Paper: A Model to predict the Controller Stratewgic-impact Performance

Diego Vallarino - Not to Reprint

2022-03-21

Main Idea:

Managers and investors ask for higher and faster management information, not only accounting information. Technology, new regulations (i.e., Sarbanes Oxley), unique business model's irruptions, and organizational complexity increase decision-making difficulty (Byrne and Pierce 2006). The controller must support these demands and understand the company's whole business, process, and departments. Only with this knowledge can cooperate with other managers and distribute essential information at every organization level.

This new approach to the controller's role has added substantial duties and tasks (Morales & Lambert 2013). See (Ahrens 1996, Burns & Baldvinsdottir 2005, Friedman & Lyne 1997, Mourtisen 1996, Herzog 1999, Northcott and Scapens 2007,) for reviews. These studies frequently define two separate functions for controllers: the bean-counter; a controller focused on accounting where numbers are an end in themselves, and the business controller or business advisor, where the role is focused on helping and advising other areas, give an opinion in front of different kind of operational decisions, participate in strategic decisions and measuring ESG (environmental, social and governance) company impact (Järvenpää 2007, Zoni & Merchgant 2007, Vaivio Kokko 2006, Semenova N. 2021, Serafeim G. 2021).

Even the controller's professional organizations (ICV-IGC 2007) tend to see this role change positively. This progression means that the controller role plays a more significant function in management control, and generates more predictive management accounting methods, complements budgetary and power process, designs, more convenient control systems for organization, creates closer cooperation between accountants and managers, permits a broader administrative control, and becomes a facilitator of strategic decision making (Byrne and Pierce, 2007).

Therefore, it is important to ask ourselves which machine learning model best allows us to predict the behavior of a "new generation" Controller, with a more strategic than accounting vision. Next, I present some prediction model options and their associated performance. Just for preliminary analysis.

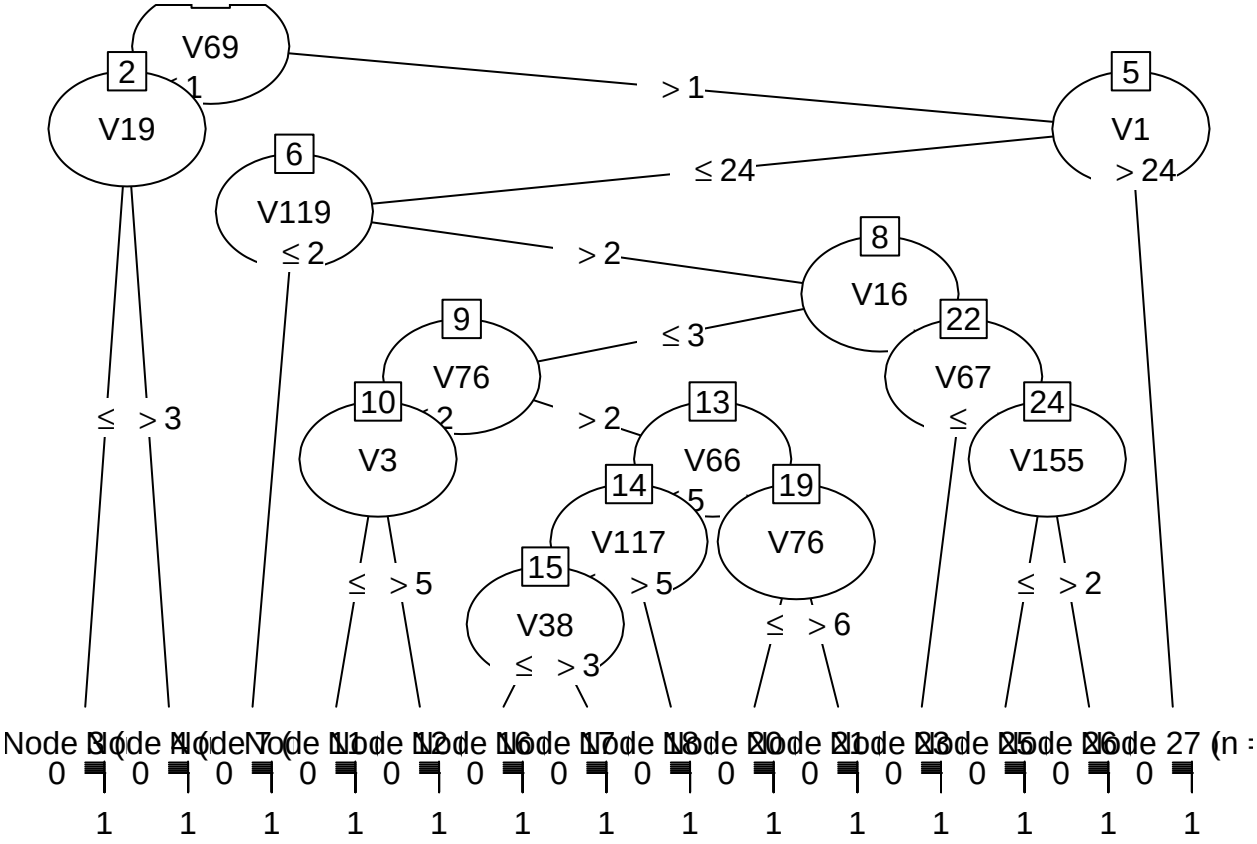
1 Desarrollo del train y test data

In this section we will only do an ETL process, to later be able to separate the data groups in training and in test.

2 Tree Model

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction 1 2 3 4 5 6 7
```

```
##          1 3 0 0 0 0 0 0
##          2 0 1 0 0 0 0 0
##          3 0 0 3 0 0 0 0
##          4 0 0 1 2 1 0 0
##          5 0 0 0 0 4 0 0
##          6 0 0 0 0 0 2 1
##          7 0 0 0 0 1 0 3
##
## Overall Statistics
##
##                Accuracy : 0.8182
##                  95% CI : (0.5972, 0.9481)
##     No Information Rate : 0.2727
##     P-Value [Acc > NIR] : 1.545e-07
##
##                   Kappa : 0.7843
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: 1 Class: 2 Class: 3 Class: 4 Class: 5 Class: 6
## Sensitivity            1.0000  1.00000   0.7500  1.00000   0.6667  1.00000
## Specificity            1.0000  1.00000   1.0000  0.90000   1.0000  0.95000
## Pos Pred Value         1.0000  1.00000   1.0000  0.50000   1.0000  0.66667
## Neg Pred Value         1.0000  1.00000   0.9474  1.00000   0.8889  1.00000
## Prevalence             0.1364  0.04545   0.1818  0.09091   0.2727  0.09091
## Detection Rate         0.1364  0.04545   0.1364  0.09091   0.1818  0.09091
## Detection Prevalence   0.1364  0.04545   0.1364  0.18182   0.1818  0.13636
## Balanced Accuracy      1.0000  1.00000   0.8750  0.95000   0.8333  0.97500
##                      Class: 7
## Sensitivity            0.7500
## Specificity            0.9444
## Pos Pred Value         0.7500
## Neg Pred Value         0.9444
## Prevalence             0.1818
## Detection Rate         0.1364
## Detection Prevalence   0.1818
## Balanced Accuracy      0.8472
```



3 Random Forest Model

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction 1 2 3 4 5 6 7
##          1 3 0 0 0 0 0 0
##          2 0 1 0 0 0 0 0
##          3 0 0 3 0 0 0 0
##          4 0 0 1 2 1 0 0
##          5 0 0 0 0 4 0 0
##          6 0 0 0 0 0 2 0
##          7 0 0 0 0 1 0 4
##
## Overall Statistics
##
##                Accuracy : 0.8636
##                  95% CI : (0.6509, 0.9709)
##     No Information Rate : 0.2727
##     P-Value [Acc > NIR] : 1.191e-08
##
##                   Kappa : 0.8374
##
##  Mcnemar's Test P-Value : NA
```

```
##
## Statistics by Class:
##
##                      Class: 1 Class: 2 Class: 3 Class: 4 Class: 5 Class: 6
## Sensitivity            1.0000  1.00000   0.7500  1.00000   0.6667  1.00000
## Specificity            1.0000  1.00000   1.0000  0.90000   1.0000  1.00000
## Pos Pred Value         1.0000  1.00000   1.0000  0.50000   1.0000  1.00000
## Neg Pred Value         1.0000  1.00000   0.9474  1.00000   0.8889  1.00000
## Prevalence             0.1364  0.04545   0.1818  0.09091   0.2727  0.09091
## Detection Rate         0.1364  0.04545   0.1364  0.09091   0.1818  0.09091
## Detection Prevalence   0.1364  0.04545   0.1364  0.18182   0.1818  0.09091
## Balanced Accuracy      1.0000  1.00000   0.8750  0.95000   0.8333  1.00000
##                      Class: 7
## Sensitivity            1.0000
## Specificity            0.9444
## Pos Pred Value         0.8000
## Neg Pred Value         1.0000
## Prevalence             0.1818
## Detection Rate         0.1818
## Detection Prevalence   0.2273
## Balanced Accuracy      0.9722
```

4 Naive Bayes Model

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction 1 2 3 4 5 6 7
##          1 2 0 0 0 0 0 0
##          2 1 1 0 0 0 0 0
##          3 0 0 2 0 0 0 0
##          4 0 0 1 2 1 0 1
##          5 0 0 1 0 5 0 0
##          6 0 0 0 0 0 2 0
##          7 0 0 0 0 0 0 3
##
## Overall Statistics
##
##                Accuracy : 0.7727
##                  95% CI : (0.5463, 0.9218)
##     No Information Rate : 0.2727
##     P-Value [Acc > NIR] : 1.523e-06
##
##                   Kappa : 0.7291
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: 1 Class: 2 Class: 3 Class: 4 Class: 5 Class: 6
## Sensitivity           0.66667  1.00000  0.50000  1.00000   0.8333  1.00000
## Specificity           1.00000  0.95238  1.00000  0.85000   0.9375  1.00000
## Pos Pred Value        1.00000  0.50000  1.00000  0.40000   0.8333  1.00000
```

```
## Neg Pred Value         0.95000  1.00000  0.90000  1.00000   0.9375  1.00000
## Prevalence             0.13636  0.04545  0.18182  0.09091   0.2727  0.09091
## Detection Rate         0.09091  0.04545  0.09091  0.09091   0.2273  0.09091
## Detection Prevalence   0.09091  0.09091  0.09091  0.22727   0.2727  0.09091
## Balanced Accuracy      0.83333  0.97619  0.75000  0.92500   0.8854  1.00000
##                       Class: 7
## Sensitivity             0.7500
## Specificity             1.0000
## Pos Pred Value          1.0000
## Neg Pred Value          0.9474
## Prevalence              0.1818
## Detection Rate          0.1364
## Detection Prevalence    0.1364
## Balanced Accuracy       0.8750
```

5 SVA Model

```
## Support Vector Machine object of class "ksvm"
##
## SV type: C-svc  (classification)
##  parameter : cost C = 1
##
## Gaussian Radial Basis kernel function.
##  Hyperparameter : sigma =  0.00114786804737896
##
## Number of Support Vectors : 50
##
## Objective Function Value : -6.0635 -8.9034 -6.8235 -6.0536 -7.9432 -4.7253 -6.3101 -5.5766 -5.0469 -5
## Training error : 0.3

## Confusion Matrix and Statistics
##
##           Reference
## Prediction 1 2 3 4 5 6 7
##          1 2 0 0 0 0 0 0
##          2 0 0 0 0 0 0 0
##          3 1 1 2 0 0 0 0
##          4 0 0 2 2 4 0 2
##          5 0 0 0 0 1 0 0
##          6 0 0 0 0 1 2 2
##          7 0 0 0 0 0 0 0
##
## Overall Statistics
##
##                Accuracy : 0.4091
##                  95% CI : (0.2071, 0.6365)
##     No Information Rate : 0.2727
##     P-Value [Acc > NIR] : 0.1178
##
##                   Kappa : 0.3286
##
##  Mcnemar's Test P-Value : NA
```

```
##
## Statistics by Class:
##
##                      Class: 1 Class: 2 Class: 3 Class: 4 Class: 5 Class: 6
## Sensitivity           0.66667  0.00000  0.50000  1.00000  0.16667  1.00000
## Specificity           1.00000  1.00000  0.88889  0.60000  1.00000  0.85000
## Pos Pred Value        1.00000      NaN  0.50000  0.20000  1.00000  0.40000
## Neg Pred Value        0.95000  0.95455  0.88889  1.00000  0.76190  1.00000
## Prevalence            0.13636  0.04545  0.18182  0.09091  0.27273  0.09091
## Detection Rate        0.09091  0.00000  0.09091  0.09091  0.04545  0.09091
## Detection Prevalence  0.09091  0.00000  0.18182  0.45455  0.04545  0.22727
## Balanced Accuracy     0.83333  0.50000  0.69444  0.80000  0.58333  0.92500
##                      Class: 7
## Sensitivity            0.0000
## Specificity            1.0000
## Pos Pred Value            NaN
## Neg Pred Value         0.8182
## Prevalence             0.1818
## Detection Rate         0.0000
## Detection Prevalence   0.0000
## Balanced Accuracy      0.5000
```

Based on the evidence presented, we can affirm that for this specific case, there is evidence that the **Random Forest model** is the one with the best Accuracy. Therefore, a priori, it would be the one that companies should use to be able to define which Controller to choose.