Application of machine learning using R-programming for financial forecasting

Dr Nitin Untwal
Associate Professor, MIT Aurangabad

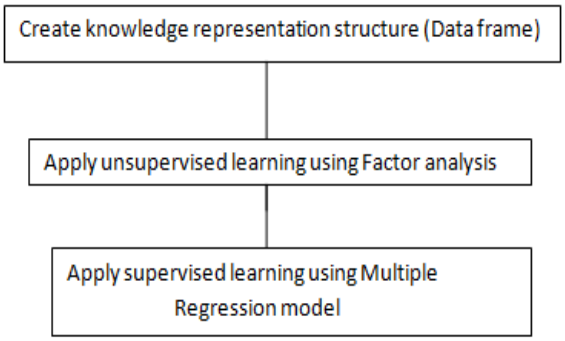
***Abstract*- Machine learning is an important branch of Artificial intelligence which includes designing and construction of algorithms. It allows computers to behave in a way based on the empirical dat. It automatically trains the system for pattern recognition which helps in decision making based on input data. The main objective of machine learning is to develop systems that are too complex or costly to develop manually. Here we had applied machine learning for financial forecasting. The researcher had applied knowledge representation structure, supervised and unsupervised learning in the form of data frame with multiple regression model and factor analysis. The unsupervised learning is applied by way of factor analysis to group the variables used in the regression model. The multiple regression model is created where financial ratios viz, Debt to equity ratio, Inventory turnover ratio, current ratio, receivables turn over ratios are independent variable and return on investment is a dependent variable.**

***Index terms*- Machine learning, Supervised learning, unsupervised learning, Knowledge Representation, working capital, return on investment, R-programming**

INTRODUCTION

Machine learning is an important branch of Artificial intelligence which includes designing and construction of algorithms. It allows computers to behave in a way based on the empirical dat. It automatically trains the system for pattern recognition which helps in decision making based on input data. The main objective of machine learning is to develop system that are too complex or costly to develop manually. Here we had applied machine learning for financial forecasting. The researcher had applied machine learning in three steps 1.Appling knowledge representation structure 2.Supervised 3. Unsupervised learning in the form of data frame with multiple regression model and factor analysis.

1.1 Structure of machine learning model applied



RESEARCH METHODOLOGY

Period of study Year 2008 –Year 2018

Data collected – Financial data is collected from screener financial data base. Stock screener is a research tool for Indian stocks analysis.

Statistical test applied – Factor analysis, Correlation, Multiple regression, residual analysis.

Financial analysis- analysis is done by applying working capital ratios and return on asset ratio

Research question: Is there any significant relationship between profitability (ROA) and DTO, DER, ITO, CR of Marutisuzuki India).

Create knowledge representation structure (Data frame)

A large amount of knowledge and some mechanism for manipulating that knowledge is the key of problem solving. The basic goal of knowledge representation is to assist inference (conclusions) from knowledge further knowledge representation structure helps in achieving complex data structure which contains information about concept being described.

We use knowledge representation structure in the form of data frame. Frames were introduced by Mervin Minsky in 1975. Knowledge about an item is stored in a single unit called frame. Frames are structures about collection of slots and slot values. Slots includes values, names and such field called links. After applying data representation we create a data frame in R-programming by name Working capital where we use Debt to equity ratio(D/E), Current ratio (CR), Inventory turnover ratio (ITR), Debtor turnover ratio (DTR), Return on Asset (ROA) as slots. The frame Working capital is created by using the syntax- create data frame (Working capital) and slot values in the form of ratios discussed above are structured with values for different years i.e 2008 to 2019.

Apply unsupervised learning using Factor analysis

Unsupervised learning is used to develop classification automatically. It is applied to find out similarity in a given data sets. It can be only used when we have sufficient past data. Classification learning gives precise finding to draw meaningful conclusion. Here we have used factor analysis in order to group the different variable like Debt to equity ratio(D/E), Current ratio (CR), Inventory turnover ratio (ITR), Debtor turnover ratio (DTR),Return on Asset (ROA) which have an relationship. The factor analysis is used as a machine learning tool for determining relationship and selecting a variable by grouping them together for developing a model. We followed three steps 1.Selection of variables 2 Measuring the relationship 3. Building model. we applied factor analysis using R- programming.

INTERPRETATION OF FACTOR ANALYSIS

Uniqueness -

It measures the variance that is unique and not shared with other variables. It measures the important or relevance of variable in the factor model. The variable ROA 0.005, DTO 0.082 and ITO 0.005 are important and have more relevance in factor model while DER 0.560 CR 0.737 have less relevance in comparison since the thumb rule is greater the uniqueness lower the relevance of the variable in the factor model.

Factor loading

It indicates the weights and co-relation between each variable of the factor. The thumb rule is higher the factor loading the more relevant in defining the factor dimensionality. A negative value indicates an inverse impact on the factor model. In this case two factors are retained because both have eigenvalues more than 1. It is observed that ROA with 0.966 and CR with 0.512 defines Factor-1. Variable ROA with 0.24 and variable ITO with 0.98 defines factor 2

Proportion variance

It indicates relative weights of each factor in total variance. It explains percentage of total variance. Factor1 with 0.466 as proportion and factor 2 with 0.256 as proportion indicates total variance higher in factor 1.

Cumulative variance

It indicates the amount of variance explained by n+(n-1)factors. Factor 1 and Factor 2 accounts for of 0.722 total variance.

SS Loading

Is used to determine the value of particular factor. The thumb rule says it is worth keeping factor which has the SS loading greater than 1. SS loading for loadings Factor 1 with 2.330 and Factor 2 with 1.280 indicates that its worth to keep factor 1 in model.

Appling supervised learning using Multiple Regression model

Modeling framework:

After reviewing theoretical literature, the following best fitted variables have been driven to measure the impact of working capital management on profitability, and the equation to investigate the relationship between working capital management and profitability is as follows:

$ROA = \beta o + \beta 1 DTO + \beta 2 ITO + \beta 3 DER + \beta 4 CR + \varepsilon$

Where ROA is the return on assets ratio, DER is the Debt-Equity ratio, DTO is the debtors turnover ratio, ITO is the inventory turnover ratio and CR is the current ratio. Where ROA is dependent and the remaining are independent variables: The ε is the error term. In the above equation, β1, β2, β3, β4 are expected to be positive ($\beta 1 DTO \geq 0$, $\beta 2$ $ITO \geq 0, \beta 3 DER \geq 0$ $\beta 4 CR \geq 0$). All data was obtained

from Screener data base, sample size is 11 i.e. from 2008 to 2018.Following is the hypothesis which has to be tested in this study.

H^o: $\beta1 = \beta2 = \beta3 + \beta4 = 0$

H1: $\beta1 + \beta2 + \beta3 + \beta4 \neq 0$

Applying supervised learning using Multiple Regression model in R- Programming we get the results as –

Y = 0.1824 + (0.7232) X1 + (0.01053)X2 + (0.03124)X3 + (-0.01483)X4

CONCLUSIONS AND IMPLICATIONS

After Appling machine learning methods like knowledge representation structure ,Unsupervised learning and supervised learning we get the results as there is a positive relationship between debtors turnover (DTO) and return on assets(ROA), between inventory turnover(ITO) and ROA and DER, but there is negative relationship between Current ratio and ROA. The interpretation of results shows that by increasing debtors turnover and inventory turnover, the company can increase its profitability but there negative effect of increasing the current ratio on profitability. Therefore, the results of the research indicate that through management of ITO, DTO, DER the company can increase its profitability. Multiple R-squared is 0.9373 which indicate fitted regression lines.

REFERENCES

[1] Salauddin, D. A. (2001). "Profitability of Pharmaceutical Companies of Bangladesh" The Chittagong University Journal of Commerce, Vol.16, pp.54.

[2] Shin, H. H. and Soenen, L. (1998). "Efficiency of Working Capital and Corporate Profitability", Financial Practice and Education, Vol. 8, No. 2, pp. 37-45.

[3] Personali. "UX Optimization Glossary > Data Science > Web Analytics > Predictive Analytics". www.personali.com. Retrieved 2018-10-22.

[4] Finlay, Steven (2014). Predictive Analytics, Data Mining and Big Data. Myths, Misconceptions and Methods (1st ed.). Basingstoke: Palgrave Macmillan. p. 237. ISBN 1137379278.

Create knowledge representation structure (Data frame) using R- Programming

```
> # Create the data frame.
> Workingcapital <- data.frame (
+ Year = c (2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018),
+ ROA = c (0.63,1.04,0.67,0.58,0.52,0.80,0.83,0.40,0.42,0.37,0.46),
+ 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
1  2008 0.63 27.30 17.24 0.47 1.19
2  2009 1.04 21.81 22.67 0.47 1.47
3  2010 0.67 35.93 24.07 0.39 0.91
4  2011 0.58 44.41 25.88 0.33 1.65
5  2012 0.52 37.96 19.81 0.47 1.34
6  2013 0.80 29.65 23.68 0.44 0.92
7  2014 0.83 30.98 25.67 0.46 0.69
8  2015 0.40 46.71 18.60 0.42 0.63
9  2016 0.42 43.52 18.37 0.42 0.56
10 2017 0.37 56.73 20.86 0.42 0.49
11 2018 0.46 54.56 25.24 0.43 0.41
```

Applying unsupervised learning using Factor analysis in R-programming

```
> # Create the data frame.
> Workingcapital <- data.frame (
+ ROA = c (0.63,1.04,0.67,0.58,0.52,0.80,0.83,0.40,0.42,0.37,0.46),
+ 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)
   ROA   DTO   ITO  DER   CR
1  0.63 27.30 17.24 0.47 1.19
2  1.04 21.81 22.67 0.47 1.47
3  0.67 35.93 24.07 0.39 0.91
4  0.58 44.41 25.88 0.33 1.65
5  0.52 37.96 19.81 0.47 1.34
6  0.80 29.65 23.68 0.44 0.92
7  0.83 30.98 25.67 0.46 0.69
8  0.40 46.71 18.60 0.42 0.63
9  0.42 43.52 18.37 0.42 0.56
10 0.37 56.73 20.86 0.42 0.49
11 0.46 54.56 25.24 0.43 0.41
> Workingcapital.fa <- factanal(Workingcapital, factors=2)
>  Workingcapital.fa

Call:
factanal(x = Workingcapital, factors = 2)

Uniquenesses:
ROA   DTO    ITO    DER      CR
0.005   0.082   0.005   0.560     0.737

Loadings:
           Factor1 Factor2
ROA     0.966   0.248
DTO    -0.942   0.173
ITO     0.170   0.983
DER     0.466  -0.472
CR      0.512

                 Factor1    Factor2
SS loadings        2.330      1.280
Proportion Var     0.466      0.256
Cumulative Var     0.466      0.722

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 2.79 on 1 degree of freedom.
The p-value is 0.095
```

Applying supervised learning using Multiple Regression model in R- Programming

```
> # Create the data frame.
> Workingcapital <- data.frame (
+ Year = c (2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018),
+ ROA = c (0.63,1.04,0.67,0.58,0.52,0.80,0.83,0.40,0.42,0.37,0.46),
+ 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
1  2008 0.63 27.30 17.24 0.47 1.19
2  2009 1.04 21.81 22.67 0.47 1.47
3  2010 0.67 35.93 24.07 0.39 0.91
4  2011 0.58 44.41 25.88 0.33 1.65
5  2012 0.52 37.96 19.81 0.47 1.34
6  2013 0.80 29.65 23.68 0.44 0.92
7  2014 0.83 30.98 25.67 0.46 0.69
8  2015 0.40 46.71 18.60 0.42 0.63
9  2016 0.42 43.52 18.37 0.42 0.56
10 2017 0.37 56.73 20.86 0.42 0.49
11 2018 0.46 54.56 25.24 0.43 0.41
> ##
> ## call :
> model <- lm(ROA ~ DTO + ITO + DER + CR , data = Workingcapital)
> summary(model)

Call:
lm(formula = ROA ~ DTO + ITO + DER + CR, data = Workingcapital)

Residuals:
Min        1Q    Median        3Q       Max
-0.072296 -0.028943 -0.010314  0.009213  0.117474

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.182438   0.547649   0.333  0.75037
DTO         -0.014831   0.003142  -4.721  0.00326 **
ITO          0.031238   0.007611   4.104  0.00633 **
DER          0.723237   0.752713   0.961  0.37374
CR           0.010526   0.072035   0.146  0.88861
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06801 on 6 degrees of freedom
Multiple R-squared:  0.9373,    Adjusted R-squared:  0.8955
F-statistic: 22.43 on 4 and 6 DF,  p-value: 0.0009394
```