New Metrics and Approaches in Bankruptcy Prediction ☆

Flavio Barboza[a], Leonardo Fernando Cruz Basso[b], Herbert Kimura[c]

[a] *Federal University of Uberlândia, Uberlândia-MG, 38408–100, Brazil. E-mail: flmbarboza@ufu.br*
[b] *Mackenzie Presbyterian Univ., São Paulo-SP, 01302–907, Brazil. E-mail: leonardofernando.basso@mackenzie.br*
[c] *University of Brasília, Darcy Ribeiro Campus, Brasília-FD, 70910–900, Brazil. E-mail: herbert.kimura@gmail.com*

Abstract

Credit risk models, particularly those that seek to understand what signals are given by companies on the verge of bankruptcy, are under constant discussion and continually updated. Statistical models, for example, Discriminant Analysis and Logistic Regression, are traditional and easy to understand. However, new non-statistical techniques have been recently tested in the financial context, such as the case of machine learning mechanisms. In this paper, we developed bankruptcy forecasting models for non-financial companies, using a data set that covers the period from 1980 to 2014. We examined static variables, growth variables and also growth variation variables in order to discriminate two groups: firms that did not go bankrupt and firms that went bankrupt within one (fiscal) year after analysis of their data. To study the discretionary capacity, we applied seven techniques, among statistics and machine learning. In view of the number of models evaluated, the analyses are rich and varied, highlighting the finding that traditional techniques with the appropriate variables are able to obtain performance better than machine-learning models, opening a debate with the current literature, which is has been stating just the opposite by highlighting the importance of computational techniques. This article contributes to the literature in many ways; it uses a comprehensive sample and a comparison of performance for the modeling, enabling a discerning view for both risk managers and for researchers in this field. Additionally, it is the first study to test not only growth measures (assets, sales, etc.) but also growth variation rates, which have relevant bases and impacts.

Key words: credit risk, bankruptcy prediction, machine learning, discriminant analysis, logistic regression.

JEL codes: C45; C52; C63; G33; L25.

☆This document is a collaborative effort. The authors are thankful to financial support from the Federal University of Uberlândia (UFU-PROPP), Mackenzie Presbyterian University (MackPesquisa) and CNPq. Additional acknowledgments to Credit & Debt Markets Research Program from Salomon Center, New York University; Brenda Kuehne (Credit & Debt Markets Research Specialist) for the list of bankrupt firms; and, participants at seminars in Uberlândia, São Paulo as well as the discussants in the 2017 Global Finance Conference for their valuable comments on earlier versions of this paper. The views expressed here are those of the authors and not necessarily those of our institutions. All remaining errors are solely authors' responsibility.

1. Introduction

In crisis times or in good times, bankruptcies happen. Both researchers and market practitioners (investors, managers, auditors, and others) have an interest in understanding the possibility of companies' not meeting their financial obligations. This interest ultimately leads to the development of new models for estimating potential bankruptcy, seeking to establish appropriate mechanisms to manage an elementary part of the credit risk.

Throughout this and the last decade, maintaining financial stability and preventing credit risk were shown to be highly relevant issues (Yim and Mitchell, 2009; Erdogan, 2016). The financial crisis of 2008/2009 appears as a red flag to the inconsistencies relating to credit analysis, serving as an important example of the need for well-structured risk management, ranging from credit policy to the granting of credit based on a consistent and appropriate modeling.

Ever since the seminal articles by Beaver (1966) and Altman (1968), there are models focused exclusively on predicting bankruptcy (Sandin and Porporato, 2008). While Beaver (1966) worked with univariate analyses to recognize the variables with the greatest impact on identifying bankruptcy events based on a list of 30 different metrics, Altman (1968) found that out of the total of 22 financial indices – including liquidity, profitability, leverage, solvency and activity – there was a linear combination of these measures capable of differentiating between bankrupt and non-bankrupt firms, in the industrial sector only, with an accuracy of approximately 93%. After applying the multivariate discriminant analysis (MDA) to the data set, five financial indices were selected to build the model, called Z-Score, sufficient to achieve a 90% accuracy level.

Furthering the study by Altman (1968), Ohlson (1980) found that the use of MDA could have been enhanced, since the variance of the groups of those with good credit history and those with bad credit history were uneven, and that the violation of assumptions of distribution normality could compromising the accuracy of predictions. Accordingly, Ohlson presents Logistic Regression (LR), one of the traditional techniques used to measure credit risk (Virág and Kristóf, 2005).

Based on these studies, several studies were developed, discussing the most appropriate variables to be incorporated into predictive models. For example, Ong et al. (2011) and Li (2012), following the same idea of the seminal articles, used usual and static accounting indicators, such as indebtedness; Agarwal and Taffler (2008), and Lyandres and Zhdanov (2013) included more market variables; and du Jardin (2010) added "change in other debts" and "change in equity position" as predictive variables, but none of these studies advanced the scope by focusing on growth measures, one of the proposals of this paper.

Some studies tested new methodologies in databases in different regions, for example, Altman et al. (1994) in Italy, Agarwal and Taffler (2008) in the UK, du Jardin (2010) in France, Etemadi et al. (2009) in Iran, Erdogan (2016) in Turkey, and Sandin and Porporato (2008) in Argentina. Tsai et al. (2014b) as well as Nanni and Lumini (2009) analyzed the credit analysis models with data from Germany, Australia and

Japan, while Gepp et al. (2009) and many others have studied the US context. In Brazil, studies conducted by Altman et al. (1979) and more recently by Yim and Mitchell (2009) also explored multivariate models for credit analysis.

The issue associated with bankruptcy prediction is widely studied. For example, Ravi Kumar and Ravi (2007) conducted a comprehensive review of the studies on this subject, poring over the literature spanning nearly 40 years (1968–2005) and analyzing studies that explore applications of statistical and computational techniques. The results showed that modeling through statistical techniques independently is no longer used, and that techniques using artificial intelligence, such as artificial neural networks (ANN), support vector machines (SVM), among others, and hybrid models were found more frequently. The authors also stressed that new machine learning techniques are promising, such as Random Forest and SVM using (non-linear) radial basis function.

Other more recent revisions also feature analyses in this area, such as those by Appiah et al. (2015) and Pereira and Martins (2016). The first one carries out a systematic review since 1966 in 83 articles, stating that there is still room for new models. The second one covers the Brazilian and foreign academic literature since 1930, and suggests that available models be applied, but with new dimensions.

Recent studies on the subject can be found in the literature. Olson et al. (2012) comparatively explore data mining methods, Li (2012) analyzes the predictive power of Altman models currently, and Lyandres and Zhdanov (2013) discuss opportunities of investment in potential bankruptcies. **?** combine accounting and stock market data with variables that represent macroeconomic changes and compare results with benchmark models based on neural networks and on the original study of Altman (1968). The analysis of delinquency/default is also studied in several recent articles such as: Bauer and Agarwal (2014), Tsai et al. (2014a), du Jardin (2010), López-Iturriaga and Sanz (2015), Kim et al. (2015), Cultrera and Brédart (2016) and Erdogan (2016).

Considering the relevance of bankruptcy analysis and the debate in the literature on credit risk, this article evaluates the predictive power of models designed to verify whether a company will go bankrupt within one year from the time of the analysis, by applying traditional statistical methods (MDA and LR) and comparing them to techniques based on artificial intelligence such as ANN, SVM, AdaBoost, Bagging, and Random Forest. Taking into account the context of identifying variables that best assist in the prediction, models were built based on a combination of 45 variables, including financial indicators, growth measures, and rates that measure growth variation, selected after a correlation analysis among such measures.

The sample shows financial information of US companies, and the data on bankruptcy were extracted on a specific basis. During the research, the sample was divided into two subgroups: the first one for the construction of models (training phase), using company data (bankrupt and non-bankrupt) between 1980 and 2005; and the second one with data from 2006 to 2014, in order to recognize the predictive power of the models, also called a test sample or validation sample.

Thus, 108 different models were generated for each technique used, showing results that allow several analyses, both in terms of relevant measures that signal the bankruptcy event, and in terms of quality of outputs produced by each method. The results point to models with high predictive power for all techniques, including the traditional ones. The best result for MDA achieved 88% accuracy, while the AdaBoost technique had many of its models with accuracy above 90%, in addition to being the technique with lowest variation (in terms of accuracy) in the validation samples.

This article is organized as follows. Immediately after this section, we present a review of the research that explores issues of credit risk and bankruptcy prediction. Variables with potential to explain business insolvency are investigated, as well as techniques used in modeling capable of making predictions in different time horizons and different financial contexts. In section 3, there is a description of the methodology, discussing the choice of the sample, the variables and the techniques applied in the study. The results are discussed in section 4. In closing, section 5 presents the conclusions, study limitations, and suggestions for future research.

2. Literature Review

The seminal articles (Beaver, 1966; Altman, 1968; Ohlson, 1980) used statistical methods in their analyses, reaching a bankruptcy predictive capacity of nearly 90%. While Beaver (1966) performed univariate analyses focusing on tests of equal means, Altman (1968) explored a multivariate technique based on discriminant analysis that has some restrictions in its application Sandin and Porporato (2008), such as linearity, normality and independence of the variables, which makes the optimization proposed by the method be achieved only under such conditions (Ong et al., 2011). With the use of Logistic Regression, initially used in research by Ohlson (1980), several assumption limitations were overcome. It is worth noting that Logistic Regression allows the estimation of a response in terms of probability, i.e., what are the chances that a company will go bankrupt, and not a score whereby it is necessary to know specific reference values, as was done by Altman (1968).

Building on his study, Altman et al. (1994) discussed the most common way of modeling by intelligent systems, neural networks, to compare performance in relation to MDA. To do so, they used a sample of Italian companies to diagnose bankruptcy up to year in advance. Little is said about the variables used for the prediction, since the focus of the work was to determine whether ANNs exceed MDA. The authors found that the performance was similar and incredibly high: around 90% accuracy. Accordingly, the authors conclude that ANNs are shown to be useful and that further research in the area could bring breakthroughs based on this method.

Yim and Mitchell (2009), with data from Brazilian companies, conducted a similar analysis. The survey was conducted based on 29 companies in financial difficulties and nearly 100 solvent companies using MDA, LR and ANNs in standard and hybrid form, in which the ANN uses metrics derived from prior treatment via LR. The study showed that such modeling showed good performance, despite the fact that the classification index of failed businesses remained at 67% and the hybrid model reached 89% accuracy in this regard. In the same year, Virág and Kristóf (2005) , conducting a very similar study, but with Hungarian firms, reached similar conclusions, attaining just over 92% accuracy in their tests.

Ong et al. (2011) investigated bankruptcies at publically-traded companies in Malaysia. Using data from a period between 2001 and 2007, the authors tested 11 static variables (measured at the same instant, i.e., not taking into account the fluctuations over time) to build the model, based on LR. Comparing 105 solvent companies with 105 insolvent ones, the results reached 91.5% accuracy in classification (between bankrupt and non-bankrupt), and the authors thus assert that LR is a technique that is superior to MDA, despite not having conducted tests in the same database with any other technique.

Sandin and Porporato (2008) investigated Argentine companies. Even with a limited number of bankruptcies (11), the authors examined the three Altman models (Z, Z' and Z") and developed other methods using a combination of models and other variables. Thus, they were able to verify that it is possible to use Z' to predict bankruptcies in Argentina, even though the model was made for US companies.

The computational/intelligent models start becoming frequent in the literature in the early 2000s. For example, Etemadi et al. (2009) test a genetic programming category, a technique belonging to the family of evolutionary algorithms and very similar to genetic algorithms. Tsai and Wu (2008) analyzed bankruptcy prediction with ensemble classifiers, originating in the ANNs, using three public databases, however identifying that a simple classifier performed better than ensemble classifiers, i.e. than combined classifiers. Working with the same database as Tsai et al. (2014b), Nanni and Lumini (2009) applied other techniques with ensemble classifiers involving bagging, SVM and ANNs, and obtained better results than the previous research.

Other advances were obtained with the study by Olson et al. (2012), in which decision trees (DT) were the technique proposed. Therefore, the authors collected a sample of 100 bankrupt US companies and 100 non-bankrupt companies. The study, not unlike Altman (1968), used traditional financial indicators of the company as well as market indicators. In the model's validation, accuracy was nearly 95%. The authors stress that a recurring problem is the presentation of trees with multiple branches, i.e., several decision points, but that this situation can be corrected with some restrictions during the modeling.

Chuang (2013) also developed models based on decision trees. Just as Olson et al. (2012), the author justifies that the most widely used models have several restrictions and some of the computational techniques also have their problems, as is the case of ANNs, which is much criticized by the obscure treatment of the data, compromising the understanding of the relationships between variables. Chuang (2013) suggests that his model brought improvements in predicting bankruptcy and, at the same time, allows a simple understanding of the studied phenomenon. Such criticisms notwithstanding, ANN-based credit studies have also been conducted, for example, in the work of López-Iturriaga and Sanz (2015).

The article by Tsai et al. (2014b) makes a comparison between various computational methods, such as ANN, SVM and DT based on combinations of Bagging and Boosting, two traditional ensemble techniques. The models were developed using Taiwanese data. The classifiers with the best performance presented by the authors were DT combined with boosting.

Amid so many methods, data and models, Appiah et al. (2015) conducted a systematic review of the literature from 1966 to 2012 and found 137 different models, and realized that the vast majority (61%) of the studies still use statistical techniques, while only 34% apply machine-learning methods and 5% discuss other aspects. Given this theoretical framework, it is possible to understand that computational techniques have a frequent challenge to overcome traditional statistical models.

With regard to the variables tested, the literature provides plenty of proposals. Financial indicators are often used and most models seek to identify how key aspects at companies – such as profit, liquidity, indebtedness, company size, risk, business volume, among others – affect the potential for payment of debts. Among the statistical models, Altman (1968) defined his model with five measures, Ohlson (1980) uses nine indicators, and Virág and Kristóf (2005) use five measures.

After extensively poring over previous studies that used ANNs, du Jardin (2010) listed a number of variables. Most of the studies have shown measures of profitability, liquidity, indebtedness, productivity and efficiency, extracted from data provided in financial statements. In various studies, the reason why explanatory variables are present is not elucidated by the authors, since the objective involves improving the prediction and not necessarily the formulation of hypotheses about the variables that influence bankruptcy.

In particular, de Andrés et al. (2012) found 22 potential variables for information on asset turnover, margins, growth, and cost financing, in addition to the aforementioned measures. After excluding some of them to avoid collinearity, the model worked with eight of these variables. Focusing on the technique and its performance, Yu et al. (2014) worked with nine variables from an initial sample of 41 candidates for entering the model. The authors point out that the models evaluate several indicators and that the sample may therefore reach more than 50 measurements. It is interesting to observe that in this study there are two measures of variation, coincidentally the same ones applied by du Jardin (2010), but were discarded during the selection process. Based on such overview, this article presents research that is aligned with the debate found in the literature mentioned herein.

3. Method

The objective of this research is to test combinations of techniques and variables seeking to build bankruptcy prediction models to improve the predictive performance, based on machine-learning techniques using new variables that have not been applied in any study. Statistical methods of MDA, Logit and ANNs are often used as a reference, and are therefore employed in this study. In all, seven methods were used: SVM, AdaBoost, Bagging, Random Forest (computational), e, RNA, Logit, and MDA.

3.1. Data and Descriptive Statistics

Data collection was performed in two databases. Accounting information, ranging from 1980 to 2014, was extracted from COMPUSTAT. Companies in the financial sector were excluded, because they treat financial information differently from other sectors and have very specific characteristics of indebtedness. The bankrupt companies used in this research are companies available in the database of the Credit & Debt Markets Research Program of NYU's Salomon Center.

As the models need to be developed and then evaluated, the database was separated into two sub-samples, one for training – with data from all companies up to fiscal year 2005 – and one for testing, with the rest of the data. Thus, a database was generated based on the following variables: Total Assets (TA), Earnings before Interest and Taxes (EBIT), Earnings before Interest, Taxes, Depreciation and Amortization (EBITDA), Number of Employees (EMP), Total Sales (SALE), Gross Profit (GP), Net Income (NI), Operating Income (OIBDP), Cost of Goods Sold (COGS), Current Assets (ACT), Current Liabilities (LCT), Inventory (INV), and Operating Cash Flow (OANCF). We then composed 45 variables, divided into six dimensions: Added Value, Profitability, Liquidity, Growth, and Growth of Growth Rate. To facilitate understanding, Table 1 summarizes the measurements.

***** Table 1 here (available at the end of this article) ******

The concept of financial performance used in this article involves the scope of various dimensions, since focusing on only one variable can obscure the relevant analysis of a company's performance. Take, for example, profitability and growth of a company's sales. The company, if it has opted to increase its market share, can forego profitability, for example, by reducing prices, to be consistent with its strategy. Therefore, both variables must be taken into consideration. Another underestimated variable is liquidity, which may be irrelevant in periods of economic stability, but can be crucial in times of depression, when banks no longer offer loans. Thus, liquidity can be the difference between perishing and surviving.

In this sense, our approach is not unlike the approach adopted by Altman (1968), who chose a limited number of variables to build his models. In particular, the design adopted here differs from that of Altman in two crucial aspects. First, the variables that express financial performance were expanded to include six dimensions: added value, profitability, liquidity, cash flow, growth, and growth of growth.

Second, we incorporated ideas of function studies to improve the models capability to predict bankruptcy. The first derivative and the second derivative of a function add relevant information to the variables that explain financial performance. In particular, we can exemplify the concept using the growth rate of sales. A relevant piece of information is knowing in a given period of time that the growth rate is positive. The company can appraise its situation vis-à-vis the competition by comparing, in the same period of time, its performance with that of others from a transversal perspective. Another even more relevant piece of information is to incorporate a longitudinal perspective and verify whether the growth rate increases at constant, increasing or decreasing rates. A company that experiences declines in growth rates may have difficulty in guaranteeing its survival.

Incorporating an analysis of the behavior of growth rates significantly increased the number of models tested, because for each original variable we introduced two more metrics, for a total of 45 explanatory variables. To the best of our knowledge, the variables associated with the constructs of Growth and Growth of Growth, combined with a growth rate, are being tested in credit risk modeling for the first time. The argument for using these constructs involves the design of a variation rate of the variation rate. In the case of infinitesimal variations, we intend to measure the impact of the second derivative to predict bankruptcies. In theory, the sign (positive or negative) of the derivative indicates that this function can be optimized and has a critical point, at which the first derivative has a value of zero at this point. Based on this, we sought to understand which characteristics, evidenced by optimal results, can more efficiently predict the companies that will face bankruptcy and companies that are healthy, with data from the previous fiscal year.

The sample allowed the use of 484 observations (company-year) of bankrupt companies and approximately 115,000 cases for the second group. One can therefore see, within the context of the US stock market, a small number of companies that actually went bankrupt in the period analyzed.

Thus, there was a correlation analysis between variables to avoid models with excessive information. Having reconfigured the information and the constructs, the number of possible combinations is still too large, so we adopted as a criterion for input in the model, selection of one variable for each of the first four dimensions, thereby making a simple combination. The subsamples were created for the training phase, with 77,000 observations and 379 cases of bankruptcy, and in the testing phase, 105 events of bankruptcy and 30,500 solvent firms. Computer models may be impaired when trained by many variables and many observations. Moreover, this measure of "many", which should be objective, it is not discussed in the literature. However, by including all variables and roughly 1,000 observations, the construction of the models takes a considerably long time. Given this computational limitation, we chose to extract a balanced sample – a common procedure in studies of bankruptcy prediction. The selection of non-bankrupt companies was conducted randomly by computer, without any selection bias.

With regard to the execution of the research, the modeling was performed using R software. The packages used in this study were: Performance Analytics (to measure the correlation), ada (AdaBoost), el071 (SVM), rpart (support of other packages), randomForest, MASS (MDA), ipred (bagging), aod (Logit), and nnet (ANN). To verify the performance of the models, two metrics were used: the ROC (receiver operating characteristic) curve, and the overall accuracy (sum of hits divided by the total number of observations).

It should be remembered that the area under the ROC curve, known as AUC, provides a measure of the effectiveness of the model. The ideal value would be 1 (when there are no errors) and the minimum acceptance value would be 0.5, because starting at this point the model has more misses than hits. For discussion purposes, Type I errors (classifying a company that went bankrupt as non-bankrupt) and Type II errors (the opposite of the previous case) were also evaluated. In both cases, the smaller the error, the better the method. However, simultaneously achieving minimum values of the two errors is not so simple in practice.

Below, we present a brief methodological description of the techniques applied in this study.

3.2. Statistical Techniques

Discriminant analysis attempts to derive the linear combination of two or more independent variables that best discriminate between two groups a priori. According to the decision rule, the method seeks to maximize the distance between the groups while minimizing intra-group variation. Accordingly, each company receives a score that is then compared to a cutoff value, which determines to which group the company belongs.

MDA works well as long as the variables in each group follow a normal multivariate distribution and the covariance matrices for all the groups are equal. However, empirical experience has shown that companies, particularly bankrupt ones, violate the condition of normality. Moreover, the second condition is also generally violated. However, empirical studies have shown that the problems related to the assumptions of normality did not weaken their rating capacity, but rather their predictive power.

Logistic regression has also been widely used to investigate the relationship between the probability of binary or ordinal response and explanatory variables. The method adapts logistic regression model of the linear response for binary or ordinal data by the maximum likelihood method. As in discriminant analysis, this technique determines the weights of the independent variables and assigns a score in terms of probability of going bankrupt for each company in the sample. The advantage of this method is that it does not assume the normality covariance matrices and equal multivariate as discriminant analysis does. Hence, the technique incorporates non-linear effects, using the logistic function to measure the chance that bankruptcy will occur.

3.3. Machine-Learning Techniques

ANN: There is a large number of different types of networks, but they are all characterized by the following components: a set of nodes (neurons) and links between the nodes. This process basically consists of inputs (as synapses), which are multiplied by weights (strength of the respective signals), and then calculated by a mathematical function that determines neuron activation. Another function calculates the output of the artificial neuron. In summary, ANNs combine artificial neurons in order to process the information.

SVM: In this algorithm, each data item is marked as a point in n-dimensional space (where n is the number of variables in each observation) with the value of each one providing the coordinates of the point. Then, we carried out the classification to find the hyper-plane that differentiates the two classes. During the classification stage, the goal is to maximize the distances between the nearest points of data, but that are of different classes, so the hyper-plane between them can then be built. This distance allows the creation of a margin. SVM has a characteristic of ignoring outliers and find the hyper-plane that has an optimal margin. Thus, we can say that SVM is robust to outliers. Another important point is that the hyper-plane need not necessarily be linear. It is at this moment that the feature called kernel enters the process. Its role is to transform the space into another, so that it can convert a non-separable problem into a separable one, which is very useful in situations of non-linear separation, usually found in financial data. In this study, we chose to test two kernels, one linear and the other exponential (non-linear) and, therefore, our tests show two models based on SVM.

Bagging: This method is an ensemble method. Its algorithm is quite simple: build several instances of an estimator in random subsets of the original data set, and then add their individual predictions to form a final prediction. These methods are used as a means to reduce the variance of a base estimator, by introducing randomization. In many cases, bagging constitutes a very simple way to improve prediction, with regard to a single model, without making it necessary to adapt the underlying base algorithm.

AdaBoost: The general principle of this method is to fit a sequence of weak classifiers (i.e., models that are only slightly better than random guessing, such as small decision trees) into repeatedly modified versions of the data. Predictions from all of them are then combined through a weighted majority vote (or sum) to produce the final prediction. The modifications of data in each iteration consist of applying weights to each one of the training samples. Initially, these weights are the same for all observations, in such a way that the first step simply trains a weak classifier in the original data. For each successive iteration, sample weights are individually modified and the learning algorithm is reapplied. At any given time, the training examples that were incorrectly predicted in the previous step have their weights increased, whereas the weights are reduced for those that were predicted correctly. Thus, each subsequent weak classifier is thereby forced to focus on getting the cases right that were mistakenly classified previously.

Random Forest: Every tree in the set is constructed from a sample taken with replacement (via bootstrap) from the initial database. When a node is defined during the construction of the tree, it may be that the division is not the best separation between all functionalities. Instead, the division that is chosen is the best separation of a random subset of resources. As a result of this randomization, the polarization of the forest generally increases (compared to a single non-random tree), but due to averaging, the variability also decreases, typically compensating the increase in polarization, thereby yielding a better model in general. Below are the models created, including a review of the main aspects involving the variables and the techniques employed.

4. Results and Discussion

In this study, 108 models were developed and tested for each technique (statistical and computational). It is noteworthy, in the case of SVM, they were applied in two kernels. Thus, in all, the study worked with the generation of 864 different models. Table 2 below shows a descriptive analysis of the training subsample used to construct the models.

***** Table 2 here (available at the end of this article) ******

As the outliers were not extracted from the sample, the median provides a comparative analysis with the mean, even to observe a possible normal distribution of the data. One can see that the task of separating the bankrupt businesses of non-bankrupt ones is quite complicated, because some variables indicate a certain difference. However, this is relative and, for the modeling per se, this imbroglio of observations is a difficulty factor, thus requiring techniques specializing in discrimination.

With regard to models, we present the best results of each technique, describing the relevant aspects. Table 3 shows the most accurate models (ACC), the variables used, the AUC (area under the ROC curve), and measurement of Type I and Type II errors (EI and EII, respectively).

***** Table 3 here (available at the end of this article) ******

We can observe that MDA had the second best performance, driven by the low level of Type II errors, but with an unsuitable index of type I errors. In this case, there would be many bankruptcies that this model would not predict (more than half). It is noteworthy that SVM with linear kernel brought similar results, obtaining a smaller Eli and still committing the same error. However, Random Forest showed results with low mean EI and Eli, which does not overly impair its performance, since the number of solvent companies in the sample is much higher.

It is also worth noting that the type I errors may be more important than the average accuracy of prediction, since a higher Type I error rate requires the lender to increase the cost of the lending operation to compensate for the inability to predict such bankruptcies. From another perspective, type II errors are more difficult to inform whether there is any loss, since it cannot be known if any loss would actually occurs.

***** Table 4 here (available at the end of this article) ******

The graphics shown in Figure 1 illustrate the ROC curves of each method, comparing them with other methods using the same variables.

***** Figure 1 here (available at the end of this article) ******

With graphics of the ROC curves, the superior performance of the computational models is more evident, with higher accuracy in identifying bankrupt firms and healthy firms. We highlight the AdaBoosting, Bagging and Random Forest lines, which visually cause doubts for the analyst to know which one has the best performance.

Table 5 focuses on the analysis of Type I errors. The contrast between the measure and the total accuracy in the best performing models in this regard, RNA and LR, calls for a more cautious analysis. This divergence can be explained by the high number of solvent companies in the test sample, but the type II error virtually invalidates the model. In a practical way, this means to deny credit to almost 3 out of 4 companies suspected of bankruptcy. However, the last three models show a relatively satisfactory metric, through a balance in the measurement of errors.

***** Table 5 here (available at the end of this article) ******

Table 6 shows the models having, as a choice parameter, the best performance in terms of the type II errors.

***** Table 6 here (available at the end of this article) ******

Tables 3 to 6 indicate a number of positive and negative points of each technique, based on 864 different ways to predict bankruptcy. Additionally, the study allows for the monitoring of the influence of financial variables constantly applied in similar studies as well as the new metrics used in this study, given the growth rates of growth.

The results suggest that the MDA models are indeed outperformed. Of the four metrics of performance evaluation, in all of them, the MDA technique was outperformed by another technique or has serious flaws, such as excessive errors. Logistic Regression also showed poor performance vis-à-vis the computational methods. ANN had a very unstable behavior and a performance analogous to MDA, and was even worse than LR. Regarding SVM, the linear model is not one of the best, with critiques equivalent to traditional techniques. On the other hand, SVM with non-linear kernel added little to the study, since it almost always presented median results. Accordingly, we found that SVM needs improvement to be used in this context.

Regarding the three most recent computational techniques – AdaBoost, Bagging and Random Forest – all had milder errors. The best performance often alternated between them, both in terms of AUC as well as accuracy. Thus, the results suggest that the choice between them to choose the best one requires a deeper analysis. Regarding the variables, the tables show the debatable importance of growth rates of growth, given that only 5 out of the 19 best one use these measures. However, the growth dimension was more participatory, present in almost half of the cases, which provides indicia that such measures may better discriminate bankruptcy from non-bankruptcy.

5. Conclusion

The aim of this research was to test new variables such as growth rates and rate of variation of these rates, in credit risk analysis, by comparing the performance of certain computational techniques to predict bankruptcies up to a year in advance at US companies. The techniques of AdaBoosting, Bagging, Random Forest and SVM are considered more recent and have been applied in the financial context, primarily in bankruptcy prediction. Accordingly, we compared their predictive power with the accuracy of methods considered traditional, in this case: MDA, LR and ANN. The results were extracted from more than 800 different models that were developed based on a training sample with nearly 400 cases of bankruptcy; more than 100 bankruptcies and over 30,000 observations of companies in a healthy situation were separated, seeking to evaluate four performance metrics: Type I error, Type II error, Total Accuracy.

We found that the traditional methods were unable to outperform the most recent techniques, for various reasons. In particular, we can mention the issue of instability, since the traditional techniques show great variation in error measurements. Although simpler, i.e., with greater clarity in their interpretation, MDA and LR showed lower results in most tests. One can see that the selection of variables can induce a bias, but vis-à-vis so many models analyzed, this problem was minimized.

This study has some limitations. One such limitation is associated with the sample, which is only of companies from a place with high availability of data and, perhaps for this reason, the analysis is restricted to regions with a similar size database. Another issue is associated with machine-learning techniques, which are addressed in their standard form, without making any changes to the algorithms.

As a suggestion for future work, one could investigate bankruptcies in longer periods (two, three or more years) and/or dividing by industry category, as well as cases of default, expanding the scope of the work in the context of credit risk. Within the scope of the methods, the combination of techniques can also be investigated. In addition, the sample could bring greater wealth of details such as the size of the debt that is involved and that probably will not be paid, i.e., the loss due to the default, which would feature a discussion about type I error. Finally the different costs of misclassification could also be included in future studies.

References

Agarwal, V., Taffler, R., Aug 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. Journal of Banking & Finance 32 (8), 1541–1551.

Altman, E. I., Sep 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance 23 (4), 589–609.

Altman, E. I., Baidya, T. K. N., Dias, L. M. R., Mar 1979. PrevisÃ£o de problemas financeiros em empresas. Rev. adm. empres. 19 (1), 17–28.

Altman, E. I., Marco, G., Varetto, F., May 1994. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). Journal of Banking & Finance 18 (3), 505–529.

Appiah, K. O., Chizema, A., Arthur, J., Sep 2015. Predicting corporate failure: a systematic literature review of methodological issues. Int Jnl Law Management 57 (5), 461–485.

Bauer, J., Agarwal, V., Mar 2014. Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. Journal of Banking & Finance 40, 432–442.

Beaver, W. H., 1966. Financial Ratios As Predictors of Failure. Journal of Accounting Research 4, 71–111.

Chuang, C.-L., Jul 2013. Application of hybrid case-based reasoning for enhanced performance in bankruptcy prediction. Information Sciences 236, 174–185.

Cultrera, L., Brédart, X., Feb 2016. Bankruptcy prediction: the case of Belgian SMEs. Review of Accounting Finance 15 (1), 101–119.

de Andrés, J., Landajo, M., Lorca, P., Jun 2012. Bankruptcy prediction models based on multinorm analysis: An alternative to accounting ratios. Knowledge-Based Systems 30, 67–77.

du Jardin, P., Jun 2010. Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. Neurocomputing 73 (10-12), 2047–2060.

Erdogan, B. E., Apr 2016. Long-term Examination of Bank Crashes Using Panel Logistic Regression: Turkish Banks Failure Case. International Journal of Statistics and Probability 5 (3), 42.

Etemadi, H., Anvary Rostamy, A. A., Dehkordi, H. F., Mar 2009. A genetic programming model for bankruptcy prediction: Empirical evidence from Iran. Expert Systems with Applications 36 (2), 3199–3207.

Gepp, A., Kumar, K., Bhattacharya, S., Nov 2009. Business failure prediction using decision trees. J. Forecast. 29 (6), 536–555.

Kim, M.-J., Kang, D.-K., Kim, H. B., Feb 2015. Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction. Expert Systems with Applications 42 (3), 1074–1082.

Li, J., 2012. Prediction of corporate bankruptcy from 2008 through 2011. Journal of Accounting and Finance 12 (1), 31–41.

López-Iturriaga, F. J., Sanz, I. P., Apr 2015. Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. Expert Systems with Applications 42 (6), 2857–2869.

Lyandres, E., Zhdanov, A., Aug 2013. Investment opportunities and bankruptcy prediction. Journal of Financial Markets 16 (3), 439–476.

Nanni, L., Lumini, A., Mar 2009. An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. Expert Systems with Applications 36 (2), 3028–3033.

Ohlson, J. A., 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. Journal of Accounting Research 18 (1), 109–131.

Olson, D. L., Delen, D., Meng, Y., Jan 2012. Comparative analysis of data mining methods for bankruptcy prediction. Decision Support Systems 52 (2), 464–473.

Ong, S.-W., Choong Yap, V., Khong, R. W., May 2011. Corporate failure prediction: a study of public listed companies in Malaysia. Managerial Finance 37 (6), 553–564.

Pereira, V. S., Martins, V. F., Feb 2016. Estudos de previsão de falências: uma revisão das publicacóes internacionais e brasileiras de 1930 a 2015. R. Contemp. Contab. 12 (26), 163.

Ravi Kumar, P., Ravi, V., Jul 2007. Bankruptcy prediction in banks and firms via statistical and intelligent techniques: A review. European Journal of Operational Research 180 (1), 1–28.

Sandin, A. R., Porporato, M., Nov 2008. Corporate bankruptcy prediction models applied to emerging economies. International Journal of Commerce and Management 17 (4), 295–311.

Tsai, C., Wu, J., May 2008. Using neural network ensembles for bankruptcy prediction and credit scoring. Expert Systems with Applications 34 (4), 2639–2649.

Tsai, C.-F., Hsu, Y.-F., Yen, D. C., Nov 2014a. A comparative study of classifier ensembles for bankruptcy prediction. Applied Soft Computing 24, 977–984.

Tsai, C.-F., Hsu, Y.-F., Yen, D. C., Nov 2014b. A comparative study of classifier ensembles for bankruptcy prediction. Applied Soft Computing 24, 977–984.

Virág, M., Kristóf, T., Nov 2005. Neural Networks in Bankruptcy Prediction - A Comparative Study on the Basis of the First Hungarian Bankruptcy Model. Acta Oeconomica 55 (4), 403–426.

Yim, J., Mitchell, H., 2009. A comparison of corporate distress prediction models in Brazil: hybrid neural networks, logit models and discriminant analysis. Nova Economia 15 (1).

Yu, Q., Miche, Y., Séverin, E., Lendasse, A., Mar 2014. Bankruptcy prediction using Extreme Learning Machine and financial expertise. Neurocomputing 128, 296–302.

Table 1: Description of the Variables used in the research, indicating type and code.

Variable	Construct	Description
X1A	Added Value	SALES – COGS / EMP
X2A	Profitability	EBIT / AT
X2B		GP / AT
X2C		OIBDP / AT
X2D		NI / AT
X3A	Liquidity	WCAP / AT
X3B		ACT / LCT
X3C		ACT - INV / LCT
X4A	Cash Flow	OANCF
X4B		OANCF / EMP
X4C		OANCF / AT
GX1A to GX4C	Growth	These are the growth rates of 11 variables described above, example of formula X1A$_t$ - X1A$_{t-1}$ / X1A$_{t-1}$
GoA		AT$_t - 1$ – AT$_{t-1}$ / AT$_{t-1}$, Growth of Assets
GoS		SALE$_t - 1$ – SALE$_{t-1}$ / SALE$_{t-1}$, Growth of Sales
GoE		EMP$_t - 1$ – EMP$_{t-1}$ / EMP$_{t-1}$, Growth of Employees
GGI		EBITDA$_t$ – EBITDA$_{t-1}$ / EBITDA$_{t-1}$
GOI		OIBDP$_t$ – OIBDP$_{t-1}$ / OIBDP$_{t-1}$
GNI		NI$_t$ – NI$_{t-1}$ / NIt$_{t-1}$
GGX1A to GGNI	Growth of Growth	Example for X1A: GX1A$_t$ – GX1A$_{t-1}$ / GX1A$_{t-1}$; a total of 17 indicators.

Table 2: Descriptive Statistic of the training sample data separating bankrupt companies (F) from non-bankrupt companies (NF).

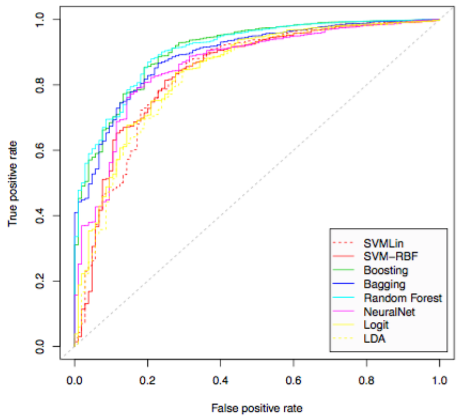
Variables	Mean		Median		1st Quartile		3rd. Quartile	
	F	NF	F	NF	F	NF	F	NF
X1A	41.13	90.67	32.2	57.48	14.21	24.81	55.46	104.52
X2A	-0.08	-0.1	-0.02	0.06	-0.11	-0.03	0.03	0.11
X2B	0.26	0.34	0.19	0.33	0.09	0.17	0.41	0.51
X2C	0	-0.04	0.04	0.11	-0.03	0.01	0.09	0.17
X2D	-0.3	-0.2	-0.16	0.02	-0.34	-0.09	-0.04	0.07
X3A	-0.29	0.04	-0.06	0.23	-0.61	0.05	0.09	0.4
X3B	1.11	2.43	0.82	1.72	0.38	1.07	1.39	2.84
X3C	0.7	1.85	0.42	1.14	0.2	0.71	0.84	2.08
X4A	6.43	220.96	0.31	6.64	-21.84	-0.12	15.24	42.77
X4B	9.72	-15.2	0.04	6.45	-9.83	-3.91	5.18	24.35
X4C	-0.03	-0.04	0	0.06	-0.07	-0.03	0.04	0.13
GX1A	-0.95	0.35	-0.03	0.06	-0.25	-0.09	0.17	0.23
GX2A	-2.06	-0.45	-0.61	0.01	-1.82	-0.36	0.15	0.34
GX2B	-1.69	0.34	-0.01	0.01	-0.19	-0.13	0.28	0.16
GX2C	-4.13	1.98	-0.23	0.01	-0.92	-0.26	0.2	0.25
GX2D	-11.91	-0.68	-1.72	-0.06	-7.01	-0.74	-0.13	0.53
GX3A	-7.61	-0.07	-0.86	-0.02	-4.17	-0.24	-0.13	0.25
GX3B	-0.23	0.31	-0.29	-0.01	-0.69	-0.18	-0.07	0.15
GX3C	-0.22	0.37	-0.35	0	-0.71	-0.19	-0.08	0.2
GX4A	-5.34	0.11	-0.37	0.04	-1.28	-0.52	0.67	0.6
GX4B	-6.28	0.02	-0.42	0	-1.26	-0.56	0.86	0.58
GX4C	-9.18	0.09	-0.45	-0.07	-1.38	-0.57	0.83	0.44
GoA	-0.05	0.12	-0.12	0.05	-0.25	-0.05	0	0.19
GoS	0.06	0.13	-0.03	0.08	-0.14	-0.04	0.09	0.23
GoE	0.47	0.06	-0.07	0.01	-0.18	-0.07	0.02	0.15
GOI	-3.62	0.18	-0.35	0.07	-0.92	-0.23	0.16	0.41
GNI	-8.19	-1.23	-1.3	0.07	-5.35	-0.77	-0.01	0.62
GGX1A	-1.66	1.26	-0.71	-0.03	-1.78	-1.21	1.37	1.9
GGX2A	-8.17	0.17	-0.95	-0.19	-4.63	-1.51	0.89	1.26
GGX2B	-2.01	0.45	-0.27	-0.01	-1.6	-1.38	1.64	1.55
GGX2C	-6.82	-2.63	-0.82	-0.16	-3.3	-1.5	1.04	1.4
GGX2D	-21.76	-3.06	-1.21	-0.14	-12.22	-1.62	0.93	1.19
GGX3A	-41.23	-0.42	-2.23	0	-12.47	-1.41	0.34	1.49
GGX3B	-3.91	2.69	-1.25	0.07	-3.05	-1.38	0.5	1.59
GGX3C	-4.89	1.17	-1.18	0.29	-2.79	-1.2	0.51	1.62
GGX4A	-5.53	-1.58	-0.51	-0.48	-2.13	-1.33	1.32	1.78
GGX4B	-7.91	-1.32	-0.7	-0.52	-2.64	-1.57	1.29	1.62
GGX4C	-93.46	24.17	-0.58	-0.61	-2.66	-1.54	1.35	1.52
GGoA	-4.26	12.21	-1.01	-0.04	-2.33	-0.96	0.16	0.94
GGoS	1.5	3.36	-0.68	0.01	-1.44	-0.95	0.41	1.28
GGoE	-0.21	1.25	-0.53	-0.24	-1.45	-1.11	0.7	0.97
GGOI	-18.03	2.88	-1	-0.04	-2.93	-1.08	0.86	1.53
GGNI	-24.57	8.46	-0.51	-0.1	-7.29	-1.41	0.96	1.25

Table 3: Best results found for each technique in terms of accuracy. All measures are in percentage terms.

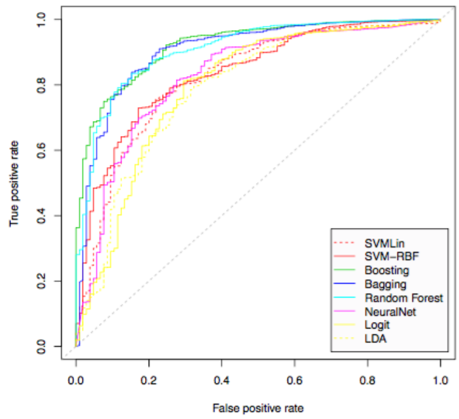
Teqnique	Variables	EI	EII	AUC	ACC
ADM	X1A,X2B,X3A,X4B,GoA, GoS,GoE,GOI,GNI,GX1A, GX2B,GX3A,GX4B,GGoA, GGoS,GGoE,GGOI,GGNI	52.38	10.88	73.41	88.98
RL	X1A,X2D,X3A,X4A,GoA, GoS,GoE,GOI,GNI,GX1A, GX2D,GX3A,GX4A,GGoA, GGoS,GGoE,GGOI,GGNI	39.05	13.12	79.41	86.79
RNA	X1A,X2C,X3A,X4A,GoA, GoS, GoE,GOI,GNI,GX1A, GX2C, GX3A,GX4A, GGX1A,GGX2C, GGX3A, GGX4A,GGoA, GGoS, GGoE, GGOI, GGNI	29.52	24.7	77.68	75.28
SVM (Linear)	X1A, X2D, X3A, X4B, GoA, GoS, GoE, GOI, GNI	53.33	6.86	76.78	92.98
SVM (Non-Lin.)	X1A, X2D, X3B, X4C, GoA, GoS, GoE, GOI, GNI	46.67	11.31	83.27	88.57
AdaBoost	X1A, X2A, X3C, X4B, GoA, GoS, GoE, GOI, GNI	26.67	12.34	88.31	87.61
Bagging	X1A, X2A, X3C, X4C, GoA, GoS, GoE, GOI, GNI	31.43	12.57	87.58	87.36
Random Forest	X1A, X2C, X3C, X4A, GoA, GoS, GoE, GOI, GNI	20.95	12.98	89.99	86.99

Table 4: Best results found for each technique in terms of area under ROC curve. All measures are in percentage terms.

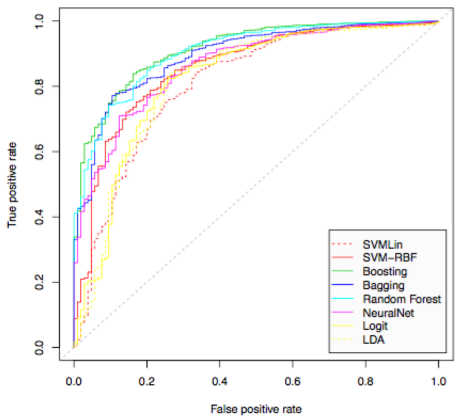
Teqnique	Variables	EI	EII	AUC	ACC
ADM	X1A,X2A,X3C,X4A,GoA, GoS, GoE, GOI, GNI	20.95	30.19	82.88	69.84
RL	X1A,X2D,X3A,X4A,GoA, GoS,GoE,GOI,GNI,GX1A, GX2D,GX3A,GX4A,GGoA, GGoS,GGoE,GGOI, GGNI	39.05	13.12	79.41	86.79
RNA	X1A, X2C, X3A, X4A, GoA, GoS, GoE, GOI, GNI	11.43	38.4	86.54	61.69
SVM (Linear)	X1A, X2C, X3C, X4A, GoA, GoS, GoE, GOI, GNI, GX1A, GX2C, GX3C, GX4A, GGoA, GGoS,GGoE, GGOI, GGNI	24.76	21.76	84.36	78.23
SVM (Non-Linear)	X1A, X2C, X3A, X4A, GoA, GoS, GoE, GOI, GNI	27.62	13.77	85.65	86.18
AdaBoost	X1A, X2D, X3C, X4A, GoA, GoS, GoE, GOI, GNI, GX1A, GX2D, GX3C, GX4A, GGX1A, GGX2D, GGX3C, GGX4A, GGoA, GGoS,GGoE, GGOI, GGNI	19.05	14.08	92.42	85.91
Bagging	X1A, X2D, X3A, X4A, GoA, GoS, GoE, GOI, GNI	17.14	15.68	91.58	84.32
Random Forest	X1A, X2D, X3A, X4A, GoA, GoS, GoE, GOI, GNI	14.29	15.87	91.82	84.14



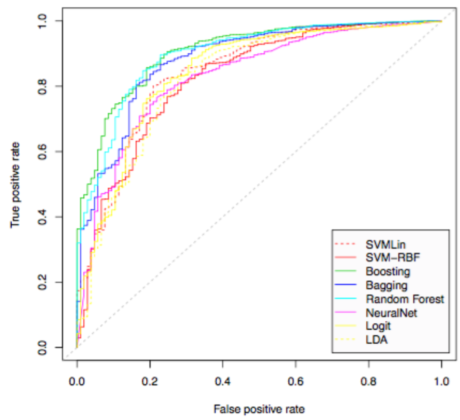
(a) Larger AUC with MDA.



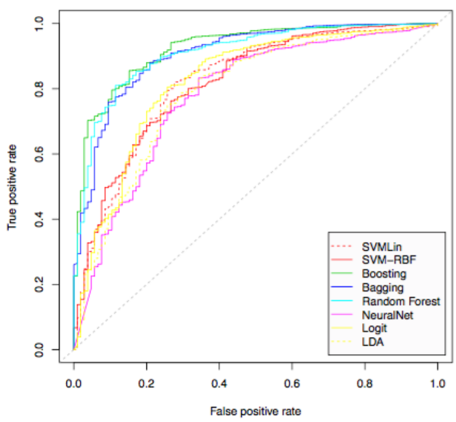
(a) Larger AUC with Logit.



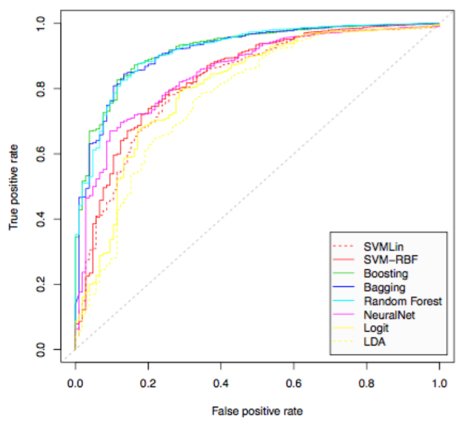
(c) Larger AUC with ANN and SVM (Non-linear).



(d) Larger AUC with SVM – Linear Kernel.



(e) (a) Larger AUC with Adaboost.



(f) Larger AUC with Bagging and Random Forest.

Figure 1: Graphics of the ROC curves of the three top performing computational models, illustrating the best results, in terms of AUC. The curves in each graph also indicate the performance of the other methods, but applied under the same conditions – subsamples and variables.

Table 5: Best results found for each technique in terms of Type I Error. All measures are in percentage terms.

Tecnique	Variables	EI	EII	AUC	ACC
ADM	X1A, X2A, X3C, X4C, GoA, GoS, GoE,GOI,GNI	19.05	34.02	80.93	66.03
RL	X1A,X2A,X3A,X4A, GoA, GoS, GoE, GOI, GNI, GX1A, GX2A, GX3A, GX4A, GGoA, GGoS,GGoE, GGOI, GGNI	6.67	72.1	60.61	28.12
RNA	X1A, X2D, X3B, X4A, GoA, GoS, GoE, GOI, GNI, GX1A, GX2D, GX3B, GX4A, GGX1A, GGX2D, GGX3B, GGX4A, GGoA, GGoS,GGoE, GGOI, GGNI	4.76	79.87	73.55	20.39
SVM (Linear)	X1A, X2D, X3C, X4C, GoA, GoS, GoE, GOI, GNI	21.9	30.2	79.24	69.83
SVM (Non-Lin.)	X1A, X2D, X3A, X4C, GoA, GoS, GoE, GOI, GNI, GX1A, GX2D, GX3A, GX4C, GGX1A, GGX2D, GGX3A, GGX4C,GGoA, GGoS,GGoE, GGOI, GGNI	19.05	26.11	82.9	73.91
AdaBoost	X1A, X2D, X3B, X4A, GoA, GoS, GoE, GOI, GNI	14.29	15.67	91.36	84.34
Bagging	X1A, X2D, X3B, X4A, GoA, GoS, GoE, GOI, GNI	15.24	15.1	90.72	84.9
Random Forest	X1A, X2D, X3A, X4A, GoA, GoS, GoE, GOI, GNI	14.29	15.87	91.82	84.14

Table 6: Best results found for each technique in terms of Type II Error. All measures are in percentage terms.

Tecnique	Variables	EI	EII	AUC	ACC
ADM	X1A, X2B, X3A, X4B, GoA, GoS, GoE, GOI, GNI,GX1A, GX2B, GX3A, GX4B, GGoA, GGoS,GGoE, GGOI, GGNI	52.38	10.88	73.41	88.98
RL	X1A, X2D, X3A, X4A, GoA, GoS, GoE, GOI, GNI, GX1A, GX2D, GX3A, GX4A, GGoA, GGoS,GGoE, GGOI, GGNI	39.05	13.12	79.41	86.79
RNA	X1A, X2A, X3A, X4B, GoA, GoS, GoE, GOI, GNI, GX1A, GX2A, GX3A, GX4B, GGX1A, GGX2A, GGX3A, GGX4B, GGoA, GGoS, GGoE, GGOI, GGNI	27.62	32.01	72.15	68
SVM (Linear)	X1A, X2D, X3A, X4B, GoA, GoS, GoE, GOI, GNI	53.33	6.86	76.78	92.98
SVM (Non-Lin.)	X1A, X2D, X3B, X4C, GoA, GoS, GoE, GOI, GNI	46.67	11.31	83.27	88.57
AdaBoost	X1A, X2A, X3C, X4B, GoA, GoS, GoE, GOI, GNI	26.67	12.34	88.31	87.61
Bagging	X1A, X2A, X3C, X4C, GoA, GoS, GoE, GOI, GNI	31.43	12.57	87.58	87.36
Random Forest	X1A, X2C, X3C, X4C, GoA, GoS, GoE, GOI, GNI	29.52	12.96	87.77	86.99