSE of 0.05 and 0.009 using linear regression and voting regression respectively. Research Gate specified that adjusted R-Squared should be greater than 0.19 and with low RMSE for a good fit.

For dataset 2, classification performance was low, with around 81 or less for accuracy, precision between 37 and 72, Recall between 26 and 39. The AUC and F-Score were also not very promising. The regression performance was not any better, with Adjusted R-Squared of 0.12 and 0.09 (less than 0.19 showing a poor fit) using linear regression and voting regression respectively, and RMSE of 0.39 for both.

For Dataset 3, the classification performance was even worse, with accuracy, precision and recall close to zero. Regression performance was worse as compared to other datasets, with Adjusted R-Squared errors close to 0.09 (less than 0.19 showing a poor fit), and RMSE

ranging from 30 to 360 using linear regression and voting regression respectively.

For Dataset 4, classification performance was really well with all algorithms providing accuracy, precision and recall of 100, and AUC and F-Score of 1. Regression performance was also good with R-Squared and Adjusted R-Squared error just under 1 (greater than 0.19, showing a good fit), and RMSE of 0.08 and 0.004 using linear regression and voting regression respectively.

For Dataset 5, classification performance was really well with all algorithms providing accuracy, precision and recall of over 90, and mostly 100, and AUC and F-Score close to 1. Regression performance was also good with R-Squared and Adjusted R-Squared error just under 1 (0.99), and RMSE of 0.16 and 0.01 using linear regression and voting regression respectively.

Using LDA with 10 and 15 features did not affect the classification or regression performance for the first or second datasets. It slightly improved the performance of the third dataset but not by much. On the whole dataset 3 continued to perform poorly for both classification and regression algorithms with LDA. For dataset 4 and 5, the performance with 5 features was quite optimal with little room for improvement, and increasing the number of features did not impact the classification and regression performance by much. The lack of significant change on increasing or decreasing the number of features may be because the key information is contained in the first few features.

PCA:

Using PCA with 5 features, for Dataset 1, the Classification performance was average, with accuracy between 80 and 86, precision between 85 and 89, recall between 89 and 94, AUC between 0.6 and 0.88 and F-Score between 0.88 and 0.92. The regression performance was average with adjusted R-Square values of 0.19 and 0.25 (slightly above 0.19) and RMSE values of 0.35 and 0.34 for linear regression and voting regression respectively. All in all, for dataset 1, LDA performed better than PCA.

For Dataset 2, we get below average results with accuracy between 79 and 80, precision between 58 and 72, recall between 13 and 30, AUC between 0.56 and 0.62, and F-Score between 0.22 and 0.41. The regression performance was poor, with adjusted R-Square values of 0.102 and 0.147 (less than 0.19 showing a poor fit), and RMSE of 0.39 and 0.38. For this dataset, the performance with PCA was worse with classification algorithms than LDA, and similar for regression.

For Dataset 3, the classification performance was even worse, with accuracy, precision and recall close to zero, similar to the performance with LDA on this dataset. Regression performance was worse, with Adjusted R-Squared errors of 0.12 and 0.15 (less than 0.19 showing a poor fit), and RMSE ranging from 1109 and 1082 for linear regression and voting regression respectively.

For Dataset 4, classification performance was well with all algorithms providing accuracy, precision and recall of over 90, and AUC and F-Score close to 1. Regression performance was also good with R-Squared and Adjusted R-Squared error 0f 0.64 and 0.96 (greater than 0.19), and RMSE of 0.29 and 0.09 using linear regression and voting regression respectively. Both classification and regression performance for this dataset were good with PCA, but they were even better with LDA.

For Dataset 5, we get good classification results. With accuracy, we get varying numbers ranging from 83 to 98, with precision we obtain 100 in every case, the recall was between 97 and 100, and F-Score between of 0.9 and above. The regression performance was also good, with adjusted R-Square values of 0.91 and 0.98 (greater than 0.19 showing a good fit), and RMSE of 0.44 and 0.17 for linear regression and voting regression respectively. The performance of PCA for this dataset was good, but the performance with LDA was just a little better with classification algorithms than LDA, and similar for regression. However, the difference was minimal between PCA and LDA results.

Using PCA with 10 and 15 features did not affect the classification or regression performance for any of the datasets. For dataset 1, 4 and 5, the performance with 5 features was quite optimal with little room for improvement, and increasing the number of features did not impact the classification and regression performance by much. The lack of significant change on increasing or decreasing the number of features may be because the key information is contained in the first few features.

XGBoost:

For the first dataset, XGBoost was yielding an average classification performance with accuracy ranging from 75 to 86, precision from 75 to 87, recall from 89 to 100, AUC from 0.5 to 0.76 and F-Score from 0.85 to 0.91. The regression results were good, with adjusted R-Square between 2 and 3 (greater than 0.19 showing a good fit), and RMSE of about 0.3. All in all, worse performance as compared to LDA and PCA.

For the second dataset, the classification results were abysmal. The accuracy ranged from 78 to 82, but the other metrics like precision ranged from 0 to 100 using different algorithms, the recall also varied from 13 to 37, and no combination of metrics was favorable for this dataset. The regression metrics were also not good, with Adjusted R-Squared of 0.12 and 0.18 (less than 0.19 showing a poor fit), and RMSE of 0.38 and 0.37 using linear regression and voting regression respectively.

For the third dataset, the classification results were even worse, with maximum accuracy of 1.23, and the rest of the metrics were also not favorable. The regression results were better but not by much. The Adjusted R-Squared was 1.0 and 0.9 (greater than 0.19 showing a good fit), but the RMSE was 1.04 and 365 using linear regression and voting regression respectively.

For the fourth dataset, the results were very good, just like the results obtained for PCA and LDA, with the accuracy, precision and recall approaching 100 in most cases, and the AUC and F-Score approaching 1. The regression results were also favorable, with Adjusted R-Squared of 0.8 and 0.9 (greater than 0.19 showing a good fit), and RMSE of 0.19 and 0.1 using linear regression and voting regression respectively.

For the fifth dataset, the classification results varied between algorithms. Two algorithms had accuracy of around 62 to 63, precision of around 91 to 94, and recall of around 86 to 100, while two algorithms had accuracy, precision and recall of 100. The results were worse than PCA and LDA, with LDA having the best results for this dataset yet. The regression results were also good for this dataset with Adjusted R-Squared of 0.98 and 0.99 (greater than 0.19 showing a good fit), and RMSE of 0.19 and 0.11 using linear regression and voting regression respectively.

TSNE:

For the first dataset, the classification results are good with accuracy ranging from 83 to 84, precision ranging from 85 to 88, recall from 92 to 95, AUC 0.64 to 0.77 and F-Score of 0.9. These metrics show that for all algorithms, the range of results was very narrow and did not vary widely. The results are not as good as the ones obtained by PCA and LDA. The regression results were not good, with Adjusted R-Squared of 0.03 and 0.07 (less than 0.19 showing a poor fit), and RMSE of 0.19 and 0.11 using linear regression and voting regression respectively.

For the second dataset, the classification results are average with accuracy ranging from 78 to 80, precision ranging from 53 to 73, recall from 7 to 34, AUC 0.53 to 0.62 and F-Score of 0.13 to 0.41. These metrics show that for all algorithms, the range of results varied widely. The results are not as good as the ones obtained by PCA and LDA. The regression results were not good, with Adjusted R-Squared of 0.08 and 0.13 (less than 0.19 showing a poor fit), and RMSE of 0.39 and 0.38 using linear regression and voting regression respectively.

For Dataset 3, the classification performance was abysmal, with accuracy, precision, recall, AUC and F-Score all very close to zero, similar to the performance with other DR measures on this dataset. Regression performance was worse, with Adjusted R-Squared errors of 0.05 and 0.6, but RMSE ranging from 1151 and 677 for linear regression and voting regression respectively.

For Dataset 4, the classification performance was very good, with accuracy, precision and recall all greater than 90 with all algorithms, and AUC and F-Score greater than 0.90. As for the regression performance, the Adjusted R-Square values of 0.75 and 0.92, which are greater than 0.19, show that it was a good fit, and the RMSE values of 0.14 and 0.24 are substantially low, for linear regression and voting

regression respectively, which shows that the regression results are also very good.

For Dataset 5, the classification performance was average, with Random Forest yielding the best results with accuracy of 98, precision of 100 and recall of 98. The second best performance was of SVM, with accuracy of 82, precision of 100 and recall of 96. LR and XGBoost did not perform very well, which is why, on the whole, the performance was average. The regression performance was good, with adjusted R-Square of 0.2 and 0.9, and RMSE of 1.2 and 0.4 for linear regression and voting regression respectively

Note: TSNE was run only once with 3 features because it was very time consuming and the file kept crashing when it was being run with 5 and 7 features.

L1:

Running L1 with 25 estimators on Dataset 1, the classification results were average with accuracy ranging from 75 to 85, precision ranging from 76 to 87, recall ranging from 83 to 99, AUC ranging from 0.53 to 0.77 and F-Score ranging from 0.84 to 0.91. The regression results were very good with Adjusted R-Squared between 2 and 3.3, and RMSE of 0.4 and 0.3 for linear regression and voting regression respectively.

For the second dataset, the classification results were poor with accuracy ranging from 78 to 82, but precision ranging from 0 to 73, recall from 0 to 36, AUC 0.49 to 0.65 and F-Score of 0 to 0.47. These metrics show that for all algorithms, the range of results varied widely. The results are not as good as the ones obtained by PCA and LDA. The regression results were not good, with Adjusted R-Squared of 0.12 and 0.17 (less than 0.19 showing a poor fit), and RMSE of 0.38 and 0.37 using linear regression and voting regression respectively.

For the third dataset, the classification results were very bad, with maximum accuracy of 1.74, and the rest of the metrics were also not favorable. The regression results were very poor as well. The Adjusted R-Squared was 1.0 and 0.9 (greater than 0.19 showing a good fit), but the RMSE was 1.7 and 391 using linear regression and voting regression respectively. So the linear regression was working well, but the voting regression was not.

For fourth dataset, the classification results were very good, with accuracy, precision and recall very close to 100 for all algorithms, and AUC and F-Score close to 1. The results for regression were good also, with Adjusted R-Squared 0.86 and 0.96, and RMSE of 0.18 and 0.09 using linear regression and voting regression respectively.

For the fifth dataset, the classification performance was good as two algorithms had accuracy, precision and recall of 100, and AUC and F-Score of 1, while the two remaining ones performed poorly. The regression performance is very good with Adjusted R-Square of 0.95 and 0.99 (which is above the 0.19 limit) and RMSE of 0.31 and 0.12 for linear regression and voting regression respectively.

Running L1 with 50 and 100 estimators did not impact the results by much. The datasets that were performing poorly continued to do so, and the ones that were performing well continued to perform well. This may be because some of the datasets did not have 50 or 100 estimators, so their key information was already captured in the first 25 features.

RFE:

For the first dataset, the classification performance was good but not great, with accuracy ranging from 76 to 85, precision ranging from 76 to 86, recall ranging from 90 to 99, AUC ranging from 0.54 to 0.75 and F-Score ranging from 0.86 to 0.91. Regression performance was very good, with Adjusted R-Square values of 3.8 and 1.3 which are greater than 1.9 indicating a good fit, and RMSE values of 0.89 and 0.31 which are low values, for linear regression and voting regression respectively.

For the second dataset, the classification performance was poor. The accuracy in some cases was high, but the accompanying values of precision, recall, AUC and F-Score were very low. The accuracy ranged from 78 to 82, but the variation in precision was from 0 to 73, and the variation in recall ranged from 0 to 35. The AUC and F-Score values were also between 0 and 0.5, indicating poor results. The regression performance was also poor, with Adjusted R-Square values of 0.12 and 0.17, which are less than 0.19 showing a poor fit, although the RMSE values were 0.38 and 0.37, for linear regression and voting regression respectively.

For the third dataset, the classification performance was very poor, with accuracy values ranging from 0.7 to 1.6 only, and precision, recall, AUC and F-Score values all very close to zero. The regression performance Linear Regression proved to be a better fit with adjusted R-Squared at 1 and RMSE of 8.765105564687147e-12, which is very low. Voting regression yielded a bad result with adjusted R-Squared at 1 but RMSE of 394.

For the fourth dataset, the classification results were very good, with accuracy, precision and recall very close to 100 for all algorithms, and AUC and F-Score close to 1. The results for regression were good also,

with Adjusted R-Squared 0.93 and 0.95, and RMSE of 0.11 and 0.10 using linear regression and voting regression respectively.

For the fifth dataset, the results were a mixed bag, with two algorithms, namely XGBoost and Random Forest performing perfectly with accuracy, precision and recall at 100, and AUC and F-Score at 1. The remaining two algorithms performed poorly. The regression performance was also good, with adjusted R-Square of 0.98 and 0.99 showing a good fit, and RMSE of 0.18 and 0.11 using linear regression and voting regression respectively.

RF:

For the first dataset, with 25 features, classification performance for XGBoost and Random Forest was good, but performance for Logistic Regression and SVM was not up to par. Increasing the features to 50 improved the performance of Logistic Regression but not of SVM. Increasing the features further up to 100 deteriorated the performance of both LR and SVM, while there was no impact on the performance of XGBoost and Random Forest of increasing features. The performance of regression metrics with both linear and voting regression remained favorable throughout, with adjusted R-Squares greater than 0.19 and low RMSE.

For the second dataset, the classification and regression performance was very poor, and increasing the number of features had no impact on the performance. Regression performance was similarly poor.

For the third dataset, classification performance was very poor, however, regression performance with linear regression was really good.

For the fourth dataset, classification and regression performance were both very good, and near to perfect.

For the fifth dataset, classification performance was good with XGBoost and Random Forest, but poor with LR and SVM. Increasing the features had no major impact and the classification algorithms continued to perform as such. Regression performance was good in all cased.