Marking rubric for labsheet 4

| Criteria | weight | Failed (0) | Passed (50) | Good (70) | Excellent (100) |
|---|---|---|---|---|---|
| **Project 1 (a)**<br><br>**Data loading, cleaning, visualization, and feature scaling** | 8.00% | Not attempted or incomplete | Basic code was there but needed some modification in order to run. Some basic visualization code was written but no explanation was provided. Feature scaling not done or not explained. | Basic code was there and the code ran without problems. Data were randomly split into 85/15 for the training and testing sets. The code was a bit messy and not explained well. Some limited explanation was given about feature scaling. | Basic code was there and the code ran without problems. Data were randomly split into 85/15 for the training and testing sets. Some suitable visualization code was written and well commented. Explanation was clear and easy to understand. Feature scaling, if implemented, was done **correctly.**. |
| **Project 1 (b)**<br><br>**AdaBoost Regressor** | 23.00% | Not attempted or incomplete | Basic code was there but needed some modification in order to run. The code was inefficient. Some regression results were provided. There was no (or minimal) explanation on what was done. | Basic code was there and ran okay. Some regression results were provided and some explanation about the results was given. Regression results were in integers and some MSEs were shown. Explanation was not clear. Not all the required computations and displays were implemented. | Basic code was there and ran okay and not too slow. The code was efficient. Regression results were in integers and all the MSEs of intermediate models were shown both numerically and graphically on the training and testing sets. Histograms of raw errors were shown for both sets and explained well. |
| **Project 1 (c)**<br><br>**Gradient Boosting Regressor** | 23.00% | Not attempted or incomplete | Basic code was there but needed some modification in order to run. The code was inefficient. Some regression results were provided. There was no (or minimal) explanation on what was done. | Basic code was there and ran okay. Some regression results were provided and some explanation about the results was given. Regression results were in integers and some MSEs were shown. Explanation was not clear. Not all the required displays were implemented. | Basic code was there and ran okay and not too slow. The code was efficient. Regression results were in integers and all the MSE of intermediate models were shown both numerically and graphically on the training and testing sets. Histograms of raw errors were shown for both sets and explained well. |
| **Project 1 (d)**<br><br>**Conclusions and overall presentation** | 6.00% | Not attempted or incomplete | Incomplete conclusion. Overall presentation could be improved. | Some comparison and conclusion was given. | Good comparison and good conclusion. Overall presentation was excellent with good use of Markdown cell(s). |
| **Project 2 (a)**<br><br>**Data loading, visualization, cleaning, and feature scaling** | 5.00% | Not attempted or incomplete | Code needed some modification to run. Code very messy and there was a lack of explanation. | Data cleaning code was there and ran okay. Basic code was there and the code ran without problems. Data were randomly split into 85/15 for the training and testing sets. There was some explanation about the data cleaning done in the code.Feature scaling was not done or not explained. | Data cleaning code was there and ran well. Basic code was there and the code ran without problems. Data were randomly split into 85/15 for the training and testing sets. Good explanation was given in the markdown cell(s). Feature scaling was either implemented or explanation was supplied if not implemented. |
| **Project 2 (b)**<br><br>**Random Forest Regressor on full-dimensional data** | 15.00% | Not attempted or incomplete | Code required some modification to run or insufficient explanation. An RF Regressor was implemented. Most required computations and results were missing. | Code ran okay and was not too messy. A RF Regression was implemented with appropriate hyperparameters and some explanation was given. Predicted results were rounded into integers. Not all required prediction results were shown. | Code ran okay and was neat. A RF Regression was implemented with appropriate hyperparameters and some explanation was given. Predicted outputs were rounded to integers. MSEs on the training and testing sets were computed and reported. Histograms (or bar charts) on the raw prediction errors on both sets were displayed and explained. |
| **Project 2 (c)**<br><br>**Random Forest Regressor on reduced-dimensional data** | 15.00% | Not attempted or incomplete | Code required some modification to run or insufficient explanation. An RF Regressor was implemented. Most required computations and results were missing. | Feature importance were appropriately used for reducing the feature dimension. Code ran okay and was not too messy. An RF Regression on the reduced dimensional features with the same hyperparameters as part (b) was implemented.<br><br>Same as the cell for part (b) above | Feature importance were appropriately used for reducing the feature dimension. Code was neat and explained well. An RF Regression on the reduced dimensional features with the same hyperparameters as part (b) was implemented.<br><br>Same as the cell for part (b) above. |
| **Project 2 (d)**<br><br>**Comparison and Conclusion** | 5.00% | 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). |