

Statistics for Data Analytics Final Project: Code 25%, Report 25%

City Colleges

December 2023

Task Instructions

Using Python, complete the following tasks:

1. Load your data into Jupyter Notebook.
2. Create pairwise scatterplots of variables of interest. Describe your discoveries and the relationships, if any.
3. Create box plots to identify outliers.
4. Perform data cleaning and preprocessing, this might include data amputation of the extreme variables.
5. Conduct data analysis: calculate the median, mean, and standard deviation of variables of interest. Discuss the implications of the results.
6. Create new box plots after data imputation and analyse the results.
7. Explain the concept of multicollinearity in the context of the dataset. Apply PCA and visualize Principal Component 1 and Principal Component 2.
8. Using Logistic Regression
 - In the SICU dataset given, your target variable is `icu_exp_flg`, which is the ICU Mortality.
 - Separate the variable `icu_exp_flg` from the main dataset to create two distinct datasets: one containing all independent variables (X) and another containing only the target variable (y). This process ensures that the dataset for X includes data for all features excluding `icu_exp_flg`, while the dataset for y exclusively holds the values of the target variable `icu_exp_flg`.
 - After doing this, Split your dataset into train and test data.
 - Normalize the data using `StandardScaler`.

- Create and train a logistic regression model.
- Make predictions on the test set and Evaluate the performance of the logistic regression model.
- Print the Accuracy and Confusion Matrix.
- Visualize your confusion matrix, so that it appears as follows.

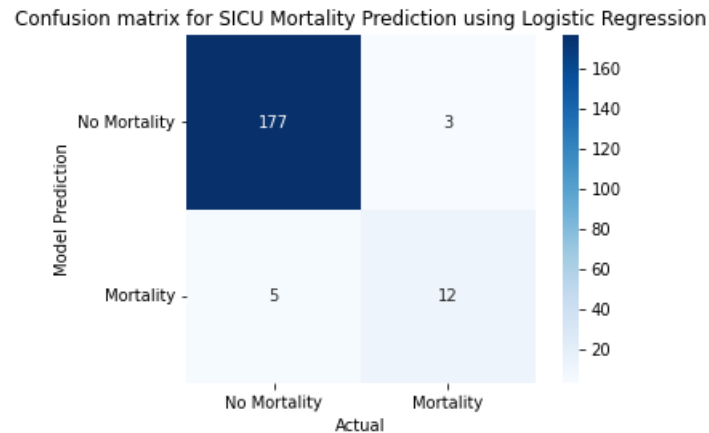


Figure 1: Confusion Matrix

9. Your results can be different from my values in the image above, but no problem about that since models can perform differently.
10. Explain the results in your confusion matrix.

Please write a report of 1400 Words, this will include all the steps you have done, explain why you have used some libraries, elaborate on some of the steps you have undertaken, and why. add some screenshots of your findings. The report will be inform of a Pdf.

Please Plagiarism is strictly prohibited.