PROG8450 Assignment2

8680643

Fei Yun

From the dataset categories we can see “Ever in ICU” might be a important measure to classify the if the client has been severe case.

Here we found 2 categories may influence “Ever in ICU” which is “Client Gender” and “Age Group”. From general research we already know that compare female, male is easier to become severe cases, people in age of 55 or higher has the highest risk of becoming severe cases.

From the dataset we can see the categories are may not have linear shape, then the leaner model cannot capture a linear feature. Random forest model is fit for this situation that it is stronger for classifications, at same time it has better performance on big volume data.

To identify the correlation of gender, age and ICU case I transferred the age, gender,I CU data to numeric data. And used random forest to learnthe study data and predictive the test

Data.

After applied on random forest model, the test dataset accuracy is 99.1%. This result on test dataset shows that the “Client Gender” and “Age Group” have influence on “Every in ICU”. In conclusion, client gender and age have strong correlation with client’s severe case risk.

The Screenshot

Graphical user interface, text, website

Description automatically generated