EDUPRISTINE

Empowering Professionals

Predictive Business Analytics

Linear Regression in R – Case Study

How do you identify the student with chances of default

You are data analyst with the police department. The police wants to identify the number of crimes that will happen in a particular region so that it can manned accordingly. You are given a task to estimate the crimes based on historical data

Please use Linear Regression to solve the problem

Details of the dataset

- **R:** Crime rate: # of offenses reported to police per million population (target variable)
- **Age:** The number of males of age 14-24 per 1000 population
- **S:** Indicator variable for Southern states (0 = No, 1 = Yes)
- **Ed:** Mean # of years of schooling x 10 for persons of age 25 or older
- **Ex0:** 1960 per capita expenditure on police by state and local government
- **Ex1:** 1959 per capita expenditure on police by state and local government
- **LF:** Labor force participation rate per 1000 civilian urban males age 14-24
- **M:** The number of males per 1000 females
- **N:** State population size in hundred thousands
- **NW:** The number of non-whites per 1000 population
- **U1:** Unemployment rate of urban males per 1000 of age 14-24
- **U2:** Unemployment rate of urban males per 1000 of age 35-39
- **W:** Median value of transferable goods and assets or family income in tens of $
- **X:** The number of families per 1000 earning below 1/2 the median income

High level steps to be followed for solving the problem

- Conversion of business problem into analytical problem
  - Identification of the dependent variable
- Data Import
- Exploratory data analysis
  - Missing Value Detection and Treatment
  - Outlier Detection and Treatment
  - Dummy variable Creation
  - Creation of train and test samples
- Checking Correlation
- Checking Multicollinearity
- Model Development
  - Removal of insignificant variables
  - Checking the diagnostics
- Prediction on Test data
  - Diagnostics of the test data
  - Checking Accuracy and quality of the model