

# Machine Learning Approach By Applying Logistic Regression Model Using R-Programming For Predictive Financial Analytics

Dr Nitin Jaglal Untwal ,  
Associate Professor, MIT, Aurangabad

## Abstract

*Machine learning is an important aspect of artificial intelligence which emphasize on design of algorithm. The algorithms are design to act for pattern recognition which ultimately helps in decision making. The algorithms makes complex and costly manual process into a easy manageable predictive modeling. Here we have develop a predictive modeling algorithms by using r-programming for financial analytics. The method applied includes supervised learning approach. The supervised learning is applied by way of logistic regression model. The independent variable includes financials ratios and dependent variable is Return on investment which is binary. The model will be useful for predicting Return on investment.*

**Key words** Machine learning, Supervised learning, unsupervised learning, Knowledge Representation, Financial ratios, return on investment, R-programming

## Introduction

The financial data analytics is an upcoming area in today's data science arena. The predictive analytics is used to predict the outcome for a given financial data set which helps in financial decision making. The predictive analytics is a part of machine learning which includes unsupervised and supervised learning. The study emphasis on application of supervised learning for financial decision making. The logistic regression model is a tool used for supervised learning. The logistic regression model is applied to predict the outcome which is binary.

## Objective of the study

To predict the Return on investment (super ROI above 50 %) by creating a logistic regression model for Maruti Suzuki.(India)

## Period of study

The study concentrates a period starting 2008 to 2018

## Data source

Screeners financial database

## Tools used for analysis

Financial ratios , logistic regression model.

## Software used for analysis

R – programming

## Creating a knowledge representation structure by applying R-programming

The knowledge representation structure is used to prepare a raw data into reliable data structure which is used in programming as per the need and requirement of programming. The data structure is the basic need of any programming. The knowledge representation structure (data frame) is the first step towards creating and applying an algorithms. The syntax used for creating a data frame.

Data Frames

The function `data.frame ( )` creates data frames

Here we are using financial ratios as dependent variable and Return on Investment (ROI) as independent variable.

`Workingcapital <- data.frame (`

```

Year = c (2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018),
ROA = c (1,1.,1,1,1,1,0,0.,0,0),
DTO = c (27.30,21.81,35.93,44.41,37.96,29.65,30.98,46.71,43.52,56.73,54.56),
ITO = c (17.24,22.67,24.07,25.88,19.81,23.68,25.67,18.60,18.37,20.86,25.24),
DER = c (0.47,0.47,0.39,0.33,0.47,0.44,0.46,0.42,0.42,0.42,0.43),
CR = c (1.19,1.47,0.91,1.65,1.34,0.92,0.69,0.63,0.56,0.49,0.41)
    )
    
```

```
print(Workingcapital)
```

Year	ROA	DTO	ITO	DER	CR
2008	1	27.30	17.24	0.47	1.19
2009	1	21.81	22.67	0.47	1.47
2010	1	35.93	24.07	0.39	0.91
2011	1	44.41	25.88	0.33	1.65
2012	1	37.96	19.81	0.47	1.34
2013	1	29.65	23.68	0.44	0.92
2014	1	30.98	25.67	0.46	0.69
2015	0	46.71	18.60	0.42	0.63
2016	0	43.52	18.37	0.42	0.56
2017	0	56.73	20.86	0.42	0.49
2018	0	54.56	25.24	0.43	0.41

**Creating logistic model**

**Creating dependent variable is binary**

Logistic regression model is used when the dependent variable is binary (Y/N, Low/High, Male/Female).

For purpose of study we had classified Return on Investment (ROI) into a binary outcome. The criteria used for creating Y (ROI) as dependent variable on the basis of criteria.

**Criteria for converting Return on Investment (ROI) into a binary –**

ROI above 50 % denoted by 1

ROI below 50 % denoted by 0

**Mathematically expressed as**

ROI Ratio  $\geq 0.50 = 1$

ROI Ratio  $< 0.50 = 0$

**Table 1.1 Showing converted ROI into a binary**

Year	ROA	Y
2008	0.63	1
2009	1.04	1
2010	0.67	1

2011	0.58	1
2012	0.52	1
2013	0.80	1
2014	0.83	1
2015	0.40	0
2016	0.42	0
2017	0.37	0
2018	0.46	0

### Creating independent variables

The different financial ratios like Debt to equity ratio(DER), Current ratio (CR), Inventory turn-over ratio (ITR), Debtor turn-over ratio (DTO), are calculated from financial database Screener and are applied in model to predict the ROI.

The logistic regression model is represented as –

```
glm.fit <- glm(ROA~DTO+ITO+DER+CR, data = Workingcapital, family = binomial)
```

### Results

We have used the residual deviance to test whether the null hypothesis is true (i.e. Logistic regression model provides an adequate fit for the data). This is possible because the deviance is given by the chi-squared value at a certain degrees of freedom. In order to test for significance, we can find out associated p-values using the below formula in R:

**p-value = 1 - pchisq(deviance, degrees of freedom) we get results as -**

The Null Deviance of 1-pchisq (1.4421,10) and residual Deviance of 1-pchisq(3.2753,6) gives conclusive results that the model is fit.

Coefficients:				
	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-3.788e+01	1.024e+06	0	1
DTO	-1.532e+00	8.861e+03	0	1
ITO	3.424e+00	1.719e+04	0	1
DER	-2.023e+01	1.834e+06	0	1
CR	4.629e+01	2.346e+05	0	1
(Dispersion parameter for binomial family taken to be 1)				
Null deviance: 1.4421e+01 on 10 degrees of freedom				
Residual deviance: 3.2753e-10 on 6 degrees of freedom				
AIC: 10				
> 1-pchisq(1.4421,10)				
[1] 0.9991046				
> 1-pchisq(3.2753,6)				
[1] 0.7735845				
> 1-pchisq(1.4421-3.2753,10-6)				
[1] 1				

### Conclusions

The logistic regression model is fit for predicting ROI with independent variables Debt to equity ratio(DER), Current ratio (CR), Inventory turn-over ratio (ITR), Debtor turn-over ratio (DTO)

### References

1. Agarwal, S. (2017). Portfolio Selection Using Multi-Objective Optimisation. Cham, Switzerland: Springer Nature
2. Agresti, A. (2002). Categorical Data Analysis. USA: John Wiley & Sons, Inc.
3. Chavent, M., Simonet-Kuentz, V., Labenne, A. & Jérôme, S. (2017). Multivariate Analysis of Mixed Data: The R Package PCAmixdata. Available from: <https://arxiv.org/pdf/1411.4911.pdf> .
4. Jolliffe, T., I. (2002). Principal Component Analysis, 2nd Edition. New York, USA: Springer,
5. Hosmer, W., D., Lemeshow, S. & Sturdivant, X., R. (2013). Applied Logistic Regression 3rd Edition. New Jersey, USA: John Wiley & Sons, Inc.