

Assessment 2: Supplement Sales Prediction

Approach:

The approach used here was supervised learning. As the data contained was labeled data. We need to predict the sales of each store, which was a numerical outcome. Multiple models could be used in predicting the sales data like Linear Regression, Ridge Regression, Lasso, K-Neighbours and Decision tree model.

After importing all the necessary libraries, I did data inspection on checking whether data has any null values and if the volume of null values is going to effect the predictive power of the model. Once data cleanup was done, next thing was to understand the data itself.

I used Data visualisation technique like checking volume of sales for each store type, location type and region code and also visualising store type and sales in barplot. Which gives a brief knowledge on sales is higher for store type (S4), location type (L2) and region code (R1).

After visualising checked the categorical variables and created dummy variables and concatenated it with test data frame.

Split the data into X and Y variable using train_test_split method for regression technique. Applied LinearRegression model to fit X_train and Y_train. Once the fit was done applied Predict method on X_train data.

Also calculated the root mean square error using metrics method. Also used RFE (recursive feature elimination) to reduce some variables. Once that was done calculated the error distribution which was found to be normally distributed. Once the model was developed on train data, used the model to predict the sales value on test data.