Decision Support Systems 49 (2010) 365–385

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Learning a board Balanced Scorecard to improve corporate performance

Germán Creamer [a,*], Yoav Freund [b]

[a] Stevens Institute of Technology, Castle Point on Hudson, Hoboken, NJ 07030, United States
[b] University of California, San Diego, Computer Science Department, 9500 Gilman Drive, La Jolla, CA 92093-0114, United States

ARTICLE INFO

Article history:
Received 14 April 2008
Received in revised form 14 March 2010
Accepted 4 April 2010
Available online 14 April 2010

Keywords:
Boosting
Machine learning
Corporate governance
Balanced scorecard
Planning
Performance management

ABSTRACT

The objective of this paper is to demonstrate how the boosting approach can be used to define a data-driven board Balanced Scorecard (BSC) with applications to S&P 500 companies. Using Adaboost, we can generate *alternating decision trees* (ADTs) that explain the relationship between corporate governance variables, and firm performance.
We also propose an algorithm to build a representative ADT based on cross-validation experiments. The representative ADT selects the most important indicators for the board BSC. As a final result, we propose a partially automated strategic planning system combining Adaboost with the board BSC for board-level or investment decisions.

1. Introduction

Kaplan and Norton [20] introduced the Balanced Scorecard (BSC) as a management system that helps organizations define their vision and strategy, and translate them into specific actions. The BSC provides feedback on internal business processes, performance, and market conditions in order to review the strategy and future plans [21–24,28]. Large U.S. companies, such as General Electric and Federal Express, and non-profit and public organizations have implemented the BSC approach [2,36].

The strategy of an organization, its main objectives, and its key business drivers define the indicators of the BSC. However, the choice of indicators is, in general, highly subjective and is often driven by company management or industry practices. Youngblood and Collins [39] describe a method based on indicators using multi-attribute utility theory. Clinton et al. [6] base their method on Analytic Hierarchy Process; nevertheless, these methods still require a mix of quantitative measures with a qualitative evaluation by managers or experts.

The main objective of this paper is to adapt a machine learning method, such as Adaboost, to define the core variables and the structure of the board BSC. The criterion used to design the board BSC is the firm performance. We compare the predictive capacity of Adaboost with several other algorithms such as logistic regression, and other decision trees.

The rest of the paper is organized as follows: Section 2 presents the basic concepts of a board BSC; Section 3 presents the methods used in this paper; Section 4 introduces the data and variables used in this research; Section 5 explains in detail our experiments; Section 6 presents the results of our forecast; Section 7 examines the results and the transformation of a representative ADT to a board BSC, and Section 8 presents the conclusions.

2. The Balanced Scorecard

The BSC suggests that an organization should be evaluated from four perspectives:

1. *The financial perspective* emphasizes the long-term objectives of the company in terms of revenue growth and productivity improvement. The financial objectives should be the final goals for the other perspectives.
2. *The customer perspective* emphasizes the lifetime relationship and service delivery with clients.
3. *The internal process perspective* focuses on the use of client information to sell new products and services according to their needs.
4. *The learning and growth perspective* is the foundation of the BSC. This perspective looks at the motivation, training, and capacity to innovate that employees need in order to implement new strategies.

The BSC is generally implemented at the corporate, business unit, and individual level. A missing element in these BSC implementations is the corporate governance dimension. In response to the recent

* Corresponding author. Current address: Howe School of Technology Management & School of Systems and Enterprises, Stevens Institute of Technology, Castle Point on Hudson, Hoboken, NJ 07030, United States. Tel.: +1 201 216 8986; fax: +1 201 216 5385.
E-mail addresses: gcreamer@stevens.edu (G. Creamer), yfreund@cs.ucsd.edu (Y. Freund).

0167-9236/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2010.04.004

corporate scandals in the U.S., several organizations and researchers have proposed corporate governance scorecards. Gompers et al. [14] use 24 different provisions related to takeover defense and shareholder rights to create a governance index. They show that a trading strategy based on this index outperforms the market. Standard & Poor's Governance Services [32] have developed a method which combines macro and micro variables and uses qualitative and quantitative analysis.[1] The German Society of Financial Analysts [33] and to some extent Standard & Poor's, use a qualitative framework based on "best practices" and require a lengthy due diligence process for each company under study, while the one proposed by Gompers et al. [14] is purely quantitative. Besides these corporate governance scorecards which emphasize corporate governance scoring, Kaplan and Nagel [19] proposed the creation of a board BSC that includes corporate governance variables and is oriented to strategic planning at the board level.

According to Kaplan and Nagel [19] an effective BSC program should include three parts:

1. *An enterprise BSC* that presents the company strategy, with detailed description of objectives, performance measures, targets, and initiatives to be implemented by the CEO and managers throughout the organization. The enterprise BSC also becomes a powerful tool for the directors to monitor the implementation of the corporate strategy.
2. *A board BSC* which defines its strategic contribution, includes the data necessary for the board operation, and offers an instrument to monitor the structure and performance of the board and its committees. Epstein and Roy [9,10] explain the importance of the board BSC as an instrument to monitor and implement the best-practices of corporate governance, and also as a mechanism for stakeholders to evaluate the board of directors.The enterprise BSC and the board BSC share the same financial objectives because the final role of the board and senior managers is to maximize the long-term return to shareholders. Additionally, an important element that differentiates the board BSC from the enterprise BSC is the perspective of "stakeholder" instead of "consumer". The reason to include the "stakeholder" perspective is that the stakeholders–such as shareholders, senior managers and financial analysts–are the consumers or clients of the board of directors. As a result, one of the key roles of the board is its responsibility to evaluate and motivate the senior management team.
3. *An executive BSC* allows the board of directors and the compensation committee to evaluate the performance of the top managers of the organization.

There is no theoretical support to indicate the selection and optimal combination of organizational variables such as executive compensation and insider ownership in a board BSC. Moreover, these variables may change from industry to industry and from country to country. Therefore a system that is able to recognize the optimal combination and mechanism that connects these variables, would contribute significantly to an efficient planning process.

The main hypothesis evaluated in this paper is the following: The definition of a board BSC can be partially automated[2] through a machine learning method such as boosting if this method is adapted to: a) select the most important variables, b) forecast corporate performance, c) establish the relationship among the relevant variables, d) define the minimum target (threshold) that each variable should have to contribute to a sufficient corporate performance, and e) build a board strategy map and a board BSC using the identified variables.

We evaluate this hypothesis as follows:

1. Select a group of well-known accounting and corporate governance variables that affect corporate performance.
2. Evaluate the capacity of Adaboost, logistic regression, single tree using boosting, and boosting decision stumps[3] for forecasting, variable selection, identification of relationships among variables, and definition of a minimum (maximum) threshold for each variable. We choose the algorithm with the best forecasting results and, thus, capable of selecting the most important variables, and establish how these variables interact to explain corporate performance.
3. Evaluate if the variables selected in the previous step can be converted into objectives in order to build a board strategy map.
4. Evaluate if the objectives defined in the previous step can be converted into indicators to build a board BSC.

The next section introduces the main methods that are used in this paper.

3. Methods

This section introduces two main forecasting approaches: logistic regression and boosting. In both cases, the training set consists of pairs $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_m, y_m)$ where $\mathbf{x}_i$ corresponds to the vector of features or variables of an instance i and belongs to the instance space X, and y_i is the binary label to be predicted of an instance i and belongs to the label set Y. For this paper, we assume $Y = \{0, 1\}$ because we are trying to evaluate if each firm under study is a low-middle (0) or high performance company (1). The features are the accounting and corporate governance variables described in Section 4.1 and we refer to them in a generic way as the vector $\mathbf{x}$.

3.1. Logistic regression

The logistic regression models [15] the posterior probabilities of Y using linear regression in the observed features $\mathbf{x}$. As in this paper the label set Y is binary, the model is specified in terms of the following log–odds ratio:

$$\log\frac{\Pr(Y=1|X=x_i)}{\Pr(Y=0|X=x)} = \alpha + \beta_T x.$$

3.2. Boosting

Adaboost is a general discriminative learning algorithm invented by Freund and Schapire [12]. The basic idea of Adaboost is to repeatedly apply a simple learning algorithm, called the *weak* or *base* learner,[4] to different weightings of the same training set. In its simplest form, Adaboost is intended for binary prediction problems. A *weighting* of the training examples is an assignment of a non-negative real value w_i to each example $(\mathbf{x}_i, y_i)$.

On iteration t of the boosting process, the weak learner is applied to the training sample with a set of weights $w_1^t, \ldots, w_m^t$ and produces a prediction rule h_t that maps $\mathbf{x}$ to $\{0, 1\}$. The requirement on the weak learner is for $h_t(x)$ to have a small but significant correlation with the example labels y when measured using the *current weighting of the examples*. After the rule h_t is generated, the example weights are changed so that the weak predictions $h_t(x)$ and the labels y are

[1] Even though the Standard & Poor's corporate governance scoring has been very successful in emerging markets, Standard & Poor's corporate governance services decided to pull out of the U.S. market in September 2005.

[2] For the presentation of a fully automated enterprise modeling system and its application to electronic commerce see Refs. [3,4].

[3] Boosting decision stumps algorithm is the boosting algorithm that uses decision stumps as the weak learner.

[4] Intuitively, a weak learner is an algorithm with a performance at least slightly better than random guessing.

$$
\begin{array}{l}
F_0(x) \equiv 0 \\
\text{for } t = 1 \ldots T \\
\qquad w_i^t = e^{-y_i F_{t-1}(\mathbf{x}_i)} \\
\qquad \text{Get } h_t \text{ from } \textit{weak learner} \\
\qquad \alpha_t = \frac{1}{2} \ln \left(\frac{\sum_{i:h_t(\mathbf{x}_i)=1, y_i=1} w_i^t}{\sum_{i:h_t(\mathbf{x}_i)=1, y_i=0} w_i^t} \right) \\
\qquad F_{t+1} = F_t + \alpha_t h_t
\end{array}
$$

Fig. 1. The Adaboost algorithm [12]. y_i is the binary label to be predicted, x_i corresponds to the features of an instance i, w_i^t is the weight of instance i at time t, h_t and $F_t(x)$ are the prediction rule and the prediction score at time t respectively.

decorrelated. The weak learner is then called with the new weights over the training examples, and the process repeats. Finally, all of the weak prediction rules are combined into a single *strong* rule using a weighted majority vote. One can prove that if the rules generated in the iterations are all slightly correlated with the label, then the strong rule will have a very high correlation with the label — in other words, it will predict the label very accurately.

The whole process can be seen as a variational method in which an approximation $F(x)$ is repeatedly changed by adding to it small corrections given by the weak prediction functions. In Fig. 1, we describe Adaboost in these terms. We shall refer to $F(x)$ as the *prediction score* in the rest of the document. The strong prediction rule learned by Adaboost is sign($F(x)$).

3.3. Alternating decision trees

Boosting has been widely used in practical learning algorithms such as boosting decision trees and boosting stumps [13]. Freund and Mason [11] proposed the use of boosting both to learn the decision rules and to combine these rules through a weighted majority vote in a decision tree called an *alternating decision tree* (ADT). Even though in an ADT each node can be understood as a decision rule in isolation, the main contribution of an ADT is a prediction score $F(x)$ which is the result of the combination of the individual rules. This prediction score can be transformed in several ways. The simplest version uses the sign of the prediction score to generate binary labels {0, 1} (this is the approach followed in this paper).

We explain the structure of ADTs using Fig. 2. The problem domain is corporate performance prediction, and the goal is to separate high and low performance stocks based on different features. The tree consists of alternating levels of ovals (*prediction nodes*) and rectangles (*splitter nodes*) (hence the word "alternating" in the name). The first number within the ovals defines contributions to the prediction score, and the second number (between parentheses) indicates the amount of instances that reaches each node. In this example, positive contributions are evidence of high performance, while negative contributions indicate corporate financial problems. To evaluate the prediction for a particular company we start at the top oval (0.04) and follow the arrows down. We follow *all* of the dotted arrows that emanate from prediction nodes, but we follow *only one* of the solid-line arrows emanating from a splitter node, corresponding to the answer (yes or no) to the condition stated in rectangle. We sum the values in all the prediction nodes that we reach. This sum represents the prediction score $F(x)$ above, and its sign is the final, or strong, prediction. Positive and negative values correspond to high and low performance companies respectively.

For example, suppose we had a company for which LnMarketCap = 7, KS = 0.86, RuleOfLaw = 6.01, Efficiency = 0.3, and YS = 0.28. In this case, the prediction nodes that we reach in the tree are 0.04, 0.62, −0.53, 0.69 and 1.56. Summing gives a score of 2.38, i.e., a very confident indicator that the company has a high market value.

This example demonstrates how alternating decision trees combine the contributions of many indicators to generate a prediction. The ADT in the above figure was generated by Adaboost from training data. In Adaboost's terms, each prediction node represents a weak prediction rule, and at every boosting iteration a new splitter node together with its two prediction nodes is added to

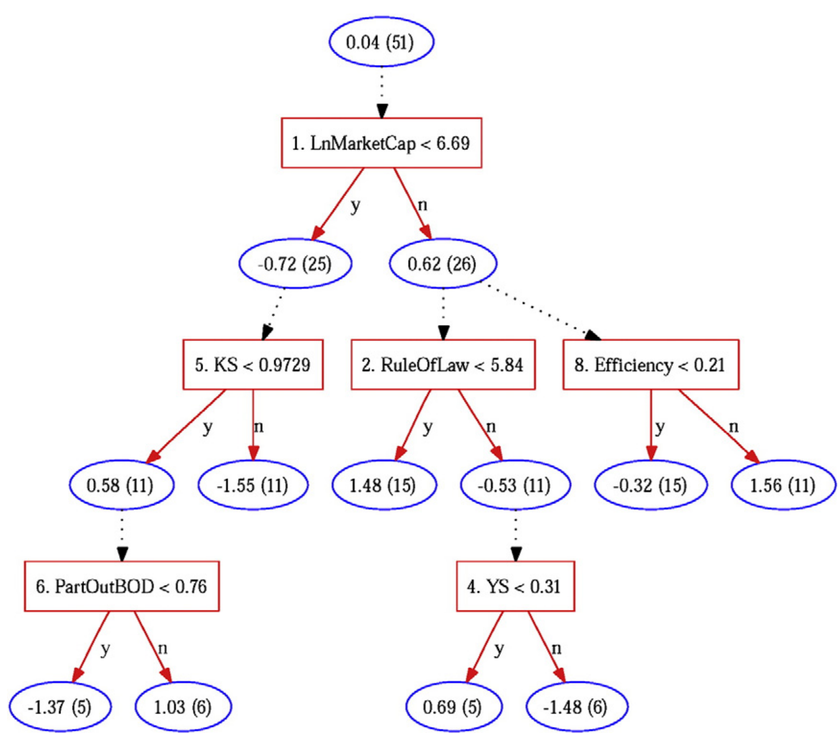

Fig. 2. Alternating decision tree (ADT) example for corporate performance. The tree has ovals (*prediction nodes*) and rectangles (*splitter nodes*). The first number within the rectangles indicates the iteration number on which the node was added. In general, lower iteration numbers indicate that the decision rule is more important. The first number within the ovals defines contributions to the prediction score, and the second number (between parentheses) indicates the number of instances that reaches each node. The root node has the initial prediction score for all cases. The sum of first values in all relevant prediction nodes, including the root, is the prediction score. The sign of prediction score is the prediction.

Please cite this article as: G. Creamer, Y. Freund, Learning a board Balanced Scorecard to improve corporate performance, Decision Support Systems (2010), doi:10.1016/j.dss.2010.04.004

the model. The splitter node can be connected to any previous prediction node, or leaf node, unless it already has a splitter node attached. The algorithm selects the new splitter node with its two prediction nodes and their position in the tree based on the minimization of the weighted error of the new rules generated. Each prediction node is associated with a weight α that contributes to the prediction score of every example reaching it. The value of α reflects the contribution of this weak rule in the overall prediction and its formula is included in Fig. 1.

3.4. The representative ADT algorithm

A common complaint about boosting and the alternating decision tree algorithm (see Section 3.2) is that the ADTs generated do not have a clear interpretation, especially if they are very different after cross-validation [15]. To answer this we propose an algorithm to calculate a representative ADT that extracts the common variables among several trees. The main motivation to propose this algorithm is to have a simpler ADT with the main nodes and relationships that are present in each of the individual ADTs.

The representative ADT algorithm is explained in Fig. 3. We introduce some notation based on Ref. [11] and used in this algorithm:

1. A base condition c is an inequality that compares a single feature or variable with a threshold, and belongs to the set of base conditions C. The threshold should be taken as a constant that separates the instances in two categories. Two base conditions with the same variable although with different thresholds are considered as two different conditions.
2. A base rule $r_{d,c}$ has a precondition d, a condition c, and two real numbers a and b. The base rule outputs score a if $d \wedge c$, b if $d \wedge \neg c$ and 0 if $\neg c$. $r_{d,c}(\mathbf{x})$ is the real value of the rule associated with $\mathbf{x}$. The thresholds of the conditions c are simplified using the first two digits of their significand. An indicator associated with the base rule $r_{d,c}$ is identified by its indices c and d.
3. The algorithm has a set of rules R, a temporary set of rules R_temp, and a vector of ranking scores S.
4. $\mathbf{T}$ is a constant predicate that is true.

The input for the representative ADT algorithm is a set of generated ADTs using the cross-validation samples. The first step of

Input:
Set of N cross-validation samples of ADTs, maximum number of nodes (V) and maximum average iterations (A).

Initialize:
Initialize multi-key maps $it_{d,c}$ and $freq_{d,c}$ that accept keys d and c associated with an initial zero value. Set R to include a base rule that has an initial precondition and condition that are both $\mathbf{T}$. Its prediction score–the root score of the representative ADT–is the average of the scores at the roots of the set of cross-validation ADTs. Initialize R_temp and empty vector S.

Selection of base rules:

1. For every ADT of the cross-validation sample, register every new base rule $r_{d,c}$ with precondition d and base condition c into the temporary set of rules R_temp; add its iteration value to $it_{d,c}$, and increase the value of $freq_{d,c}$ by one. Calculate a and b as the average of the respective values of the rules $r_{d,c}$ existent in the set of cross-validation ADTs.

2. Obtain the ranking score[a] $ranking_{d,c} = \frac{avgIter_{d,c}}{freq_{d,c}}$ for each $r_{d,c}$ and insert it into the vector S.
   $avgIter_{d,c} = \frac{it_{d,c}}{freq_{d,c}}$ and $freq_{d,c}$ are the average iteration and frequency of the base rule $r_{d,c}$ respectively.

3. Move a base rule $r_{d,c}$ from R_temp into R if the following conditions are satisfied:
   i. $ranking_{d,c}$ is in the first V elements of the vector S sorted in ascending order or $avgIter_{d,c} \leq A$.
   ii. rule $r_{d,c}$ exists in the majority of the set of cross-validation trees (at least one more than half of the existent ADTs)
   iii. scores a and b are of opposite signs
   iv. the prediction node d is the direct parent of $r_{d,c}$ and belongs to another rule r'_{c_1,c_2} present in the set R where c_1 and c_2 are its precondition and condition respectively such that d is $c_1 \wedge c_2$ or $c_1 \wedge \neg c_2$

4. Create a new rule $r_{r,c}$ where r is the root of the representative ADT, in any of the following cases:
   (a) last condition of previous item is not satisfied while the rest of conditions are satisfied, or
   (b) all conditions of previous item are satisfied by a base rule $r_{d,c}$ while exists another base rule with the same precondition d and lower ranking score $ranking_{d,c}$.

5. If the above conditions are satisfied, the representative ADT may have several base rules in the same level where their base conditions have the same variable with different thresholds.

Output:
The representative ADT based on set of rules R.
Classification rule as the sign of the sum of the predictions of the base rules in R.

[a] A low value of the ranking score shows that it is a more important base rule because in most cases this rule is included in early iterations.

Fig. 3. The representative ADT algorithm. This algorithm selects the most important variables of a group of ADTs according to an internal ranking procedure.

the representative ADT algorithm is to create a new splitter node for each base rule that is present in the majority of the ADTs (six out of ten ADTs for this research), with the same threshold based on the first two digits of its significand and with the same predecessor node. If the predecessor node is not selected, then the new splitter node is inserted in level one of the tree.

The new nodes are ranked in ascending order according to a ranking score obtained by the ratio of the average iteration to the frequency of nodes with the precondition *d* and condition *c*. A low value of this ranking score indicates that the splitter node is more important because it is present in many nodes of the original set of ADTs and/or appears in the first iterations. The algorithm selects the first V ranked splitter nodes or those that have less than an acceptable number of average iterations (*A*). This coefficient *A* is exogenously selected to avoid overfitting the ADTs of the cross-validation samples. Empirically, boosting should stop when the test error does not decrease anymore or even increase while the training error might still be falling. We chose to work with a value of sixty for *V* and *A* because after sixty iterations the reduction of the test error is very limited for our sample in the case of Adaboost (see Fig. 4). When the algorithm selects two or more splitter nodes with the same predecessor node, the splitter node with lower ranking score is put at level one of the tree. The algorithm may select a splitter node with a variable that is already in the representative ADT, although with a different threshold or predecessor.

As an example of the representative ADT algorithm (Fig. 3), we will briefly explain part of the process that generates the representative ADT for sectors 1 and 2. Fig. 5-A includes relevant sections of one of the ADTs of the cross-validation samples. $r_{r,LnMarketCap}$ is a base rule of the above figure where the root is its parent and its condition includes the variable LnMarketCap. This rule exists in at least 60% of the cross-validation ADTs and satisfies the conditions of the algorithm, so it can be included in the set *R*. The same process leads to the inclusion of the rules that are based on Efficiency and Debt ratio. The base rule $r_{d,stockDirectors}$ has as a parent a predictor node of the rule $r_{r,LnMarketCap}$ and can also be included in *R*. Similar dynamic leads to the inclusion of the rules that are based on the variables KS and TotalCompExec.

According to Fig. 5-A, the rule with the variable payDirectors is a child of the predictor node of the rule defined by IK. However this rule is not included in *R*, so the rule defined by payDirectors is sent to level 1. Similarly, the rule defined by optionsDirectors is also sent to level 1. In all the cases explored, the thresholds and predictor nodes of these rules are very similar to those observed in the sample ADT because they are the result of averaging the same scores across the different samples of ADTs. Additionally, the average iteration of all these rules, and the total number of nodes is less than sixty, so they can be included in the representative ADT (see Fig. 5-B). Finally, the set *R* contains the rules that define the representative ADT included in Fig. 3.

The representative ADT contributes to the design of the board BSC because it selects the most relevant variables, establishes the relationship among these variables, and assigns a threshold to each one. This last capability is especially important for the board BSC because the thresholds indicate the minimum value that the variables should have to contribute to a sufficient corporate performance as we explore in the next section.

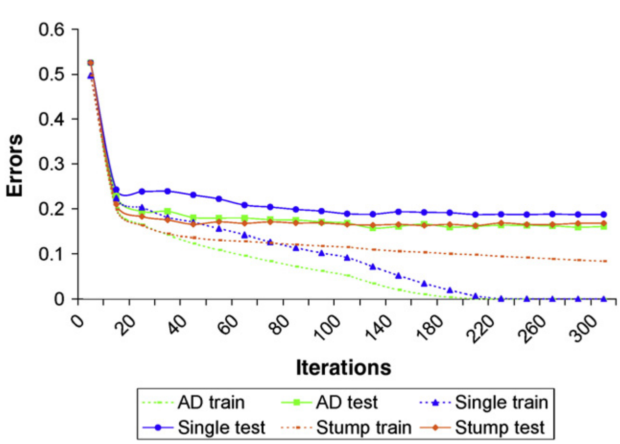

Fig. 4. Training and test error across learning algorithms for S&P 500 companies.

3.5. The Board Balanced Scorecard

The variables of the representative ADT and the relationship of these variables are used to identify key business drivers, strategic objectives and indicators of the board BSC (see Section 2). In the first step, the variables of the representative ADT are transformed into objectives of the board BSC and of the strategy map. Managers may want to evaluate if these objectives reflect strategic objectives of the company. If the representative ADT has several levels, then the relationship among nodes also determines the relationship within the objectives of the board strategy map. In the second step, the variables of the representative ADT are used as indicators, their thresholds are used as targets of the board BSC, and the objectives are imported from the board strategy map.

In the first step, the transformation from a variable of the representative ADT to an objective of the board BSC is an inductive process where the restriction imposed by the variable may indicate the satisfaction of a specific objective. For instance, the representative ADT for sectors of industrial activity 1 and 2 (Fig. 5-B) has operating expenses to sales ratio (Efficiency) as one of the most important variables according to its ranking score. This variable is an indicator of the objective to manage expenses in an adequate way. The next most important variable is market capitalization (LnMarketCap) which is an indicator of the objective of having a high level of risk management capability. In most of the cases, this transformation from variable to objective is implicit in the initial variable selection; nevertheless, the particular combination of variables may lead to different objectives. This is the case of the variable long-term assets to sales ratio (KS) which was initially selected for its capacity to detect and restrict agency conflict, and then it became an indicator of the objective that managers invest strategically considering their size and capital structure. The above variables support a high level objective of maximizing the long-term return of shareholders as presented in the board strategy map of sectors 1 and 2 (see Fig. 6).

Once there is a map between variables and objectives, the second step is a straightforward exercise where the variables become indicators and their thresholds in the representative ADT are transformed into the targets of the board BSC (Fig. 7). The sign and the importance of each indicator are established by the scores of the representative ADT that are also included in the board BSC. The summation of the scores of the selected indicators is the score of the board BSC that shows the strength of our calculations. A high positive or negative score is a strong indicator of high or low corporate performance respectively. In the above example, if a company has a long-term market capitalization (LnMarketCap) greater than its threshold of 8.3, it has a positive score of 0.217, and it is a sign of an adequate practice of risk management.

Apparently the representative Board Scorecard could be built directly from the representative ADT without creating the Strategy Map; however the Strategy Map aligns the variables selected by the representative ADT with the main corporate goals. It might be possible that certain variables selected by the representative ADT cannot be transformed into corporate goals or they are repetitive. In this case, the generation of the Strategy Map helps to select the most important variables that should be used to define the corporate goals, and therefore the Scorecard.

4. Data and variables description

Our dataset is called S&P 500 because it includes companies that are part of the S&P 500 index. The main sources of data for S&P 500 companies were ExecuComp for executive compensation information and Compustat North America for accounting information. These two datasets are products of Standard & Poor's. We restricted our dataset to S&P 500 companies with available data from 1992 to 2004. We eliminated observations that did not have enough information to calculate Tobin's Q[5] [26,34] or incomplete executive compensation information.

The main group of variables selected from the above datasets and introduced in the next subsections are variables related to corporate governance issues or to company performance. We use machine learning techniques to quantify the effect of each variable on corporate performance.

4.1. Independent variables: corporate governance factors and accounting indexes

In the experiments described in the next sections we used the following as independent variables or as features of the machine learning algorithms.

For the corporate governance variables, we include the percentage of insider ownership (T_Insider) because the separation of ownership and control is seen as an opportunity for managers to accumulate wealth at the expense of the shareholders. We also include variables related to executive compensation for the top five senior managers. The variables of executive compensation are total compensation for officers (TotalCompExec) and CEOs (totalCompCEO), value of options for officers (OptionAllValExec), CEOs (TotalValOptCEO), and directors (OptionsDirectors), value of stock options for officers (OptionAllValExec), fees paid for attendance at board of directors meetings (TotalMeetingPay), annual cash paid to each director (PayDirectors), indicator variables to specify if directors are paid additional fees for attending board committee meetings (DcommFee), and annual number of shares granted to non-employee directors (StockDirectors). The discussion about the link between executive compensation and performance is very extensive. Himmelberg et al. [17] and Palia [29] do not find any important association between Tobin's Q as a proxy for performance and equity incentives granted to managers. On the contrary, Hillegeist and Pealva [16] find that those firms with higher options incentives show better performance than the other US firms that were studied. The contradictory results of previous research as well as the importance of executive compensation in corporate governance policies led us to the study of how total, and stock options compensation for the top five officers, CEOs and directors of a broad sample of US firms affect performance.

We have selected a group of accounting variables for all companies that are well-known for their predictive power, and also are indirect indicators of corporate governance variables. These accounting variables are: the logarithm of market capitalization (LnMarketCap); long-term assets to sales ratio (KS) for their effect in the reduction of the agency conflict[6]; debt to total assets ratio (DebtRatio) as a capital structure indicator; operating expenses to sales ratio (Efficiency) as an efficiency or agency cost indicator[7]; operating income to sales ratio (YS) as a market power proxy, and an indicator of cash available from operations; and capital expenditures to long-term assets ratio (IK) as a proxy for the relationship between growth and the possibility of investing in discretionary projects. A large IK ratio may suggest agency problems if managers are developing new projects that may increase their power, but do not add market value to the company. We use region and sector as indicators of the geographical area and industrial sector in which the company operates. The Global Industry Classification Standard is used for sectors of activity. We also include Standard and Poor's index membership (SPindex) (see Fig. 8).

4.2. Dependent variables or measures of company performance

We use Tobin's Q as the measure of performance. This indicator is the preferred measure of performance in corporate governance studies such as in La-Porta et al. [25]. Tobin's Q, as a measure of the value of intangibles of a firm, is the ratio of the market value of assets to the replacement cost of assets. This is a measure of the real value created by management.[8] A higher value of Tobin's Q indicates that more value has been added or there is an expectation of greater future cash flow. Hence, the impact of management quality on performance is captured by Tobin's Q. Any difference of Tobin's Q from one indicates that the market perceives that the value of total assets is different from the value to replace their physical assets. The value of internal organization or management quality is assumed to explain the difference. We use as a proxy for Tobin's Q the ratio of book value of debt plus market value of common stocks, and preferred stocks to total assets.[9]

5. Experiments

We used Adaboost implemented with ADTs (see Section 3.2) to classify stocks with Tobin's Q below and equal or above the median. The results of ADTs must be interpreted as companies with positive scores that have high Tobin's Q, while companies with negative scores have low Tobin's Q. As independent variables we used the variables that we introduced in Section 4.1.

We performed tenfold cross-validation experiments to evaluate classification performance on held-out experiments using Adaboost. We run our experiments using 120 iterations because until this point there is a reduction of the test error in some of the learning algorithms (see Fig. 4). We used the MLJAVA package, which implements the alternating decision tree algorithm described by Freund and Mason [11].[10]

We obtained a single ADT applying the representative ADT algorithm (see Fig. 3) to the ADTs generated by Adaboost. We segmented by two main groups of economic sectors: a) 1 (energy and materials) and 2 (industrials and consumer discretionary), and b) 3 (consumer staples and health care), 4 (financials and information technology) and 5 (telecommunication services and utilities). Finally, we used the main variables and thresholds of the representative ADT as indicators and targets of the board BSC respectively (see Section 2).

We tested the performance of logistic regression (see Section 3.1), single tree using boosting, and boosting decision stumps. We eliminated variables for all the learning algorithms that indicated multicollinearity. Multicollinearity is the presence of correlation among dependent variables. In the case of logistic regression, we include indicator variables for industrial sectors. We also run individual logistic regression analysis on the examples of each of the nodes of the representative ADTs. We expect that the sign of the

[5] Tobin's Q is the ratio of the market value of assets to the replacement cost of assets.

[6] Assets can be monitored very easily and they can become collateral either for the development of new projects or to finance new acquisitions.

[7] If operating costs are too high in relation to industry peers or previous years, it might be due to excessive perquisite consumption or other direct agency costs.

[8] The intangibles can also refer to other factors such as intellectual capital or the value of information technology. The discrimination between the contribution to performance of top management and other intangibles assets requires a more detailed analysis. In this research we control for differences among economic sectors where companies may have similar technology.

[9] Several papers [5,30,31] indicate that this proxy is empirically close to the well-known Lindenberg and Ross [26] proxy.

[10] A publicly available implementation of boosting is JBoost (http://jboost.sourceforge.net). This implementation includes the MLJAVA package that we used in this research.

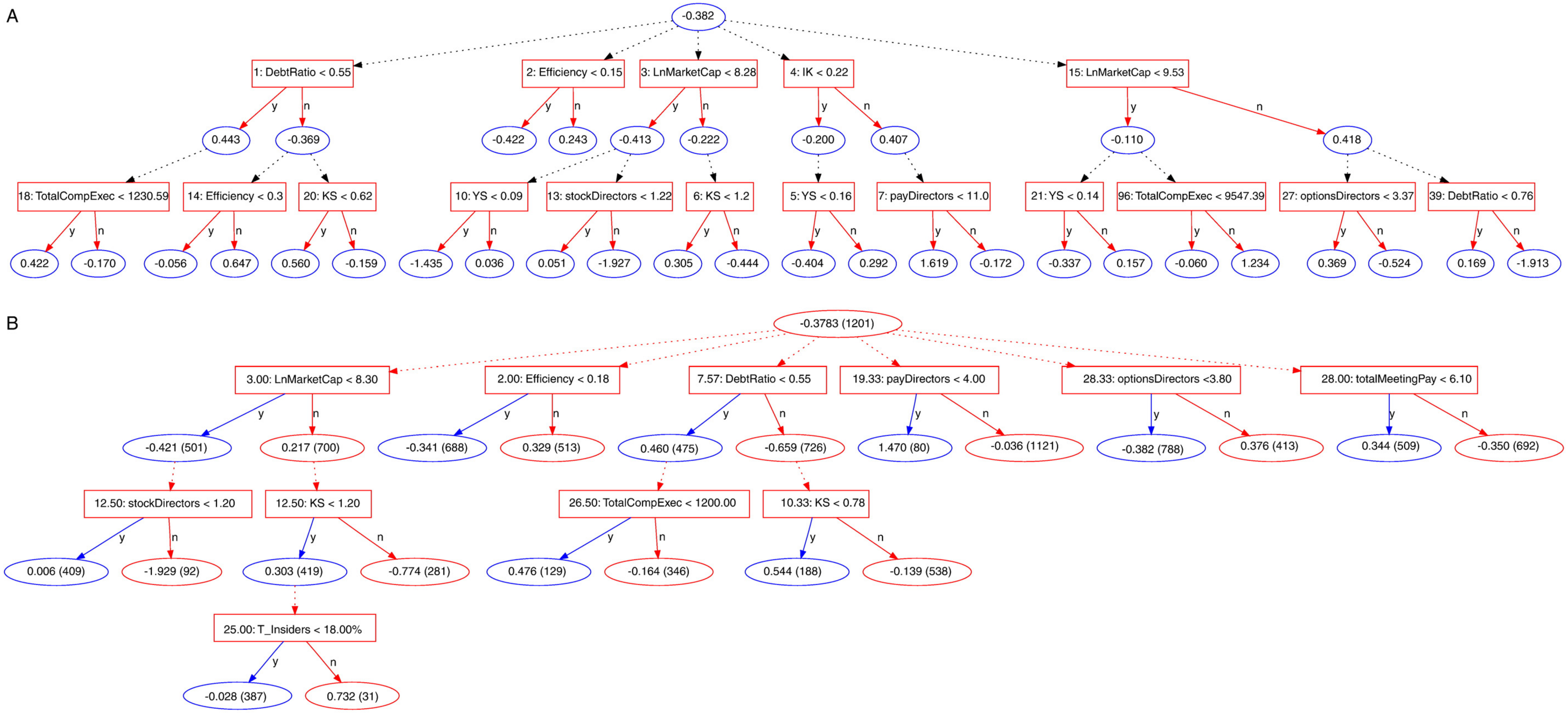

Fig. 5. S&P 500: Sample of cross-validation ADT (A) and representative ADT (B) for sectors 1 and 2. The first number in each rectangle is average iteration. The first number within the ovals defines contributions to the prediction score, and the second number (between parentheses) indicates the number of instances that reaches each node. Fig. 3 describes the procedure to calculate representative ADTs.

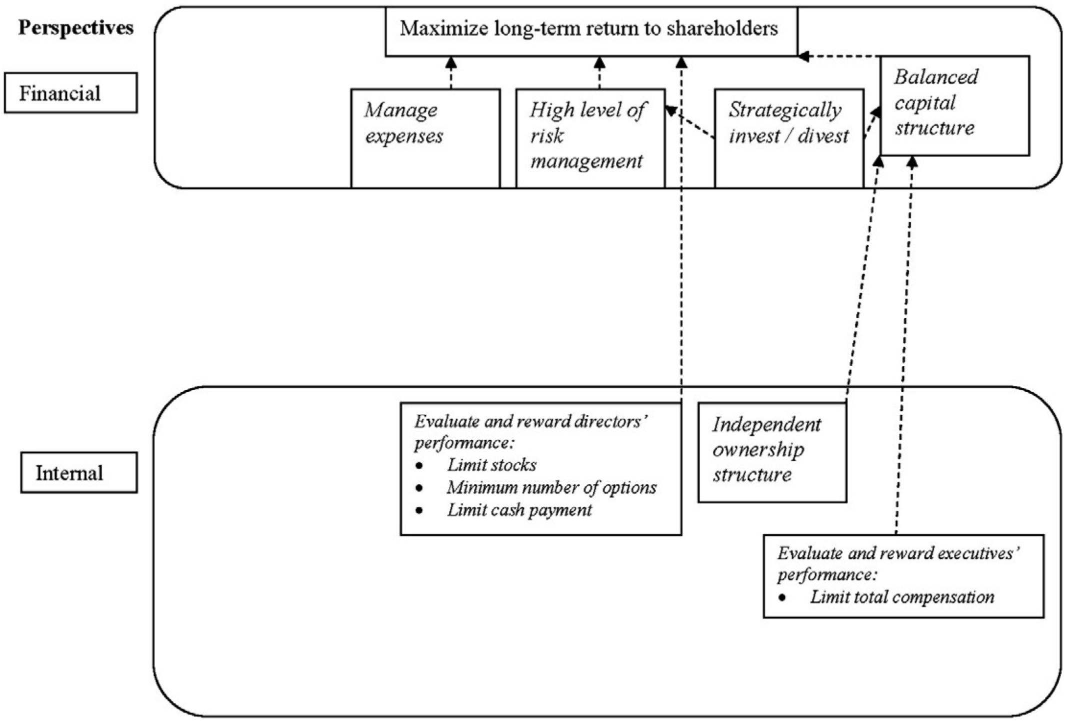

Fig. 6. S&P 500: representative Board Strategy Map for sectors 1 and 2 using only the representative ADT information. This figure shows the relationship among corporate variables in dashed lines.

Strategic objectives		Indicators	Target(s)	Scores		Owners
High-level objectives	Specific objectives			Yes	No	
Financial:						
Maximize long-term total return shareholders	Manage expenses	Operating expenses / sales (Efficiency ratio)	< 0.18	-0.341	0.329	Sr. management team
	High level of risk management	Long-term market capitalization (LnMarketCap)	< 8.3	-0.421	0.217	
	Strategically invest/divest	Long-term assets / sales (KS) (when debt ratio > 0.55)	< 0.78	0.544	-0.139	
	Strategically invest/divest	Long-term assets / sales (KS) (when LnMarketCap > 8.3)	< 1.2	0.303	-0.774	
	Balanced capital structure	Debt ratio	< 0.55	0.46	-0.659	
Internal:						
Evaluate and reward directors' performance	Limit cash payment to directors	Annual cash pay to directors (payDirectors)	< $4000	1.47	-0.036	Compensation committee
		Fees for attendance board meetings (TotalMeetingPay)	< $6100	0.344	-0.35	
	Limit stocks to directors	Number stocks granted to directors (stocksDirectors)	< 1200	0.006	-1.929	
	Minimum options to directors	Number of options granted to directors (OptionsDirectors)	< 3.8	-0.382	0.376	
Evaluate and reward executives' performance	Limit compensation top officers	Total compensation officers (TotalCompExec)	< $1200K	0.476	-0.164	
Independent ownership structure	Limit insiders' ownership	% insiders' ownership (T_Insiders)	< 18%	-0.028	0.732	Governance committee

Fig. 7. S&P 500: representative Board Scorecard for sectors 1, and 2. Adapted from Kaplan and Nagel [19]. The board scorecard assigns indicators to the objectives selected in the board strategy map. The indicators are from the representative ADT (Fig. 5-B). The targets come from the rectangle and the scores from the ovals of the representative ADTs. K is thousands and M is millions.

parameters should be consistent with the implications of the representative ADT. Finally, we conducted a χ^2 test with Yates correction[11] to evaluate the accuracy of the representative ADTs' predictions.

In this research we mainly concentrated on the financial perspective and the internal process perspective of the board BSC because of data restriction. With the information that we have available, the board BSC and the enterprise BSC are not very different. They share the same financial variables, and an important number of the executive compensation variables. The board BSC has a different group of variables related to stakeholders and to the internal operation of the board itself. However, we did not have this information. We included these variables in the board strategy map only for demonstration purposes (see Fig. 9).

[11] This conservative correction subtracts 0.5 from the absolute difference between the observed and expected value [38].

6. Results

The results of the test errors for the learning algorithms used are shown in Table 1. The test errors of Adaboost, decision stumps, and logistic regressions are not significantly different, while Adaboost is significantly better than single tree. A similar case is observed in Fig. 4 which presents the evolution of the training and test errors. The receiver operating characteristic (ROC) curves[12] (Fig. 10) also indicate that Adaboost, decision stumps, and logistic regressions are not

[12] The ROC curve shows the performance of a classifier algorithm independently of class distributions. The vertical axis is the ratio of true positives to the total number of positives, while the horizontal axis is the ratio of false positives to the total number of negatives. An approximation to the northwest corner by the ROC curve is an indication of a higher level of accuracy [37].

Indicator	Definition
TobinQ	Tobin 's Q, which is the ratio of the market value to the replacement cost of assets. We use a proxy for Tobin's Q as the ratio of book value of debt plus market value of common stocks and preferred stocks to total assets
T_Insider	% insiders' ownership.
TotalCompCEO	Total compensation for CEOs. It includes the same items as TotalCompExec (thousands of dollars).
TotalValOptCEO	Value of options for CEOs (thousands of dollars).
TotalCompExec	Total compensation for officers. It includes the following items: salary, bonus, other annual, total value of restricted stock granted, total value of stock options granted (using Black-Scholes), long-term incentive payouts, and all other total (thousands of dollars).
OptionStockValueExec	Value of stock options granted to the executive during the year as valued using S&P's Black Scholes methodology (thousands of dollars).
OptionAllValExec	The aggregate value of all options granted to the executive during the year as valued by the company (thousands of dollars).
DexecDir	Dummy variable to indicate if officer was also a director for the reference year.
OptionsDirectors	Number of options and additional options granted to each non-employee director during the year (thousands).
StockDirectors	Stock shares (including restricted stock) granted to each non-employee director (thousands).
PayDirectors	Annual cash retained paid to each director (thousands of dollars).
TotalMeetingPay	Fees paid for attendance to board of directors meeting (thousands of dollars).
DcommFee	Dummy variable to indicate if directors are paid additional fees for attending board committee meetings.
SPindex	Standard and Poor's index membership. It indicates if companies are part of S&P500 (SP), S&P midcap index (MD), S&P smallcap index (SM), or is not part of a major US index (EX).
LnMarketCap	Natural logarithm of market capitalization, used to measure firm size
KS	Ratio of long term assets (property, plant and equipment) to sales. This ratio is considered for its effect in the reduction of the agency conflict because these assets can be monitored very easily and they can become collateral for the development of new projects.
YS	The ratio of operating income to sales
DebtRatio	The ratio of debt to total assets, used as a capital structure variable. Emerging markets are much less liquid than those of developed countries. Hence, firms may give more importance to debt, rather than equity, as a source of capital.
Efficiency	The ratio of operating expenses to sales. This is the efficiency ratio and works as a proxy for market power. It also indicates cash flow available for management use. Similarly, this efficiency ratio may also reveal agency costs or agency conflicts.
IK	The ratio of capital expenditures to long term assets (stocks of property, plant and equipment)

Fig. 8. Variables used for corporate governance experiments and for a board BSC.

significantly different, although Adaboost shows a slightly higher level of accuracy in most of the cases. The ROC curve of single tree does not perform as well as the other algorithms.

Boosting decision stumps algorithm has a similar performance than Adaboost, but it puts all the variables under the root and does not consider the interaction among the variables.

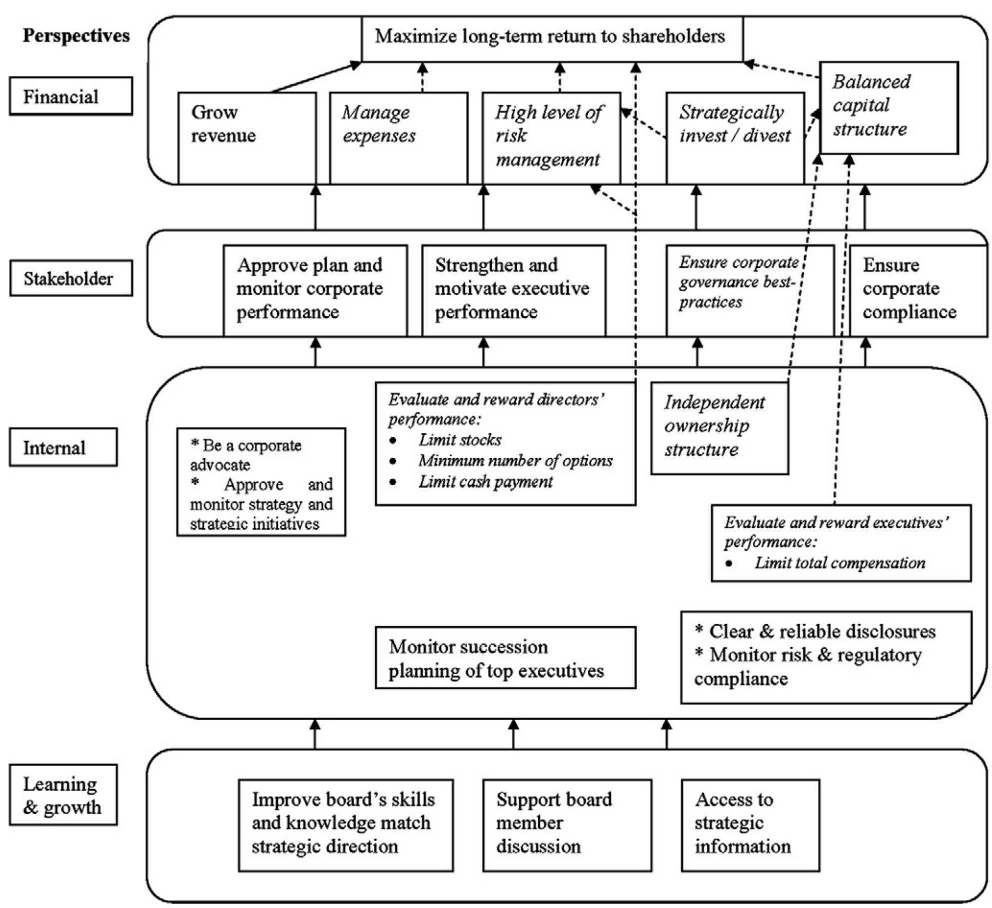

Fig. 9. S&P 500: representative Board Strategy Map for sectors 1 and 2 using the variables of representative ADT and those adapted from Kaplan and Nagel [19]. Italics are the objectives selected or modified by representative ADT and dashed lines indicate the relationships among corporate variables identified by the representative ADT.

Table 1
Test errors and standard deviations of learning algorithms. All variables are included in the calculations. * 5%, ** 1% significance level of t-test difference between test errors of algorithms and Adaboost.

	S&P 500	
	Test error	St. dev.
Adaboost	16.1%	2.0%
Single tree	20.4% **	4.2%
Stumps	16.8%	2.1%
Logistic regression	16.9%	2.9%
Number observations	2278	

Logistic regression offered some insight about the relevance of the most important variables; however it was not possible to determine appropriate thresholds for each variable. Logistic regression recognizes debt ratio and long-term assets to sales ratio as the most important variables, (see Table 2). However it assigns similar values to the corporate governance variables; hence it is not possible to recognize differences among them.

According to Adaboost, the most important variables for all the companies are operating expenses to sales ratio (efficiency ratio), operating income to sales ratio, capital expenditures to long-term assets, and long-term assets to sales ratio (see Table 2). We also segmented the companies under study by the main accounting variables when only the corporate governance variables are included (see Table 3) and we obtain similar results of those that we obtain above. The most important reduction of the test error is observed when companies have market capitalization, operating income to sales ratio, efficiency ratio, and long-term assets to sales ratio above the median. These variables were also chosen as relevant variables by the representative ADTs as we explain in the next section.

Adaboost satisfies the conditions formulated in the initial hypothesis: prediction is similar or better than other well-known algorithms such as logistic regression. Adaboost also facilitates the interpretation of the results because of the limited number of trees that were generated. For these reasons, we choose Adaboost, and its representative ADT, as the main algorithm to build a board strategy map and a board BSC.

The test errors and the χ^2 test with Yates correction of the representative ADTs are in Table 4 and their contingency tables are in Table 5. The simplification of the representative ADTs explains their reduction of accuracy in relation to the average test errors of the original ADTs, however the results of the χ^2 test are statistically significant. This indicates that the forecasts of the representative ADTs are not random.

Table 6 confirms that most of the nodes and thresholds of the representative ADTs are consistent with the underlying population. For instance, the node "LnMarketCap" in the representative ADT for sectors

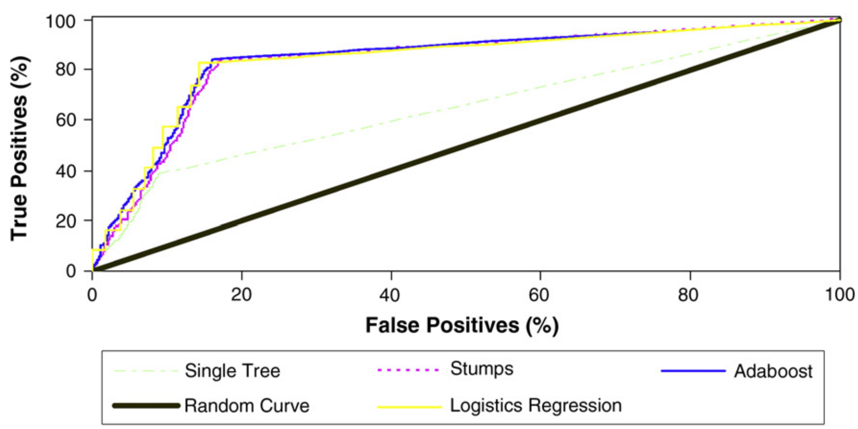

Fig. 10. ROC curve for S&P 500 companies. y-axis presents the percentage of positive observations adequately classified (true positive) and x-axis presents the percentage of negative observations misclassified (false positives).

Table 2
Results for S&P 500 companies. This table reports statistics and results of predicting Tobin's Q for S&P 500 companies using logistic regression and Adaboost. Q25: 25th percentile. Q75: 75th percentile. Logistic regression includes indicator variables to control for sector, although they are not included in the table. Corporate governance variables are in gray. Variables are ranked as an average of the iteration when each variable is selected and weighted by its frequency.

	Statistics				Logit	Boost
	Q25	Median	Q75	Mean	Odds ratios	Ranks
LnMarketCap (Nat. log market capitalization)	7.97	8.66	9.48	8.81	0.36	5
IK (Capital expenditures/long-term assets)	0.14	0.20	0.30	0.24	0.03	3
Efficiency (Operating expenses/sales)	0.12	0.22	0.34	0.24	0.00	1
YS (Operating income/sales)	0.11	0.17	0.25	0.19	0.00	2
Debtratio (Debt/total assets)	0.41	0.57	0.69	0.56	11.18	6
KS (L.T. assets/sales)	0.72	1.03	1.50	1.32	4.26	4
TobinQ (Tobin's Q: performance)	0.43	2.03	3.16	2.74	NA	
T_Insiders (% insider's equity ownership)	0.1%	0.3%	1.8%	3.27%	1.15	
totalMeetingPay (Total payment per meeting)	0	7	11	7.73	1.03	
TotalCompExec (Total compensation)	1572	2673	4782	4345.00		9
option StockValueExec (Value stock option)	466.9	1174	2687	2677.00	1.00	
payDirectors (Annual cash pay to directors)	19	26	35	27.90	1.04	8
optionAllValExec (Total value options)	543.8	1436	3088	2972.00	1.00	
optionsDirectors (Number options directors)	0	3	10	9.42	1.00	
stockDirectors (Number of stocks directors)	0.00	0.00	0.84	0.71	1.15	7
totalCompCEO (Total compensation CEO)	2012.00	4215.00	8300.00	8166.00		
totalValOptCEO (Total value options CEO)	542.30	1919.00	4741.00	5340.00	1	

Table 3
Test errors and standard deviations of Adaboost for S&P 500 companies aggregated by variables (below and equal or above the median). * 5% and ** 1% significance level of *t*-test difference with test error of subset with all observations and with only corporate governance variables.

Segments	S&P 500	
	Test error	S.D.
All variables, all sectors (2278)	16.11% **	2.02%
All variables, sectors 1 and 2 (1201)	16.24% **	4.53%
All variables, sectors 3, 4, 5 (1077)	16.53% **	2.76%
Only corporate governance variables (2278)	36.74%	3.50%
Only corporate governance variables, debt ratio < median (1139)	33.62%	1.77%
Only corporate governance variables, debt ratio ≥ median (1139)	33.74%	6.12%
Only corporate governance variables, efficiency < median (1139)	31.40%	4.88%
Only corporate governance variables, efficiency ≥ median (1139)	30.25% **	2.57%
Only corporate governance variables, cap.exp./L.T.assets < median (1139)	32.23%	4.93%
Only corporate governance variables, cap.exp./L.T.assets ≥ median (1139)	31.60%	4.31%
Only corporate governance variables, L.T. assets/sales < median (1139)	34.69%	4.38%
Only corporate governance variables, L.T. assets/sales ≥ median (1139)	30.80% *	3.80%
Only corporate governance variables, log market cap. < median (1139)	31.14% **	2.67%
Only corporate governance variables, log market cap. ≥ median (1139)	29.21% **	4.09%
Only corporate governance variables, oper.income/sales < median (1139)	29.91% *	5.50%
Only corporate governance variables, oper.income/sales ≥ median (1139)	30.00% **	5.60%

1 and 2 (Fig. 5-B) assigns a negative score (−0.421) to small companies and a positive score (0.217) to large companies. As expected, the logistic regression coefficient of this variable is positive (1.18) (see Table 6) confirming the sign of the representative ADT. For a node located in the second or higher level of the representative ADT, the logistic regression is run only with the segment defined by its parents. As an example, the node "stockDirectors" is a children of the node "LnMarketCap<8.3." Hence, the logistic regression is run only with the segment of observations characterized by "LnMarketCap<8.3." In this case, the negative sign of the parameter stockDirectors in the logistic regression is also consistent with the negative score that characterizes observations with a value of "stockDirectors<1.2" in the representative ADT. In the case of the representative ADT for sectors 1 and 2, there is only one node at the third level (T_Insiders) which has a different sign of what the logistic regression analysis suggests. Also there are only two nodes

Table 4
Test errors and χ^2 test with Yates correction of representative ADTs for S&P 500 companies aggregated by sectors. * 5% and ** 1% significance level.

Segments	Test error	χ^2 test	Number of observations
Sectors 1 and 2	24.3%	191.4262**	1201
Sectors 3, 4 and 5	25.7%	129**	1077

Table 5
Contingency tables of representative ADTs for S&P 500 companies aggregated by sectors. 0 and 1 represent observations with Tobin's Q below and equal or above the median respectively.

Actual	(a) Sectors 1 & 2		(b) Sectors 3–5	
	Forecast		Forecast	
	0	1	0	1
0	734	208	147	103
1	84	175	174	653

Table 6
Parameters of logistic regression for S&P 500 companies. The dependent variable is a binary variable that classifies stocks with Tobin's Q below and equal or above the median. The first group of columns corresponds to sectors 1 and 2, and the second group corresponds to sectors 3, 4, and 5. Each column selects a group of stocks that is consistent with the levels and nodes of the representative ADTs. The parameters that are relevant for each node are highlighted. Corporate governance variables are in gray. . 10%, * 5%, ** 1%, and *** 0.1% represent the significance level of the parameters.

Level in the representatve ADT	Sectors 1 and 2						Sectors 3, 4, and 5				
	1	2	2	2	2	3	1	2	2	3	3
	All variables	LnMarketCap < 8.3	LnMarketCap > = 8.3	Debt Ratio < 0.55	Debt Ratio > = 0	LnMarketCap > = 8.3 & KS < 1.2	All variables	YS < 0.15	Sector = 5	YS < 0.15 & stockDir. < 0.77	YS < 0.15 & stockDir. > 0.77
Sector = 5 (dummy variable)							−2.99***	−8.31		−10.13	
LnMarketCap (Nat. log market capitalization)	1.18***	1.72***	0.92***	1.52***	1.26***	1.00***	1.12***	1.13***	1.07*	1.23***	
IK (Capital expenditures/long-term asssets)	4.22***	3.34*	6.65***	4.72**	17.94***	8.54***	3.89***	5.84***		7.49***	
Efficiency (Operating expenses/sales)	5.00***	8.75***	−4.45***	3.06*	1.68	4.97*	5.93***	5.44***		5.45***	
YS (Operating income/sales)	8.53***	14.32***	3.77**	2.90*	9.14**	40.61***	10.63***	4.82		5.73*	
Debtratio (Debt/total assets)	−3.42***	−0.97	6.19***	−6.24***	5.05**	−1.90	−1.30*	−1.70		−1.03	
KS (L.T. assets/sales)	−1.35***	−2.28***	−1.24***	−0.57**	−3.16***	−6.89***	−1.47***	−1.80***		−1.52***	
T_Insiders (% insider's equity ownership)	−0.10	1.07	−0.05*	−1.73	0.00	−4.94*	1.37	0.61		0.28	
TotalMeetingPay (Total payment per meeting)	−0.02	0.05	−0.13	0.00	−0.20	−0.06*	−0.03	0.01		0.02	
TotalCompExec (Total compensation)	0.00*	0.00***	1.99**	−0.001***	2.15*	0.00	0.00***	0.00		0.00	
OptionStockValueExec (Value stock option)	0.00	0.00***	0.00	0.00**	0.00	0.00*	0.00***	0.00		0.00	
PayDirectors (Annual cash pay to directors)	−0.05***	−0.10***	0.00	−0.03***	0.00	−0.04***	−0.02**	0.00		−0.01	
OptionAllValExec (Total value options)	0.00	0.00	−0.04***	0.00	−0.06***	0.00	0.00	0.00		0.00	
OptionsDirectors (Number options directors)	0.00	−0.01	0.00	−0.01	0.00	0.04	0.00	0.01		0.01	0.04
StockDirectors (Numbers of stocks directors)	−0.27***	−1.10*	0.01	−0.05	0.00	0.04	−0.02	0.01		0.13	
TotalCompCEO (Total compensation CEO)	0.00	0.00*	−0.18	0.00*	−0.47*	0.00	0.00**	0.00		0.00	
TotalValOptCEO (Total value options CEO)	0.00	0.00*	−1.84	0.00	2.00	0.00	0.00***	0.00		0.00	
DcommFee (indicator if fees paid to directors)	−0.39	−1.36**	0.00	−1.06**	0.00	0.59	0.27	0.43		0.20	
DexecDir (indicator if officer Is also a director)	1.83***	2.54*	0.00	0.37	0.00	−0.41	0.08	1.62*		1.43	

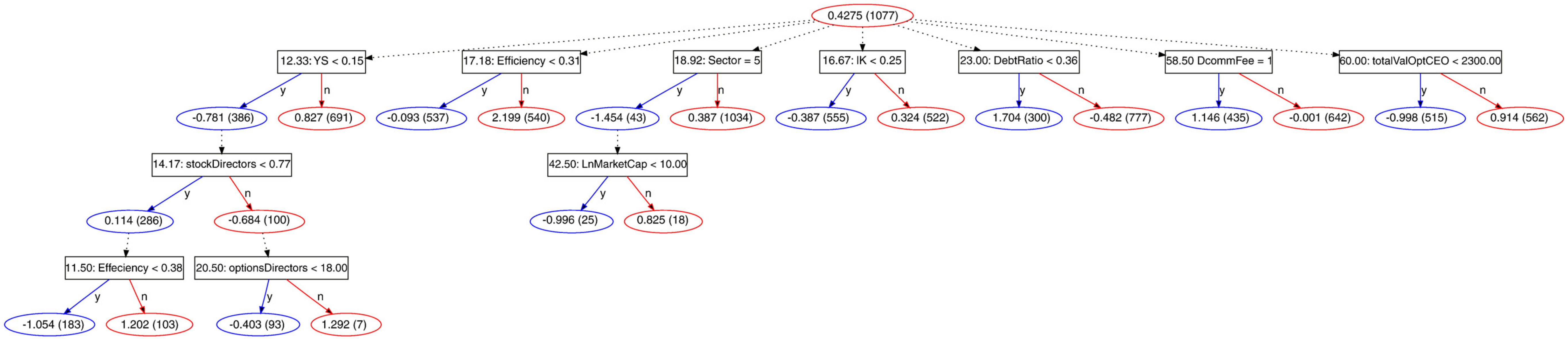

Fig. 11. S&P 500: representative ADTs for sectors 3, 4 and 5. The first number in each rectangle is average iteration. The first number within the ovals defines contributions to the prediction score, and the second number (between parentheses) indicates the number of instances that reaches each node. Fig. 3 describes the procedure to calculate representative ADTs.

(optionsDirectors and TotalMeetingPay) that are not significant according to the logistic regression analysis. The representative ADT for sectors 3, 4 and 5 has one node (stockDirectors) that is not significant, and another node (totalValOptCEO) that cannot be confirmed by the regression analysis because its parameter is near zero. At a high level, representative ADTs are able to select the most relevant variables, establish the relationship among them, and define a threshold that each variable should have to improve corporate performance.

7. Discussion

In this section we briefly discuss the variables that are most relevant or are part of a structure of several levels of the representative ADTs.

We present the representative ADT segmented by sectors 1 and 2 (Fig. 5-B), and sectors 3, 4, and 5 (Fig. 11-B). In the case of sectors 1 and 2, companies that have a maximum debt to total assets ratio of 0.55 show a superior performance. Additionally, among these high performance companies, those that show a better performance have a limit of USD 1.2 million as the total compensation for top officers. In the case of a company that is highly leveraged, performance may improve if the long-term assets to sales ratio (KS) is less than 0.78. In this segment of companies, size matters. Large companies have an advantage as long as the long-term assets to sales ratio is less than 1.2, and there is at least 18% of insiders in the board of directors. During recessions, industrial companies with large long-term assets to sales ratio may have difficulties to quickly adapt to a significant reduction in demand. Smaller companies show an improvement in their performance when the number of stocks granted to directors is restricted to 1200. Additionally, reducing the fixed compensation for directors and increasing their variable compensation through options have a positive effect in the company.

The importance of size, debt ratio, and also a minimum efficiency ratio of 0.18 in the latter companies is explained because sectors 1 (energy and materials) and 2 (industrials and consumer discretionary)[13] are mostly of an industrial type or are capital intensive, and they may have higher fixed investments than the rest of the industries. The most important aspect for companies of these sectors is to maintain a minimum level of operating expenses to cover an efficient level of operation and a healthy capital structure where about half of its assets are funded by debt. A larger proportion of debt to assets may make these companies very vulnerable during periods of downturn when liquidity and credit become more restricted, and they may not have the flexibility for easy liquidation of assets as in the case of financial institutions.

The representative ADT for sectors 3, 4, and 5 indicates that performance improves in those companies with an operating income to sales ratio larger than 0.15. If this is not the case, restricting the number of stocks or increasing the number of options granted to directors improves performance. However, the minimum number of options to directors is higher than the one observed in sectors 1 and 2. Additionally, the representative ADT of companies of sectors 3, 4, and 5 establishes a minimum for total remuneration of the CEO which does not exist in the case of companies of sector 1 and 2. These differences might be explained by the main business processes of each economic sector. Companies of sectors 1 and 2 take major investment decisions that involve direct participation of the senior managers as well as the strategic direction of the board of directors, such as the development of a new factory or the exploration of a new oil region. Companies may encourage the involvement of CEOs and directors with large compensations based on options. However, once these investment decisions are taken the role of middle managers becomes more relevant, and CEO compensation is less important.[14] The profit of companies of sectors 3, 4 and 5, especially in the case of financial services, information technology and telecommunications, are driven by the quality of customer relationship, continuous technology update, investment and risk management decisions. Therefore, their success may significantly depend on the motivation of the CEO and directors through flexible remuneration (options) to take strategic decisions that maintain the leadership of the company or the implementation of sound risk and investment management practices.

If the above rules are effective to improve company performance, the flexible part of executives' compensation might be reduced in some sectors, however it should not disappear as a result of the recent FASB rule which establishes that companies should register as expense any options granted to employees.

7.1. From the representative ADT to the board Balanced Scorecard

We built board strategy maps using the representative ADTs for two main groups of sectors: a) 1 and 2, and b) 3, 4, and 5. We converted the variables of the representative ADT into objectives, and included the relationship among the variables of the representative ADT in the board strategy map. In the case of sectors 1 and 2, the variables of the representative ADT (Fig. 5-B) are converted into objectives of the board strategy map (see Fig. 6). For instance, the variable debt ratio (debtRatio) is converted into "Balanced capital structure". We expect that the threshold of the variable debt ratio (0.55) will indicate a maximum debt ratio or combination of debt and equity that high performance companies should have. Additionally, the variable long-term assets to sales ratio (KS) indicates the maximum recommended proportion of assets in relation to sales. However, this index is relevant only in the case of those companies that have a maximum debt ratio of 0.55 or a minimum logarithm of market capitalization of 8.3. For this reason we establish in Fig. 6 a relationship between "Strategically invest/divest" and "Balanced capital structure" and "High level of risk management." We follow the same process with the rest of the variables of the representative ADT and as a result obtain a simplified board strategy map for sectors 1 and 2 (Fig. 6). This strategy map contains only information about the financial and the internal perspectives because this is the information that we have available. We extend this strategy map into a modified version of the board strategy map (Fig. 9) as proposed by Kaplan and Nagel [19]. This new version of the board strategy map includes the additional "learning and growth" and "stakeholder" perspectives. The variables and relationships derived from the representative ADT are in italics and in dashed lines respectively.

As we demonstrated for the case of sectors 1 and 2, we have expanded the board strategy map proposed by Kaplan and Nagel [19] to incorporate new objectives that were consistent with the main variables selected by the representative ADT. The new objectives that emerged are "Balanced capital structure" in the financial perspective; "Ensure corporate governance best-practices" in the stakeholder perspective; "Independent ownership structure", "Evaluate and reward directors' performance" and "Evaluate and reward executives' performance" in the internal perspective.

We translate the new board strategy map into a board BSC for sectors 1 and 2 (Fig. 7) that includes high-level objectives, specific

[13] Energy includes energy equipment and services. Materials include chemical industries, construction materials, containers and packaging, metals and mining, and paper and forest products. Industrials include capital goods; commercial services and supplies; and transportation. Consumer discretionary includes automobiles and components; consumer durables and apparel; hotels, restaurants and leisure; media, and retailing.

[14] Jensen and Murphy [18] consider that there is a major misalignment between corporate performance and compensation paid to executives, especially CEOs. In recent years, there are well-known stories of CEOs who have been paid large compensations regardless of their performance. Michael Ovitz, former president of The Walt Disney Corp., received $140 million as his severance package when he was fired by unhappy shareholders after 14 months at the company. Core et al. [7] also find that CEOs have greater compensation in companies with greater agency problems.

Strategic objectives		Indicators	Target(s)	Scores		Owners
High-level objectives	Specific objectives			Yes	No	
Financial:						
Maximize long-term total return shareholders	Grow revenues	Operating income / sales (YS)	< 0.15	-0.781	0.827	Sr. management team
	Manage expenses	Operating expenses / sales (Efficiency ratio)	< 0.38	-1.054	1.202	
	High level of risk management	Long-term market capitalization (LnMarketCap) (for sector 5)	< 10	-0.996	0.825	
	Strategically invest/divest	Capital expenditures / Long-term assets (IK)	< 0.25	-0.387	0.324	
	Balanced capital structure	Debt ratio	< 0.36	1.704	-0.482	
Internal:						
Evaluate and reward directors' performance	Compensate participation of directors in committees	Indicator of directors' fees for attending board committee meeting (Dcommfee, Y=1, N=0)	0	0.348	0.512	Compensation committee
	Limit stocks to directors	Number stocks granted to directors (StockDirectors)	< 770	0.114	-0.684	
	Minimum number of options to directors	Number of options granted to directors (OptionsDirectors)	< 18000	-0.403	1.292	
Evaluate and reward executives' performance	Minimum number of options to CEO	Total value options CEO's (totalValOptCEO)	< $2300K	-0.998	0.914	
Presence of insiders' directors	Include insiders directors	Indicator if officer is also a director (DexecDir, Y=1, N=0)	1	-1.891	0.177	Governance committee

Fig. 12. S&P 500: representative Board Scorecard for sectors 3, 4, and 5. Adapted from Kaplan and Nagel [19]. The board scorecard assigns indicators to the objectives selected in the board strategy map. The indicators are from the representative ADT (Fig. 11). The targets come from the rectangle and the scores from the ovals of the representative ADTs. K is thousands and M is millions.

objectives, indicators, targets, scores, and owners or those responsible for each objective. The objectives are obtained from the board strategy map. The indicators are the most important variables selected by the representative ADTs and their targets and scores are the threshold levels and scores calculated for each variable respectively. For instance, in the above example if a company has a debt ratio below 0.55 it has a positive score of 0.46, otherwise the score is -0.659. A high positive or negative summation of all the partial scores is a strong indicator of high or low corporate performance respectively. We do not include the board strategy map for sectors 3, 4, and 5 because it is just slightly different from the one that we presented for sectors 1 and 2. We only include the board BSC for sectors 3, 4, and 5 (Fig. 12) which has its own parameters based on its representative ADT (Fig. 11).

Finally, we can say that the representative ADT and board BSC complement each other. The representative ADT selects what are the most important variables that should be used as indicators and therefore helps to choose the key drivers and objectives of the board BSC. Additionally, the representative ADT is able to calculate the targets for every metric. The board BSC puts in perspective the findings of the representative ADT. The board BSC, as a strategic management system, integrates the four perspectives already described, and offers a framework that connects the variables recognized by the representative ADT in a logical order towards the maximization of shareholders' return.

8. Final comments and conclusions

Boosting has been applied to different business areas such as direct marketing [27], financial forecasting [1], and electronic commerce [35]. In these areas, the analysis may substantially improve if boosting is used not only for prediction but also for interpretation as this paper has demonstrated.

Adaboost, combined with the representative ADT, is an algorithm that can partially automate the definition of a board BSC as we proposed in our initial hypothesis. This algorithm is able to forecast corporate performance, select the most important variables, establish relationships among these variables, define a target for each variable to optimize corporate performance, and build a board strategy map and a board BSC. With this tool, managers can concentrate on the most important strategic issues and delegate the calculation of the targets to a semi-automated planning system supported by Adaboost.

The use of ADTs in finance requires time-series or cross-sectional data in order to calculate meaningful nodes. Indicators that do not have enough information cannot be quantified using ADTs, so, the initial versions of a board BSC still require an important participation of the board of directors, middle and senior management. However, as the planning team or the company creates its own database, then the representative ADT can select the relevant indicators and their targets. As Creamer and Freund [8] showed, Adaboost also worked adequately with small datasets. However, the variance of the test error increased as the size of the dataset decreased. We suggest that companies that use Adaboost to build board BSCs use large datasets (industrial surveys or compensation surveys) or build their own internal dataset using the company's historical information.

Finally, this paper can be enriched by the modification of the representative ADT algorithm to reduce the loss of accuracy in relation to the average test errors of the original ADTs. This difference is related to the variance of the test errors of the ADTs that also depends of the size of the dataset used. Future research can be directed to optimize the rules that define when and where a node is included in the representative ADT. For instance, we could combine several splitter nodes with the same predecessor at the same level. We could also have a more complex representative ADT that might be different from a single ADT, although may simulate more closely the diverse sample of ADTs. When there is a small training dataset, then generating an ADT with a small number of decision nodes is useful both in terms of interpretability and in terms of performance on the test set. However, when there is a large training dataset it is necessary to make a choice: either performance or interpretability. A small ADT will be easy to interpret and to translate into a BSC with an emphasis on understanding how the different variables interact as we have done in this paper. A large ADT will be more appropriate if the scorecard is used for very precise calculations such as those required in risk management or forecasting.

Acknowledgements

Authors thank the editor Andrew Whinston, two anonymous referees, David Waltz, Tony Jebara, Sal Stolfo, Vasant Dhar, John Moody, and participants of the Machine Learning in Finance Workshop of NIPS 2005, and of the Eastern Economics Association meetings, 2007 for their valuable comments, and to Patrick Jardine for proof-reading the article. The opinions presented are the exclusive responsibility of the authors.

References

[1] E. Alfaro, N. Garca, M. Gmez, D. Elizondo, Bankruptcy forecasting: an empirical comparison of adaboost and neural networks, Decision Support Systems 45 (2008) 110–122.
[2] M. Ampuero, J. Goransson, J. Scott, Solving the measurement puzzle. how EVA and the Balanced Scorecard fit together, Perspectives on Business Innovation 2 (1998) 45–52.
[3] S. Ba, R. Kalakota, A.B. Whinston, Using client-broker-server architecture for intranet decision support, Decision Support Systems 19 (1997) 171–192.

[4] S. Ba, K.R. Lang, A.B. Whinston, Enterprise decision support using intranet technology, Decision Support Systems 20 (1997) 99–134.
[5] K. Chung, S.W. Pruitt, A simple approximation to Tobin's Q, Financial Management 23 (1994) 70–74.
[6] D. Clinton, S. Webber, J. Hassell, Implementing the Balanced Scorecard using the analytic hierarchic process, Management Accounting Quarterly 3 (2002) 1–11.
[7] J.E. Core, R.W. Holthausen, D.F. Larcker, Corporate governance, chief executive officer compensation, and firm performance, Journal of Financial Economics 51 (1999) 371–406.
[8] G. Creamer, Y. Freund, Using Boosting for Financial Analysis and Performance Prediction: Application to S&P 500 Companies, Latin American ADRs and Banks, Computational Economics (2010), doi:10.1007/s10614-010-9205-3.
[9] M.J. Epstein, M.J. Roy, Measuring and improving the performance of corporate boards, Technical Report, The Society of Management Accountants of Canada, 2002.
[10] M.J. Epstein, M.J. Roy, How does your board rate? Strategic Finance (February 2004) 25–31.
[11] Y. Freund, L. Mason, The alternating decision tree learning algorithm, Machine Learning: Proceedings of the Sixteenth International Conference, Morgan Kaufmann Publishers Inc., San Francisco, 1999, pp. 124–133.
[12] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences 55 (1997) 119–139.
[13] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting, The Annals of Statistics 38 (2000) 337–374.
[14] P.A. Gompers, J.L. Ishii, A. Metrick, Corporate governance and equity prices, Quarterly Journal of Economics 118 (2003) 107–155.
[15] T. Hastie, R. Tibishirani, J. Friedman, The Elements of Statistical Learning, Springer, New York, 2003.
[16] S.A. Hillegeist, F. Pealva, Stock option incentives and firm performance, IESE Research Papers D/535, IESE Business SchoolAvailable at, http://ideas.repec.org/p/ebg/iesewp/d-0535.html2004.
[17] C. Himmelberg, G. Hubbard, D. Palia, Understanding the determinants of managerial ownership and the link between ownership and performance, Journal of Financial Economics 53 (1999) 353–384.
[18] M. Jensen, K.J. Murphy, Remuneration: where we've been, how we got to here, what are the problems, and how to fix them, Finance Working Paper 44/2004, European Corporate Governance Institute, 2004.
[19] R.S. Kaplan, M.E. Nagel, Improving corporate governance with the Balanced Scorecard, Directors Monthly, March 2004, pp. 16–20.
[20] R.S. Kaplan, D.P. Norton, The Balanced Scorecard: measures that drive performance, Harvard Business Review 70 (1992) 72–79.
[21] R.S. Kaplan, D.P. Norton, Putting the Balanced Scorecard to work, Harvard Business Review 71 (1993).
[22] R.S. Kaplan, D.P. Norton, Linking the Balanced Scorecard to strategy, California Management Review 49 (1996) 53–79.
[23] R.S. Kaplan, D.P. Norton (Eds.), The Balanced Scorecard, Harvard Business School Press, Boston, 1996.
[24] R.S. Kaplan, D.P. Norton, Using the Balanced Scorecard as a strategic management system, Harvard Business Review 74 (1996) 75–85.
[25] R. La-Porta, F.L. de Silanes, A. Shleifer, R. Vishny, Investor protection and corporate valuation, Journal of Finance 57 (2002) 1147–1170.
[26] E. Lindenberg, S.A. Ross, Tobin's Q ratio and industrial organization, Journal of Business 54 (1981) 1–32.
[27] C.X. Ling, C. Li, Data mining for direct marketing: problems and solutions, KDD '98: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, AAAI Press, New York City, 1998, pp. 73–79.
[28] M. Martinsons, R. Davison, D. Tse, The balanced scorecard: a foundation for the strategic management of information systems, Decision Support Systems 25 (1999) 71–88.
[29] D. Palia, The endogeneity of managerial compensation in firm valuation: a solution, Review of Financial Studies 14 (2001) 735–764.
[30] S.B. Perfect, K.W. Wiles, Alternative construction of Tobin's Q: an empirical comparison, Journal of Empirical Finance 1 (1994) 313–341.
[31] P. Peterson, D. Peterson, Company performance and measures of value added, The Research Foundation of the Institute of Chartered Financial Analysts, Charlottesville, VA, 1996.
[32] Standard & Poor's Governance Services, Standard & poor's corporate governance scores and evaluations, , 2004.
[33] C. Stenger, The corporate governance scorecard, Corporate Governance: An International Review 12 (2004) 11–15.
[34] J. Tobin, A general equilibrium approach to monetary theory, Journal of Money, Credit, and Banking 1 (1) (1969) 15-29.
[35] D. van Heijst, R. Potharst, M. van Wezel, A support system for predicting ebay end prices, Decision Support Systems 44 (2008) 970–982.
[36] S. Voelpel, M. Leibold, R. Eckhoff, T. Davenport, The tyranny of the balanced scorecard in the innovation economy, Journal of Intellectual Capital 7 (2006) 43–60.
[37] I.H. Witten, E. Frank, Data Mining: practical machine learning tools and techniques, 2nd. editionMorgan Kaufmann, San Francisco, 2005.
[38] F. Yates, Contingency table involving small numbers and the χ^2 test, Journal of the Royal Statistical Society 1 (1934) 217–235.
[39] A. Youngblood, T. Collins, Addressing Balanced Scorecard trade-off issues between performance metrics using multi-attribute utility theory, Engineering Management Journal 15 (2003) 11–17.

Yoav Freund is a professor of Computer Science and Engineering at UC San Diego. His work is in the area of machine learning, computational statistics and their applications in bio-informatics, computer vision, network routing and high-performance computing. Dr. Freund is an internationally known researcher in the field of machine learning — a field which bridges computer science and statistics. He is best known for his joint work with Dr. Robert Schapire on the Adaboost algorithm. For this work they were awarded the 2003 Gödel prize in Theoretical Computer Science and the Kanellakis Prize in 2004. Listed in the Thomson ISI directory of highly cited researchers.

Germán Creamer is an Associate Professor of quantitative finance and financial engineering in the School of Systems and Enterprises and in the Howe School of Technology Management at Stevens Institute of Technology. Dr. Creamer has been a senior manager in the Risk, Information and Banking Division in American Express where he worked in the enterprise-wide risk management and the information management groups. He has taught at Columbia University, Tulane University, and in several leading Latin American business schools. Dr. Creamer has a PhD in Computer Science (specialized on computational finance), a MSc in Financial Engineering, both from Columbia University, and a PhD in Economics from the University of Notre Dame. His articles have appeared in leading journals including *Quantitative Finance, Data Mining and Knowledge Discovery, and World Trade Review*. Professor Creamer is also a CFA charter holder. His current area of research is on risk management, trading systems, financial decision support systems and data mining/machine learning applications to finance.