I learned about predictive modelling using many techniques. Simple linear regression deals with one dependent variable and one independent variable. It follows the equation y=mx+b. where y is response variable. X is independent variable. M is slope and b is the intercept. We use lm() in R to create regression.

I also learned about logistic regression model which can predict binary outcomes. Unlike linear regression, which predicts continuous values, logistic regression predicts probabilities. We use glm() function in R to create logistic regression. In real world application we use this regression for spam detection, medical diagnosis etc.

I learned about multiple linear regression. It extends simple linear regression by using more than one predictor variable. It can be represented as

Y=b0​+b1​x1​+b2​x2​+...+bn​xn​+ϵ

b0 is the intercept and x1, x2,x3 are the independent variable. b1, b2 are the coefficient. ϵ is the error. We use similar lm() in R to model multiple linear regression. I learned about effect modification (also known as interaction) occurs when the effect of one independent variable (predictor) on the dependent variable (outcome) changes depending on the level of another independent variable.

Next section I learned about supervised and unsupervised learning. Supervised learning has labelled data which predicts output using regression and classification. Unsupervised learning has no labels and rather it finds patterns, structures and relationships.

I also learned about confusion matrix used to evaluate the performance of a classification model by comparing predicted vs. actual values. These are the 3 terms associated with confusion matrix.

Accuracy: Overall correctness of the model.

Precession: How many predicted positives are actually correct.

Recall: How many actual positives were correctly identified

F1 Score: Balance between Precision and Recall

Specificity: True negative rate

Sensitivity: True positive rate same as Recall