Chapter 27
Predictive Modelling for Future Technology Development

Preeti Bala
KKMAII, India & Stratford University, New Delhi, India

ABSTRACT

The prediction models in machine learning and the application of different algorithms help in developing appropriate knowledge of future behaviour based on past data. With a view to help developing insights, this chapter explains the basics of predictive analytics with its common traits. The main objective of this chapter and the topics covered in the predictive models show how different methods of tools and algorithms are crucial for predictive analytics and steps to be used to make a good model. It also describes the basic tools and techniques to be used in predictive modelling. The discussions are based on the data of live projects in predictive modelling. It attempts to handle the predictive analytics like linear regression, multivariate linear regression, nonlinear regression, and multivariate non-linear regression. These statistical tools are crucial and critical in future technology developments.

INTRODUCTION

Technology development requires accurate calculation and computation of the data that are required for the correct usage and application. This is possible only when a modelling is tried and tested before launching the technology for real time usage. A close analysis of the past data can only help in understanding the possible behaviour of the future events. Predictive modelling is the fundamental sine qua non for future technology developments and this needs predictive analytics. Predictive modelling and predictive analytics are terms which are interchangeably used (Joseph M. Carew, 2020), In predictive modelling, a mathematical process is involved leading to the prediction of future events or outcomes by analysing past patters of the data. It is always endeavoured to explore the most likely eventuality based on the past behaviour of various data. This requires data collection which has to be carried out selectively and meticulously. Once the data is available, the analyst can use statistical models using the historical data. Due care is to be taken to avoid data errors like the Alpha Error or Beta Error. According to Statistical

DOI: 10.4018/978-1-7998-8327-2.ch027

theorems, large sample data may not be improving the accuracy of prediction to the Beta Error. It is said that " All models are wrong, but some are useful" (Joseph M. Carew, 2020),

Availability of data

How does Regression works -Regression Analysis is a statistical process for estimating the relationship among variables. IT includes techniques for analyzing relationship between several variables, that is between a dependent variable and one or more independent variables. That means regression helps one understand how a value of the dependent variable varies when anyone of the independent variables is changed, while the other independent variables remain fixed. In most cases, regression estimates the average value of the dependent variable when the independent value doesn't change.

In some cases, the focus is on the other location parameter of the conditional distribution of the dependent variable given the independent variables. The goal of estimation is to obtain a function of the independent variables, called the regression function, which can be linear or nonlinear. To characterize the variation of the dependent variable around the regression function is also an agenda of regression analysis, which can be described by a probability distribution. In regression analysis, to best fit a set of data observations that you provide, the values of parameters for a function are to be determined. Equation expresses these relationships in symbols. It shows that regression is the process of estimating the value of a continuous target (y) as a function (F) of one or more predictors (x1, x, an), a set of parameters (01, 0,, O), and a measure of error (e).

$Y=f(x,O)+e$

Multivariate Linear Regression

The term multivariate linear regression refers to linear regression with two or more predictors (x_1, x_2, …., x_n). The regression line cannot be visualized in two-dimensional space, when multiple predictors are used; still the line can be computed by expanding the equation for linear regression to include the parameters for each of the predictor variables.

$$Y = \Theta_1 + \Theta_2 x_1 + \Theta_3 x_2 + \ldots + \Theta_n x_{n-1} + e$$

Regression coefficients: The regression parameters are often referred to as coefficients in multivariate linear regression. In this case, the algorithm computes a coefficient for each of the predictors used by the model. The impact of the predictor x on the target y is given by the coefficient of it. To analyze the regression coefficients and to evaluate how much the regression line fits the data, many statistics such as **R-squared value** are available.

Confidence Bounds

For each case in the scoring data, a regression model predicts a numeric target value. Also, some regression algorithms can identify confidence bonds in the form of the upper and the lower boundaries of interval front/period in which the predicted value is supposed to lie.