AY2024 Computer Based Assessment Homework

Hospital Length of Stay Estimation at Admission

Inpatient Length of Stay (LOS) in hospitals are important estimates done upon admission (after a tentative diagnosis) to plan for hospital resources (beds, specialized equipment, doctors/nurses with specific training, etc…) and to inform the patient on the estimated cost of stay so that they can get the required approval from insurance. Currently, the estimation is done using the mean LOS in each CCSR diagnosis.

A data sample (INF002v4.csv) and data dictionary from the New York State Department of Health are provided. The length of stay value of 120 + had been recoded to 120 so that you can use LOS as a continuous integer variable.

The business objective is to estimate the LOS more accurately upon admission than using the mean LOS.

1. Explore the data and report 3 notable findings.
2. State the list of potential predictor X variables that will be inputs to models and the final dimensions of the dataset (i.e. number of observations and variables).
3. Using 70-30 train-test, conduct (a) Linear Regression, and (b) CART to compare testset errors. Explain your specification of the model[1] and display the results in a table.

Model	Complexity	Testset RMSE
Linear Regression	*State number of X variables.*	
CART	*State number of terminal nodes.*	

4. Comment on your models' results and provide insights for the business application.
5. Suggest improvements.

[1] Did you optimise each model? How? Why?