

Machine Learning For Predictive Financial Analytics By Applying Naïve Bayes Algorithm

Dr Nitin Jaglal Untwal,
Associate Professor , MIT Aurangabad

Abstract

Naïve Bayes is an important tool for machine classifier. It classifies the data sets by maximum posterior decision rule. In the present research study naive bayes is applied to get the probabilistic analysis. The naive bayes predicts on the basis of prior knowledge and current evidence. The variables used are financial ratios 1.Return on asset (ROA) is measures of Profitability 2.Debtors turnover ratio (DTO) Inventory turnover ratio (ITO) as measures of efficiency 3.Debt to equity ratio(DER) measures of long term solvency 4.Current ratio as measures of liquidity. The ratios are defined by using class satisfactory (within Range) and dissatisfactory (not within Range). The study is design with five step model. Decide the range for converting financial ratios into binary as satisfactory (within Range=1) and dissatisfactory (not within Range=0), Create frequency table for class, Calculate prior probability, Create likelihood table , Calculate Posterior probability and Analysis by using Confusion Matrix . The accuracy of model is 0.45

Keywords Machine learning, Supervised learning, Naïve Bayes, Prior probabilities, Likelihood probabilities, Posterior probabilities. R-programming

Introduction

Naïve bayes is supervised machine learning tool for binary or multiclass classification. When the input is binary or categorical it becomes easy to understand the naïve bayes classification technique. Naïve Bayes is an important tool for machine learning classifier. It classifies the data sets by applying maximum posterior decision rule. In the present research study naive bayes is applied to get the probabilistic analysis. The naive bayes predicts on the basis of prior knowledge and current evidence. The Naïve Bayes technique use class probabilities and conditional probabilities for predictive modeling. Class probabilities are probabilities of each class in training data set. Conditional probabilities of each input value gives each class value.

Objective of the study

To predict the Return on investment (super ROI above 50 %) by creating a naive bayes model for Maruti Suzuki.(India) and to check its accuracy.

Period of study

The study concentrates a period starting 2008 to 2018

Data source

Screeners financial database

Tools used for analysis

Financial ratios , naïve bayes model Prior probabilities, Likelihood probability, Posterior probability.

Software used for analysis

R – programming

1. Creating financial ratio into binary

The ratios are defined by using class satisfactory (within Range=1) and dissatisfactory (not within Range=0).

Table 1.1 Criteria for converting financial ratios into a binary –

| S.no | Financial ratio | Range |
|------|-----------------|-----------------------------|
| 1 | ROA | ROI Ratio ≥ 0.50 |
| 2 | ITO | ITO Ratio between 2 to 25 |
| 3 | DTO | DTO Ratio between 2 to 25 |
| 4 | DER | DER Ratio between 0.40 to 2 |
| 5 | CR | CR Ratio between 1 to 5 |

Converting financial ratios into a binary

After creating range for different financial ratios and by applying criteria to various financial ratios following table 1.2 is prepared. The table 1.2 and table 1.3 represents the class of different financial ratios year wise and define class as satisfactory (within Range=1) and dissatisfactory (not within Range=0) and Prior probabilities.

Table 1.2 Converting financial ratios into a binary –

| Year | DTO | ITO | DER | CR | ROA |
|---------------------|-----|-----|-----|----|-----|
| 2008 | 1 | 1 | 1 | 1 | 1 |
| 2009 | 1 | 1 | 1 | 1 | 1 |
| 2010 | 1 | 1 | 0 | 0 | 1 |
| 2011 | 0 | 0 | 0 | 1 | 1 |
| 2012 | 1 | 1 | 1 | 1 | 1 |
| 2013 | 1 | 1 | 1 | 0 | 1 |
| 2014 | 1 | 0 | 1 | 0 | 1 |
| 2015 | 0 | 1 | 1 | 0 | 0 |
| 2016 | 0 | 1 | 1 | 0 | 0 |
| 2017 | 0 | 1 | 1 | 0 | 0 |
| 2018 | 0 | 0 | 1 | 0 | 0 |
| Class Frequency (1) | 6 | 8 | 9 | 4 | 7 |
| Class Frequency (0) | 5 | 3 | 2 | 7 | 4 |

2. Calculating Prior probability

Prior probability is calculated before collection of new data further. It is the best coherent analysis of the probability of outcome depending on current knowledge before an trial or experiment is performed. The

prior probability of 0.63 is found for probability of satisfactory return on investment P(SROA) and The prior probability of 0.36 is registered for probability of dissatisfactory return on investment P(DROA).

1.3 Class Frequency table

| Class | Frequency ROA |
|-------|------------------|
| 1 | 34 |
| 0 | 21 |

1.4 Table showing Prior Probability of ROA

| | Class Frequency ROA | Total Frequency of ROA | Prior Probability |
|---------|---------------------------|------------------------------|----------------------|
| P(SROA) | 7 | 11 | 0.636364 |
| P(DROA) | 4 | 11 | 0.363636 |

3. Create likelihood table

With the help of likelihood table we have calculated posterior probability. A posterior probability is the probability of the parameter given by evidence in contrast to likelihood function which is vice-versa.

Posterior probability = Prior Probability + new evidence (Likelihood)

Refer Annexure-I for four Likelihood tables showing posterior probability of different parameters.

4. Calculate Posterior Probability

The highest Posterior Probability is given by probability of satisfactory financial ratios with satisfactory ROA calculated as 31 (Positive, True)

Table 1.5 Posterior Probability of satisfactory financial ratios with satisfactory ROA(Positive, True)

| Posterior Probability | | | |
|-----------------------|---|------|-----|
| P(SDTO)/P(SROA) | 6 | 0.63 | 9 |
| P(SITO)/P(SROA) | 5 | 0.63 | 7 |
| P(SDER)/P(SROA) | 5 | 0.63 | 7 |
| P(SCR)/P(SROA) | 4 | 0.63 | 7 |
| P(SROA) | | | 060 |
| TOTAL (Round-off) | | | 31 |

The Lowest Posterior Probability is given by probability of dissatisfactory financial ratios with dissatisfactory ROA calculated as 24 (Negative, False)

Table 1.6 Posterior Probability of Dissatisfactory financial ratios with Dissatisfactory ROA(Negative,False)

| Posterior Probability | | | |
|-----------------------|---|------|------|
| P(DDTO)/P(DROA) | 4 | 0.36 | 11 |
| P(DITO)/P(DROA) | 1 | 0.36 | 2 |
| P(DDER)/P(DROA) | 0 | 0.36 | 0 |
| P(DCR)/P(DROA) | 4 | 0.36 | 11 |
| P(DROA) | | | 0.36 |
| TOTAL (Round-off) | | | 24 |

5. Analysis and results using Confusion Matrix

To test the performance of the naïve bayes classifier model we need to create a confusion matrix

Table 1.7 Confusion Matrix

| N=55 | Satisfactory | Dissatisfactory | |
|------|--------------|-----------------|----|
| 1 | 20 | 11 | 31 |
| 0 | 19 | 5 | 24 |
| | 39 | 16 | 55 |

1. Out of the 31 actual instances of Satisfactory (First row), the classifier predicted correctly 20 of them.
2. Out of the 24 actual instances of Dissatisfactory (First row), the classifier predicted correctly 19 of them.
3. Out of the all 55 instance, the classifier predicted correctly 39 instances of them.

Annexure - I

1.8 Table showing Likelihood probability with attribute DTO for each class

| Likelihood probability | | | |
|------------------------|---|------|----|
| P(SDTO)/P(SROA) | 6 | 0.63 | 9 |
| P(DDTO)/P(SROA) | 1 | 0.63 | 1 |
| P(SDTO)/P(DROA) | 0 | 0.36 | 0 |
| P(DDTO)/P(DROA) | 4 | 0.36 | 11 |

1.9 Table showing Likelihood probability with attribute ITO for each class

| Likelihood probability | | | |
|------------------------|---|------|---|
| P(SITO)/P(SROA) | 5 | 0.63 | 7 |

| | | | |
|-----------------|---|------|---|
| P(DITO)/P(SROA) | 2 | 0.63 | 3 |
| P(SITO)/P(DROA) | 3 | 0.36 | 8 |
| P(DITO)/P(DROA) | 1 | 0.36 | 2 |

1.10 Table showing Likelihood probability with attribute DER for each class

| Likelihood probability | | | |
|------------------------|---|------|----|
| P(SDER)/P(SROA) | 5 | 0.63 | 7 |
| P(DDER)/P(SROA) | 2 | 0.63 | 3 |
| P(SDER)/P(DROA) | 4 | 0.36 | 11 |
| P(DDER)/P(DROA) | 0 | 0.36 | 0 |

1.11 Table showing Likelihood probability with attribute CR for each class

| Likelihood probability | | | |
|------------------------|---|------|---|
| P(SCR)/P(SROA) | 4 | 0.63 | 6 |
| P(DCR)/P(SROA) | 3 | 0.63 | 4 |
| P(SCR)/P(DROA) | 0 | 0.36 | 0 |
| P(DCR)/P(DROA) | 0 | 0.36 | 0 |

Accuracy

Now to obtain accuracy from the confusion matrix we apply the following formulae:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Accuracy} = \frac{20+5}{20+5+11+19} = 0.45$$

Recall

Recall is the ratio of the total number of correctly classified positive examples divide to the total number of positive examples. High Recall indicates the class is correctly recognized (a small number of FN).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Recall} = \frac{20}{20+19} = 0.51$$

Precision

Precision is the measures of how often it is correct when positive result s predicted

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Precision} = \frac{20}{20+11} = 0.64$$

The accuracy of model is 0.45 and its Precision is 0.64

Reference

1. R: A Language and Environment for Statistical Computing, R Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2017}, url = <https://www.R-project.org/>
2. Confusion Matrix in Machine Learning. (2020, February 23). Retrieved from <https://www.geeksforgeeks.org/confusion-matrix-machine-learning/>
3. Brownlee, J. (2019, August 12). Naive Bayes for Machine Learning. Retrieved from <https://machinelearningmastery.com/naive-bayes-for-machine-learning/>
4. Confusion Matrix in Machine Learning. (2020, February 23). Retrieved from <https://www.geeksforgeeks.org/confusion-matrix-machine-learning/>