Marking rubric for labsheet 4

Criteria	weight	Failed (0)	Passed (50)	Good (70)	Excellent (100)
Project 1 (a) **Data loading, visualisation, and cleaning**	10.00%	Not attempted or incomplete	Basic code was there but needed some modification in order to run. * Data read in okay. * The code attempted to remove two features. It seems to be a bit random how these features were chosen. *** Data file was included in the submission or must be downloaded from the web.	Basic code was there and the code ran without problems. Some explanation was clear and easy to understand. * Data read in successfully; visualisation was illustrated and some explanation was given. * Two features were removed. Some explanation was given. * One-hot encoding was done.	Basic code was there and the code ran without problems. Explanation was clear and easy to understand. * Data read in successfully; visualisation was illustrated and explained well. * Two features were removed and described and justified clearly. * One-hot encoding was done correctly. * Dimensions of features before and after data cleaning were reported.
Project 1 (b) **AdaBoost Regressor – set up, training, and prediction**	17.00%	Not attempted or incomplete	* Data were randomly split into 85/15 for the training and test sets. * Base estimator SVR was set up with an RBF kernel. * AdaBoostRegressor was set up. Some explanation was given. * The regressor was trained successfully on the training set. * The regressor was tested successfully on the test set. Outputs from intermediate estimators were predicted. * The predicted results were integers. Only 3-4 of the 6 bullet points were implemented correctly. Some modification required to get the code running.	* Data were randomly split into 85/15 for the training and test sets. * Base estimator SVR was set up with an RBF kernel and other appropriate hyperparameter values. * AdaBoostRegressor was set up with appropriate hyperparameter values. Some explanation was given. * The regressor was trained successfully on the training set. * The regressor was tested successfully on the test set. Outputs from intermediate estimators were predicted. * The predicted results were integers.	* Data were randomly split into 85/15 for the training and test sets. * Base estimator SVR was set up with an RBF kernel and other appropriate hyperparameter values. Well explained. An operation to optimise the performance of SVR was considered and explained. * AdaBoostRegressor was set up with appropriate hyperparameter values. Code was well explained. * The regressor was trained successfully on the training set. The code was neat and efficient. * The regressor was tested successfully on the test set. Outputs from intermediate estimators were predicted. Code was neat and well explained. * The predicted results were integers. * Meaningful variable names.
Project 1 (c) **AdaBoostRegressor – results and plots**	30.00%	Not attempted or incomplete	* Code for generating the MAE plots worked okay. Some explanation was given in markdown cells. * Code for generating the prediction results on both the training and test sets ran okay. Some explanation was included. * Code for generating the raw error distributions on both sets ran okay. Explanation was given in markdown cells. Either (1) One or both functions lacked modularisation; or (2) no functions were provided. Some modifications were required.	* Code for generating the MAE plots worked okay. Some explanation was given in markdown cells. * A function for generating the prediction results on both the training and test sets ran okay. Explanation was given in markdown cells. * A function with suitable arguments and comments for generating the raw error distributions on both sets ran okay. Explanation was given in markdown cells. One or both functions lacked modularisation.	* Code for generating the MAE plots worked well. The plots were self-explanatory with proper axis labels, etc. Explanation was given in markdown cells. * A function, with suitable arguments and comments, for generating the prediction results on both the training and test sets ran okay. The code was modularised. The plots were self-explanatory with proper axis labels, etc. Explanation was given in markdown cells. * A function, with suitable arguments and comments, for generating the raw error distributions on both sets ran okay. The code was modularised. The plots were self-explanatory with proper axis labels, etc. Explanation was given in markdown cells. * Meaningful function names.
Project 1 (d) **AdaBoostRegressor – Comparison and summary**	3.00%	Not attempted or incomplete	Incomplete conclusion. Overall presentation could be improved.	Some comparison and conclusion was given.	Good comparison and summary. Overall presentation was excellent with good use of markdown cell(s).
Project 2 (a) **Data loading and visualisation**	2.00%	Not attempted or incomplete	* Code needed some modification to run. * Some visualisation was shown. * Data was split into training and test sets. *** Data file was included in the submission or must be downloaded from the web.	* Data read in without problems. * Some basic code for visualising the data was given. * Data were randomly split into 85/15 for the training and test sets. * Some explanation was given in the markdown cell(s).	* Data read in without problems. * Some basic code for visualising the data was given and explained. * Data were randomly split into 85/15 for the training and test sets. * Good explanation was given in the markdown cell(s). * Feature scaling was either implemented or not implemented, with explanation. * Code ran perfectly.
Project 2 (b) **RF Regressor on full-dimensional data**	13.00%	Not attempted or incomplete	* A RF Regressor was implemented with appropriate hyperparameters. * Predicted outputs were rounded to integers. * MAEs on the training and test sets (on the original data) were computed and reported. * Histograms (or bar charts) on the raw prediction errors on both sets were displayed and well explained. * Plots were self-explanatory with good layout (not too large/small). Only 2-3 of the above were implemented; code was messy or needed some help to run.	* Code ran okay. * A RF Regression was implemented with appropriate hyperparameters. * Predicted outputs were rounded to integers. * MAEs on the training and test sets were computed and reported. * Histograms (or bar charts) on the raw prediction errors on both sets were displayed and well explained. * Plots were self-explanatory with good layout (not too large/small).	* Code ran okay and was neat. * A RF Regressor was implemented with appropriate hyperparameters and some explanation was given. * 2-3 hyperparameters were looked at to avoid overfitting. * Predicted outputs were rounded to integers. * MAEs on the training and test sets (on the original data) were computed and reported. The displayed outputs were meaningful. * Histograms (or bar charts) on the raw prediction errors on both sets were displayed and well explained. * Plots were self-explanatory with good layout (not too large/small). * Meaningful variable names.
Project 2 (c) **Feature importance & dimensionality reduction**	11.00%	Not attempted or incomplete	* Code for extracting the feature importances need some help to run. * Dimensionality seemed to be reduced.	* Code for extracting the feature importances ran okay. * Some code was supplied to reduce the dimension of the features. * Some explanation was given alongside the code.	* Code for extracting the feature importances ran okay. * Features to be removed and to be kept were worked out in the code and displayed meaningfully as text outputs. * Some code was supplied to reduce the dimension of the features. * Code was well explained. * Meaningful variable names.
Project 2 (d) **Random Forest Regressor on reduced- dimensional data**	11.00%	Not attempted or incomplete	* A RF Regressor was implemented with appropriate hyperparameters. * Predicted outputs were rounded to integers. * MAEs on the training and test sets (on the reduced dimensional data) were computed and reported. * Histograms (or bar charts) on the raw prediction errors on both sets were displayed and well explained. * Plots were self-explanatory with good layout (not too large/small). Only 2-3 of the above were implemented; code was messy or needed some help to run.	* Code ran okay. * A RF Regression was implemented with appropriate hyperparameters. * Predicted outputs were rounded to integers. * MAEs on the training and test sets (on the reduced dimensional data) were computed and reported. * Histograms (or bar charts) on the raw prediction errors on both sets were displayed and well explained. * Plots were self-explanatory with good layout (not too large/small).	* Code ran okay and was neat. * A RF Regressor was implemented with appropriate hyperparameters and some explanation was given. * Predicted outputs were rounded to integers. * MAEs on the training and test sets (on the reduced dimensional data) were computed and reported. The displayed outputs were meaningful. * Histograms (or bar charts) on the raw prediction errors on both sets were displayed and well explained. * Plots were self-explanatory with good layout (not too large/small). * Meaningful variable names.
Project 2 (d) **Comparison and Conclusion**	3.00%	Not attempted or incomplete	Incomplete comparison. Overall presentation could be improved.	Some comparison was reported. Overall presentation was okay.	Good comparison. Overall presentation was excellent with good use of markdown cell(s).