CHAPTER 7

Statistical Model and Method in Regression Based Prediction

Nurulhuda Firdaus Mohd Azmi, Nilam Nur Amir Sjarif, Norziha Megat Zainuddin, and Istas Fahrurrazi Nursyirwan

7.1 INTRODUCTION

Regression is a statistical measure to investigate the relationship between a dependent variable or response variable, Y and one or more independent or explanatory variables, X_1, X_2..., X_k. The three (3) basic parameters of regression model are the straight-line slope, the intercept and the conditional variance of the distribution of values around the straight line. The purpose of regression model can be used to forecast, model time series, or determine and explain the relationship between variables and predict continuous values (Sharma, 2021). There are three (3) main types of regression techniques: Simple Regression, Multiple Regression and Non-Linear Regression. Simple regression uses just one (1) independent variable, X to predict the outcome of the dependent variable, Y, whereas multiple regression employs two (2) or more independent variables to predict the result of the dependent variable, Y. For simple and multiple linear regression, the response variable, Y must be a quantitative scale (continuous) observations/ record.

There are also many instances in which the relationship between two (2) variables will be curvilinear, rather than linear. To model such curvilinear relationships, it needs to incorporate terms into the multiple regression model that will create "curves" in the model. The model which possesses the curvilinear is refer as a

polynomial model. The order or degree of the model is determined by the largest exponent of the independent variable in the model. As more curves appear in the data, the order of the polynomial must be increased.

In building statistical based regression for prediction, it comprises three (3) main phases of model and method namely Model Specification, Model Building and Model Diagnostic. These phases are significant as the cycle of building prediction analysis in order to produce reliable and valid prediction.

This chapter will explain model and method in building statistical based regression for prediction whereby, each phase of model will be explicitly and described. It starts with Section 7.1 which present about regression analysis in general. Next, Section 7.2 present some fundamental aspect of regression including the differences between statistical approach with machine learning approach and the statistical assumption in building statistical based regression model. Section 7.3 describes the methodology of building regression-based prediction that include model specification, model building and model diagnostic. Finally, Section 7.4 will provide the conclusion of the chapter. We hope this chapter will shed understanding to the readers with the right process and methodology in building statistical based regression leading to effective and precise prediction model.

7.2   REGRESSION ANALYSIS

Regression analysis is a statistical tool that is used for two (2) main purposes: description explanation and prediction. There are many situations in which the prime purpose of regression analysis is description explanation. For example, economists use regression analysis for descriptive purposes as they search for a way of explaining the economy. Market researchers also use regression

analysis, among other techniques, in an effort to describe and examine the significant factors that influence the demand for products (Hernan & Robins., 2017).

Regression with the purposes of making prediction seek to develop a model that can be used to predict the values of a dependent or response variable based on the values of one or more explanatory or independent variables. In building regression-based prediction model, it involves a lifecycle of processes as shown in Figure 7.1. These processes are business understanding, model specification, choice of algorithm, model building, model diagnostics, and model deployment.
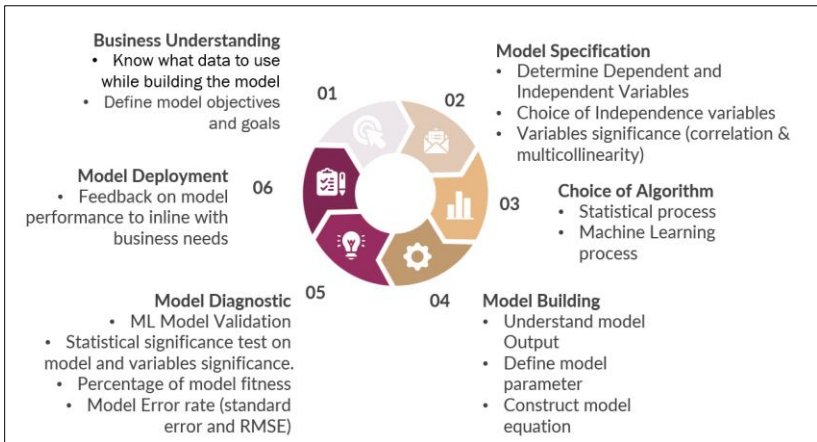

Figure 7.1 Prediction model process

Application of regression for prediction are forecasting, time series modelling and finding the explanation on the independent variable relationship between the variables. For example, in sales regression forecasting help businesses understand how their sales teams are or are not succeeding and what the future could look like based on past sales performance. The results can also be used to predict future sales based on changes that haven't yet been