OPEN ACCESS

Volume: 9

Special Issue: 1

Month: May

Year: 2022

P-ISSN: 2321-788X

E-ISSN: 2582-0397

Impact Factor: 3.025

Citation:
Jeyakumar, M., et al. "Evaluation of Logistic Regression and Multivariate Adaptive Regression Spline Models for the Analysis of Factors Determining Entrepreneurial Success." *Shanlax International Journal of Arts, Science and Humanities*, vol. 9, no. S1, 2022, pp. 6–10.

DOI:
https://doi.org/10.34293/sijash.v9iS1-May.5933

Evaluation of Logistic Regression and Multivariate Adaptive Regression Spline Models for the Analysis of Factors Determining Entrepreneurial Success

Dr. M. Jeyakumar
Associate Professor and Head, PG and Research Department of Economics
Saraswathi Narayanan College, Madurai

Dr. M. Meena
PhD and Research Department of Economics
Saraswathi Narayanan College, Madurai

Dr. K. Subramanian
Assistant Professor, PG and Research Department of Economics
Saraswathi Narayanan College, Madurai

Abstract

The present research work aims at evaluating Logistic Regression (LR) and Multivariate Adaptive Regression Spline (MARS) models through comparing the results of analysis of factors contributing to entrepreneurial success in Madurai District. The study considered two financial indicators namely Profit Rate and Reinvestment Rate for defining success in entrepreneurial venture. It analysed five factors influencing entrepreneurial success among micro and small scale manufacturing enterprises in the district. Previous business experience was the most important factor contributing to entrepreneurial success, followed by Entrepreneur's Training and Entrepreneurial orientation. Using Number of significant factors, Accuracy, Sensitivity and Specificity, the results of LR and MR models have been compared. The study validates that MARS model is more powerful and reliable than LR model in predicting the possibility of entrepreneurial success in the business venture.

Keywords: Profit Rate, Reinvestment Rate, Entrepreneurial Success, Karl Pearson Correlation, Logistic Regression, and Multivariate Adoptive Regression Splines.

Introduction

Entrepreneurship is a creative and career-oriented activity. It generates income and employment through accelerating investment. It boosts exports to make an economy self-reliant. Successful entrepreneurs have the ability to improve the standard of living of people. They emerge as National assets. They are innovative, risk-taking and proactive. On the other hand, unsuccessful entrepreneurship results in fall in employment opportunities and bank money being locked up in the form of Non-Performing Assets.

Thus, it is imperative to understand what makes an entrepreneur successful in the business venture. However, only a few studies have attempted to bring out the factors underlying entrepreneurial success and predict its possibility. The purpose of this study is to exercise MARS model, a data mining technique, with the widely used LR model for analysing the influential factors and predicting the success in the venture. The ultimate aim is to compare the results and evaluate the efficacy of MARS model in prediction.

Materials and Methods

The present research paper is completely based on primary data collected from the registered micro and small scale manufacturing enterprises in Madurai district. A sample of 60 enterprises has been chosen for making a comparative analysis of entrepreneurial success using LR and MARS models. The sample units have been selected based on purposive sampling method.

Two financial ratios namely Profit Rate and Reinvestment Rate were considered for defining entrepreneurial success. Profit rate refers to the ratio of net profit to total capital asset. Reinvestment rate is the ratio of net profit reinvested to total net profit for modernisation and expansion of the enterprise. An, entrepreneur makes earns at least 20 per cent return on capital (profit rate) annually and reinvesting at least 10 per cent of the net profit he/she has reaped has been defined as successful entrepreneur in the present analysis. Entrepreneurial Orientation (X_1), Category of Enterprise (X_2), Entrepreneur's Training (X_3), Entrepreneur's Education (X_4) and Previous Business Experience (X_5) have been taken as factors determining entrepreneurial success.

Both LR and MARS models were fitted for the comparative analysis of the influential factors of entrepreneurial success. The LR Model is a multivariate regression exercised to study the relationships between independent variables and binary dependent variable. The dependent variable is dichotomous showing the success or failure of entrepreneurship, with the value of 1 or 0 respectively. The LR model is expressed as follows:

$$P = \frac{eZ}{1 + eZ}$$

Where P is the profitability of occurrence

Z is expressed as follows:

$$Z = \alpha+\beta_1X_2 + \beta_2X_2 + \ldots\ldots\ldots..+ \beta_5X_5$$

Where α is the intercept and βs represent the Logistic regression coefficients.

The predicted value of the dependent variable is a probability estimate. As the relationship between the dependent variable and the probability cannot be expressed as a linear function when the probability value has the upper limit of one and the lower limit of zero, probability is converted into a Logit function as shown below:

$$\text{P Logit(P)} = \ln \frac{P}{1-P} = \alpha+\beta_1X_2 + \beta_2X_2 + \ldots\ldots\ldots\ldots\ldots\ldots..+ \beta_5X_5$$

where P/1-P is the odds ratio. The natural logarithm is a linear function of the independent variables ranging from -α to +α.

Many models are followed to non-linear patterns in the data by manually adding non-linear terms including squared terms. However, to do so, specific nature of non-linearities and interactions must be known apriori. Alternatively, there are some Algorithms that are inherently non-linear. Such Algorithms search for non-linearities and interactions in the data that enable to improve the predictive accuracy. MARS is such a model to capture the non-linear relationships in the data by assessing cutpoints (Knots). It is a non-linear and non-parametric regression technique given by Friedman in 1991. It is a strategy in which data sets are partitioned into separate piecewise

linear segments (Splines) of differing gradients. It consists of forward and backward Stepwise Algorithms. It generates Basic Functions (BFs) by stepwise searching overall possible univariate knots. It uses the Adaptive Regression Algorithm for the automatic selection of knot location. The MARS model can be expressed as follows:

$$f(x) = \beta_0 + \sum_{i=1}^{n} \beta_i \lambda_i(x)$$

Where $\lambda_i(x)$ is a BF and β_i is the coefficient of the i^{th} BF. Here, n is the number of BFs in the model and all the coefficients are estimated using the Least Squares method. Generalised Cross-Validation (GCV) is used to remove redundant BFs from the MARS model. It is computed as follows:

$$GCV = \frac{\sum_{i=1}^{N} \left[y_i - f(x_i) \right]^2}{\left[1 - \frac{M + d \times \frac{M-1}{2}}{N} \right]^2}$$

Where N = sample size, $f(x_i)$ is the predicted value of the MARS model, M is the number of BFs and d is the penalising parameter. Friedman states that the optimum value of d will be in the range of $2 \leq d \leq 4$. The present analysis takes the value of d as 3. The study made use of various software packages including SPSS 20 to compute the Karl Pearson Correlation Coefficients and Logistic Regression.

Results and Discussion

Multicollinearity has been tested using Karl Pearson correlation coefficients. As there has been no strong evidence for high degree of correlation among the independent variables, all the factors have been considered in both models for the comparison of the results.The results of Logistic Regression for the factors contributing to entrepreneurial success are furnished in the following Table:

Results of LR Model

Sl. No.	Factor	Coefficient	Wald Statistics	Odds Ratio
1	Entrepreneurial Orientation (X1)	0.497 (0.428)	1.343	1.643
2	Category of Enterprise(X2)	1.543 (1.423)	1.177	4.680
3	Entrepreneur's Training(X3)	-0.836 (0.815)	1.050	0.434
4	Entrepreneur's Education(X4)	0.728 (0.455)	2.558	2.070
5	Previous Business Experience(X5)	2.321* (0.680)	11.650	10.185
	Constant	-6.429* (2.432)	6.990	---

Note: Figures in the brackets are Standard Errors
* indicates one per cent level of significance
Source: Primary Data

The natural log of odds ratio can be written as follows:

$$\text{Ln}\prod(x)/ 1 - \prod(x) = -6.429 + 0.497X_1 + 1.543X_2 - 0.836X_3 + 0.728X_4 + 2.321X_5$$

The results of Logistic Regression show that the odds of success of entrepreneur in the business venture increases by a multiplicative factor of 10.19 for the increase in X5(Previous Business Experience) by one point. The other factors are statistically insignificant in the Logistic Regression. The classification matrix indicates that the overall accuracy value of the LR Model has been 91.7per cent. The Sensitivity and Specificity values have been 93.9 and 88.9 per cent. The Hosmer and Leme show Test (Chi-square = 5.268) shows that the fitted Logistic Regression Line is statistically insignificant.

The trial and error method of combination of number of Basic Function (BF), Maximum Interaction (MI) and Minimum Observation (MO) has been exercised to obtain the best MARS model. The number of MO per knot has been taken as 0 and 1, Basic Function as two to three times of the factors used in the model and MI has been taken as one to three. The results show that the best MARS model is in the combination of BF = 6, MI = 1 and MO = 0 as it has the lowest GCV value, the highest accuracy level and many significant number of the Principal Components. The MARS equation can be written as follows:

$$\text{Ln} \prod(x)/ 1 - \prod(x) = 10.699 - 0.119\ BF_1 + 0.119\ BF_3 + 0.119\ BF_4$$

Where

$BF_1 = \max(0, 117.982 - X_7)$

$BF_3 = \max(0, X_5 - 14.056)$

$BF_4 = \max(0, X_1 - 10.686)$

The results of MARS model indicate that X_7, X_5 and X_1 have relative importance of 98.2 per cent, 82.51 per cent and 56.24 per cent respectively. The results also point out that the Sensitivity and Specificity values produced in MARS modeling were 96.3 per cent and 91.8 per cent respectively.

Conclusion

When MARS has three significant factors, LR has only one significant factor. The Accuracy level of MARS has been 94.1 per cent and that of LR 91.7 per cent. The Sensitivity of MARS has been 96.3 per cent and that of LR 93.9 per cent. The Specificity of MARS has been 91.8 per cent and that of LR 88.9 per cent. Thus, the comparison of LR and MARS models in terms of these four criteria has brought out the fact that the MARS model is more powerful than LR model in predicting the possibility of success in entrepreneurial venture.

References

1. Friedman, J.H. (1991): Multivariate Adaptive Regression Splines. Ann. Stat. Vol.19, pp. 1-141.
2. Menard, S. (1995): Applied Logistic Regression Analysis, 2nd ed., SAGE University Series on Quantitative Applications in the Social Sciences, SAGE: Thousand Oaks, CA, USA.
3. Hosmer, D.W., Lemeshow, S. (2000): Applied Logistic Regression, 2nd ed., John Wiley and Sons Inc: New York, USA.
4. Mina, C.D. and Barrios, E.B. (2010): Profiling Poverty with Multivariate Adaptive Regression Splines, Philippine Journal of Development, Vol.37(2), pp.55-97.
5. Ragayu et al. (2017): Application of Principal Component Analysis (PCA) to Reduce Multicollinearity Exchange Rate Currency of Some Countries in Asia Period 2004-2014. International Journal of Educational Methodology, Vol. 3(2), pp.75-83.
6. Kilinc et al. (2017): Using Multivariate Adaptive Regression Splines to Estimate Pollution in Soil. International Journal of Advanced Applied Sciences, Vol.4(2), pp.10-16.
7. Park et al. (2017): Evaluation of Logistic Regression and Multivariate Adaptive Regression

Spline Models for Groundwater Potential Mapping Using R and Gl.Sustainability Vol.9(7), pp.1-20.

8. AnantoWibowo and M.RismawanRidha (2020): Comparison of Logistic Regression Model and MARS using Multicollinearity Data Stimulation, Vol.4(1), pp.39-47.