



Evolutionary feature selection approaches for insolvency business prediction with genetic programming

Ángel Beade¹ · Manuel Rodríguez¹ · José Santos²

Accepted: 13 June 2023 / Published online: 9 July 2023
© The Author(s) 2023

Abstract

This study uses different feature selection methods in the field of business failure prediction and tests the capability of Genetic Programming (GP) as an appropriate classifier in this field. The prediction models categorize the insolvency/non-insolvency of a firm one year in advance from a large set of financial ratios. Different selection strategies based on two evolutionary algorithms were used to reduce the dimensionality of the financial features considered. The first method considers the combination between the global search provided by an evolutionary algorithm (differential evolution) with a simple classifier, together with the possible use of classical filters in a first step of feature selection. Secondly, genetic programming is used as a feature selector. In addition, these selection approaches will be compared when GP is used exclusively as a classifier. The results show that, when using GP as a classifier method, the proposed selection method with GP stands out from the rest. Moreover, the use of GP as a classifier improves the results with respect to other classifier methods. This shows an added value to the use of GP in this field, in addition to the interpretability of GP prediction models.

Keywords Differential evolution · Genetic programming · Feature selection · Prediction of business insolvency

1 Introduction

Business failure prediction models have the main objective of anticipating the difficulties of a company and providing a useful tool for decision-making by the multiple agents involved and by the company itself. This insolvency prediction is treated as a classification problem based on the information available in the companies' financial statements. For the prediction/classification, initial works used

statistical techniques, such as discriminant or logit analysis (Altman et al. 1994), while in the last two decades machine learning methods have provided suitable approaches to obtain financial decision support systems that can help monitor the financial health of companies (Jayasekera 2018).

In this field of business failure prediction, the selection of explanatory variables is, together with the choice of the classifier method, one of the basic challenges (Volkov et al. 2017). However, there is no consensus on how to approach that selection of explanatory variables. Barnes (1987) pointed out that financial ratios (the most common type of explanatory variable in works on prediction of business insolvency) were basically selected on the basis of their popularity, together with some new ones (financial ratios) that researchers were adding. The situation has not undergone major changes as shown by du Jardin (2009) in summarizing the criteria used to select the explanatory variables to be included in models of business insolvency prediction. The author indicates that 40% of the analyzed works use, in the selection of the variables, the criterion of their popularity in previous works as well as their reported

✉ Ángel Beade
a.beade@udc.es

Manuel Rodríguez
manuel.rodriguez.lopez@udc.es

José Santos
jose.santos@udc.es

¹ Business Department, University of A Coruña, Campus de Elviña, s/n, 15071 A Coruña, Spain

² CITIC (Centre for Information and Communications Technology Research), Department of Computer Science and Information Technologies, University of A Coruña, A Coruña, Spain

predictive capacity. Alaka et al. (2018) provide an overview of the methods used in the selection of variables in 49 studies between 2010 and 2015 that may serve to highlight the above lack of consensus.

In addition, the financial information available from companies (for public and/or restricted use) has increased exponentially in recent times, at the same time as classification techniques have increased the capacity to deal with a greater number and type of variables. However, it is not uncommon for the available dataset to be overloaded with a multitude of financial features (which can be used as input variables for the prediction models) which, in addition to increasing the time and costs inherent in obtaining solutions, can lead to an excessive adjustment of the models obtained.

Consequently, given the large number of features that can be considered from that financial information for the learning models, automatic selection of features is an appropriate alternative to be considered. In classification problems, the objective of feature selection is to discover the most relevant characteristics, trying to reduce the dimensionality of a (typically) high feature space without compromising the classifier performance.

Feature selection methods are commonly categorized based on whether the selection criterion depends on the classifier (or predictor) and its associated learning algorithm (Chandrashekar and Sahin 2014):

- i. The *filter approach*, which does not depend on the classifier/learning algorithm. An external measure determines, only from the data, the most relevant features for the classification. That is, these methods use measures which only depend on intrinsic properties of the data. Typically these methods select the features based on statistical measures and information content of the features. For example, features in the datapoints that do not have correlations with the object class are primarily responsible for misclassification (Chakravarty et al. 2013) and, therefore, are candidates to be discarded.
- ii. The *wrapper approach*, which depends on the specific classifier model employed. One possibility in this approach is to train the classifier with different subsets of characteristics, selecting the features that provide the best classification performance. Another possibility is that the classifier (with all the features), can select the most relevant ones for the classification, such as the sensitivity analysis in trained artificial neural networks (Yeung et al. 2010).

Finally, when the selection is integrated as part of the learning algorithm, these approaches are sometimes referred to as *embedded* methods.

Filtering methods are typically faster than wrapper methods. However, filtering methods ignore the possible interaction of features and the performance of the characteristics in the specific classification algorithm. On the other hand, the usual drawback of the wrapper methods is their high computational burden (Zhao et al. 2018).

Another possibility is feature reduction without preserving the original features of the datapoints. The most widely used method is Principal Component Analysis (PCA) (Jolliffe 2002), which projects the features along the principal components. Usually, this projection is restricted to the (few) largest eigenvectors. This PCA reduction has a linear dependence of the original features.

Our study attempts to decipher the most representative variables for the business failure prediction problem, using different strategies for feature selection, extending our previous work presented at conference IWINAC 2022 (Santos et al. 2022). First, an Evolutionary Algorithm (EA) was used for the automatic selection of the most relevant features (financial ratios) in predicting (1-year ahead) business insolvency. Differential Evolution (DE) (Price et al. 2005) was selected as a robust evolutionary algorithm to automatically determine that subset of relevant features.

DE genetic population will encode solutions that specify the selected ratios from an ample set of predefined financial ratios. Our work considers a simple classifier (K-Nearest Neighbors - KNN) to provide a low computational cost when the fitness is computed. Thus, the fitness associated with each encoded solution or subset of selected ratios will be determined by how well that subset of ratios solves the classification task with this simple classifier model (KNN). The evolutionary algorithm obtains an optimized selected subset for the classification task. Therefore, it is a wrapped method, since the selection depends on the classifier used.

Two important properties of this method (combination between DE and KNN) can be emphasized: i) the evolutionary algorithm performs a global search over the space of possible solutions (subset of selected ratios). ii) The EA automatically selects the number of selected ratios, without an a priori decision on the appropriate number of ratios to be selected. Finally, the combined method (DE/KNN) depends on the KNN classifier used, but is easily generalizable to other classifiers only by changing the definition of the fitness function, considering the classification performance provided by other classifiers.

Furthermore, the method can be combined with filtering methods if the initial set of ratios, which is considered by the evolutionary algorithm, is chosen taking into account the relevance provided by classical statistical filtering methods. Consequently, experiments that use this alternative are considered in this study, comparing the classification performance when a first selection of the pool of ratios (with two filters, Fisher Score and T-statistic) is employed.

Secondly, this study also uses Genetic Programming (GP) (Koza 1992; Poli et al. 2008) as a classifier method in business insolvency prediction as well as a method to reduce the dimension of a set of variables in order to detect the relevant ones (explanatory variables of the prediction model). The proposed GP reduction method will be based on the frequency of selection of the explanatory variables in a set of independent GP runs (which can be considered as an embedded method). Moreover, this set of variables obtained with GP as a dimensionality reduction method is evaluated - together with the one obtained with DE/KNN - now using GP exclusively as a classifier method.

In the selection of GP as a classifier, two considerations have been taken into account: i) The interpretability of the tree or program optimized with GP. In the words of Brabazon et al. (2020), GP can provide human readable solutions, which is important in the field at hand. ii) The inherent variable selection process involved in the GP evolutionary process, always adjusted to each particular model (a property that is also used as a basis in the feature selection process with GP).

Therefore, one main goal in our study is to test the capability of GP as an appropriate classifier in the field of business failure prediction. Within this goal, we used different feature selection methods, combining different selection strategies (filter methods, DE/KNN wrapper method and GP as feature selector) to test that GP capability. These selection approaches will be compared when GP is used as a classifier. This feature selection process with these methods allows us to decipher whether dimensionality reduction improves GP classification results and to decipher whether there are financial features more relevant than others for classification.

The remainder of the paper is structured as follows. Previous work using feature selection methods in business failure prediction is discussed in Sect. 2, focusing on approaches using the evolutionary algorithms employed in our study. Section 3 details the methods used: the EA employed (DE), how the solutions are encoded in each genotype of the population, the definition of the fitness

function with the KNN classifier and the selection of the data for the experiments. Moreover, Sect. 3 also details the GP environment employed and the use of GP as a feature selection method. Section 4 expounds the experiments performed, with the comparison of different alternatives regarding the initial set of ratios with DE/KNN as variable selector, as well as using GP as selector. Then, an experiment is detailed using GP only as classifier with the subsets obtained with the different methods of variable selection. Finally, Sect. 5 includes a discussion of the main results and a summary of the main conclusions that can be drawn.

2 Previous work

Several works have experimented with the use of feature selection strategies in the specific problem of business insolvency prediction. For example, Liang et al. Liang et al. (2015) used filter and wrapper based feature selection methods on financial distress prediction. The studies by Tsai et al. Tsai et al. (2021), Lin et al. Lin et al. (2019) and Papíková and Papík Papíková and Papík (2022) analyze the effect of several feature selection methods on single and ensemble learning-based prediction models.

Our work focuses on insolvency prediction in a specific domain, small and medium-sized companies. (Sect. 3.4). Two selection strategies based on two evolutionary algorithms were used to reduce the dimensionality of the financial features considered. The first method considers the combination between differential evolution with a simple KNN classifier. In this first case, DE was selected because is a well-established, contrasted and robust method with proven advantages over other evolutionary methods (Das and Suganthan 2011; Eltaieb and Mahmood 2018).

The surveys of Xue et al. (2016) and Dokeroglu et al. (2022) provide an overview of different possibilities when EAs are applied to feature selection. Focusing on DE, in some previous works, DE was used with the filter and wrapper categories of feature selection. For example, Chakravarty et al. (2013) used DE to automatically find out the most important features with a filter approach. In their work (Chakravarty et al. 2013), the DE genetic population encodes possible feature subsets and the fitness function considers the intra-class and inter-class variation of the feature values, combined with a Lagrangian multiplier.

On the contrary, Zhao et al. (2018) used DE in combination with both feature selection approaches. First, three filtering methods (Fisher Score, T-statistic and Information Gain) were used to generate the feature pool as input to DE. Then a

Support Vector Machine (SVM) was used as the classifier in a 10-fold cross-validation method in order to set the fitness of each feature subset. Serrano-Silva et al. (2018) also used metaheuristic algorithms (including DE) for automatic feature weighting in different classification problems, including bankruptcy prediction (only with six attributes or features). The metaheuristic establishes the weights for the attributes or features that describe financial data and in order to improve the results of a Naïve Associative Classifier employed by the authors (Serrano-Silva et al. 2018).

Our work with DE follows a similar methodology to the work by Zhao et al. (2018), which employs a support vector machine as classifier model in a different application (molecular signature information for biomarker discovery). A similar method was also used by Salcedo-Sanz et al. in Salcedo-Sanz et al. (2004) for the insolvency prediction of non-life insurance companies, but using a Simulated Annealing search hybridized with a SVM to find the most relevant ratios from a limited set of 21 financial ratios.

Genetic Programming (GP) (Koza 1992; Poli et al. 2008) has also been used in our study with two functions: firstly as a feature selection method and secondly as a classifier to compare the considered feature selection strategies. GP evolves programs, usually represented as trees. As mentioned, GP is considered here given its useful properties in the field, especially the interpretability of the optimized solutions (trees), which is often a requirement for an end user of the prediction model (e.g., a credit lender such as a bank or an investor).

Regarding the use of GP as a classifier, on one hand, there are few previous works using evolutionary computation methods and, in particular GP, in this field of prediction of business insolvency. For example, Salcedo-Sanz et al. (2005) applied GP for the prediction of insolvency of non-life insurance companies, using a limited set of 19 financial ratios. As the authors state, GP provides an optimized decision tree with the important property that the tree can be inspected, interpreted and reused in different data sets (Salcedo-Sanz et al. 2005). Their results show a better performance in comparison with a support vector machine. Some other previous works has also used GP as a classifier in particular domains in business prediction, with the specific goal of comparison with other classification techniques (Alfaro-Cid et al. 2009; Etemadi et al. 2009; García-Almanza et al. 2010; Lensberg et al. 2006; McKee and Lensberg 2002). On the other hand, the use of GP as a method for automatic selection of financial ratios in insolvency prediction is novel, although there are few previous works in other applications (Neshatian and Zhang 2011; Viegas et al. 2018).

3 Methods

3.1 Differential evolution/KNN feature selection

3.1.1 Differential evolution

The selection of the financial ratios was performed by the global search provided by an evolutionary algorithm. Not being the objective here the comparison of different evolutionary algorithms or other bioinspired metaheuristics in the application, DE was selected, as previously mentioned, since it is robust and contrasted method that provides better results than other evolutionary methods in many applications (Das and Suganthan 2011; Eltaieb and Mahmood 2018) and it also has an easy implementation with few defining parameters.

DE (Price et al. 2005) maintains a set of vectors (encoded solutions) in its genetic population. The key aspect in DE is that new candidate solutions are created by simple operations between the vectors or current solutions, in particular the difference between vectors to define perturbations for the generation of the trial or candidate solutions. A simple selection process keeps the population constant through generations with the best found solutions. Algorithm 1 summarizes the main steps of the standard DE algorithm.

A limited number of parameters is needed to define the implementation of differential evolution. In addition to the size of the population, there are two parameters: the weight factor F and the crossover probability CR . These parameters are used to define a “trial” or “candidate” vector (y) for each “target” vector x of the population (main loop between lines 4-19). The factor F (standard values in $[0, 2]$) weights the difference between (randomly chosen) pairs of vectors (x_2 and x_3 in Algorithm 1). A “donor” or “mutant” vector ($x_1 + F(x_2 - x_3)$) is created from that weighted difference of those two vectors and added to the base vector x_1 (line 10 in the pseudo-code). Next, a crossover operation is performed between the target vector (x) and the mutant vector, with probability CR . This crossover operation is performed independently at each position of the vector, corresponding to the internal loop in lines 7-14 in the pseudo-code of Algorithm 1.

The result of the crossover operation defines the final trial vector (y) for each target vector x . The standard “binomial” crossover (specified in Algorithm 1) was used in our work. In this crossover step, regardless of the crossover probability, the trial vector always incorporates genetic material from the mutant vector, as guaranteed by the index R (pseudo-code, line 9).

Algorithm 1 Differential Evolution algorithm.

```

1: Initialize the population of random solutions
2: Evaluate vectors (encoded solutions) (KNN accuracy with the encoded selected ratios)
3: repeat
4:   for all vector  $x$  in the population do
5:     Let  $x_1, x_2, x_3 \in$  population, randomly obtained  $\{x_1, x_2, x_3, x$  different from each other $\}$ 
6:     Let  $R \in \{1, \dots, n\}$ , randomly obtained  $\{n$  is the dimension of the search space $\}$ 
7:     for  $i = 1$  to  $n$  do
8:       Pick  $r_i \in U(0, 1)$  uniformly from the open range  $(0, 1)$ .
9:       if  $(i = R) \vee (r_i < CR)$  then
10:         $y_i \leftarrow x_{1i} + F(x_{2i} - x_{3i})$   $\{CR -$  Crossover probability,  $F -$  Weight factor $\}$ 
11:       else
12:         $y_i = x_i$ 
13:       end if
14:     end for  $\{y = [y_1, y_2 \dots y_n]$  is a new generated candidate or trial vector $\}$ 
15:     Evaluate fitness  $f(y)$  of candidate  $y$  (KNN accuracy with its encoded selected ratios)
16:     if  $f(y) \leq f(x)$  then
17:       Replace vector  $x$  by  $y$   $\{$ if  $y$  has better or equal fitness $\}$ 
18:     end if
19:   end for
20: until termination criterion is met
21: return  $z \in$  population  $\forall t \in$  population,  $f(z) \leq f(t)$ 

```

The population size is maintained constant by the selection process. The trial vector (y) and the target vector (x) are compared, keeping in the next evolutionary generation the fittest one (pseudo-code, line 16). In this way the algorithm incorporates elitism, since the best solution (vector) is maintained or improved throughout the generations.

DE uses different possibilities for defining the candidate vector, changing the selection of the base vector, the number of vector differences or the crossover type. The standard DE scheme *DE/rand/1/bin* was used here, in which the base vector x_1 was randomly selected (*bin* denotes the crossover type and 1 denotes the number of vector differences involved when the mutant vector is defined). Therefore, this DE scheme provides a low selective pressure.

One key aspect in DE is the adaptive nature of the step length when the mutant vector is defined ($x_2 - x_3$, used in the search of new solutions), which is progressively adapted along the evolutionary process (Feoktistov 2006). This step length tends to be large in the initial generations, since individuals are largely distant from each other. However, throughout the evolutionary process, the solutions of the population tend to concentrate in the promising found areas of the search landscape, so the differences tend to be progressively smaller. Therefore, this mechanism of vector difference provides an automatic control of exploration/exploitation level in the search process.

Finally, it should also be noted that the aim of the present paper is not to compare different DE variants in the application. The standard DE algorithm is appropriate for our purpose focused on its combination with a simple classifier to perform the feature selection process in the application, without requiring modern DE versions (Eltaeib

and Mahmood 2018) with, for example, self-adaptation of the DE defining parameters.

3.1.2 Encoding of solutions

Each vector of the DE genetic population encodes a subset of selected features. On the contrary to the work of Zhao et al. (2018), which used a binary encoding for representing the selected features in the vectors of the DE population, we used a real-valued encoding and, therefore, the standard DE operators can be used, without requiring binary DE versions (Doerr and Zheng 2020).

Given a complete set of D features (financial ratios), every vector is a D real-value vector, with values in the range $[-1, 1]$. Positive or zero values denote that the corresponding features are selected, while the negative values indicate that the feature is not considered. Therefore, a given phenotype (selected subset of features) can be represented by different genotypes, since each selection/non-selection of a feature can be represented with different real values in the corresponding range. In the generation of the mutant vector, if a value exceeds such bounds ($[-1, 1]$), the resulting value is set as a random value in the range (Price et al. 2005). These aspects help to maintain diversity in the population.

Note also that the encoding allows automatically selecting the most appropriate number of selected features, i.e., it is not required to establish how many relevant features should be selected.

3.1.3 KNN classifier and fitness definition

The K-nearest neighbors algorithm is a non-parametric method used for classification and regression (Altman

1992). In the KNN supervised classification, the distance between an input instance (vector in a multidimensional feature space) and every instance of labeled (correctly classified) examples in a training set is calculated. The input instance is classified by a plurality vote of its neighbors, with the instance being assigned to the class most common among its K nearest neighbors. Thus, the complexity of the KNN classifier only depends on the number of neighbors (K) used in the classification. If $K = 1$, then the input example is simply assigned to the class of that single nearest neighbor. The most appropriate value for the parameter K is very dependent on the classification problem. Higher values of K can reduce the effect of the noise on the classification, but make boundaries between classes less distinct. Therefore, we employed two values in the ranges normally considered for the parameter K ($K = 3$ and $K = 15$).

The fitness function will be given by the classification accuracy provided by the KNN classifier with the subset of selected ratios that encodes each population solution. That accuracy is considered with a test set (Sect. 3.4), different from the training set, the latter with a priori correct classifications. The Euclidean distance between the values of the selected financial ratios was considered as measure of distance in the KNN classifier. All the ratios were normalized in $[0,1]$, taking into account the lowest and highest values in each ratio (considering all the values in the training and complete test sets detailed in Sect. 3.4).

3.2 Genetic programming as a feature selector

In GP, a population of computer “programs” is evolved (Koza 1992; Poli et al. 2008). Several features of GP are appropriate for the particular field of insolvency prediction with the goal of obtaining models with high predictive power. First, GP does not establish prior assumptions about the explanatory variables of a prediction model. In addition, as mentioned above, the GP allows direct interpretation of the optimized tree or program. Finally, the ability to somehow regulate the complexity of the optimized tree or program by parameterizing its depth, the number of nodes or the functions that can be used in the search for a solution. Given these GP properties, GP was selected as classifier to compare the different selection approaches considered. Figure 2 below shows an example of an evolved program (prediction model), while Sect. 4.2.1 will detail the possible functions used in the tree nodes.

Nevertheless, GP can also be used as a feature selection strategy. The proposed approach is based on the hypothesis that in GP the relevant variables are maintained in the population and throughout the generations of the evolutionary process, while the irrelevant ones will be eliminated due to selection pressure. For this purpose, the

relative frequency of occurrence of each input variable in the totality of the trees and in the totality of the nodes, as well as over the totality of generations of each run, will be used. By aggregating, the relative frequency of each input variable in a set of GP runs can be obtained. The measure may not be absolutely accurate, but the main GP inputs (measured by their relative frequency) are expected to be the most important variables in the finally selected solutions. The following steps will be followed for this purpose:

- A large number (1,000) of independent GP runs is performed with the totality of the input variables.
- On the results of this experiment, a subset of GP runs is established which generate the best solutions (5% of the total).
- In the runs that provide the best solutions, the relative frequencies of each variable are obtained and, depending on the p-value (null hypothesis: the selection of these variables was by chance), a decision is made as to which subset of input variables is selected. The null hypothesis implies that the typified relative frequency of occurrence of the variables should follow a standard normal distribution - $N(0,1)$. To verify this, the Kolmogorov-Smirnov test is applied and the p-value is obtained. Those variables with p-value < 0.05 have been chosen. The choice of a low p-value seeks to select variables whose results are statistically significant and not due to chance.

In addition, as mentioned, GP will be used exclusively as a classifier method in the comparison of the different dimensionality reduction methods. In this case, given the stochasticity of GP, once again 1,000 independent runs of GP are performed but with the subset of input variables determined by each dimensionality reduction method. As GP provides decision trees with a threshold associated to the tree classification, we will use the average AUC of the 5 best solutions, which is also a common measure in the field of business insolvency prediction with models with an associated classification threshold.

It should be noted that the 1,000 independent GP runs to determine the relevant variables (when GP is used as a selector) are different and unrelated to the GP runs when GP is used as a classifier. That is, variable selection with GP is a prior process, independent of the subsequent use of GP as a classifier.

The software used to use GP, both as a classifier and as a feature selection method, has been HeuristicLab (Wagner et al. 2014). This software can be downloaded from their website [18]. HeuristicLab was selected because it is an extensible, paradigm-independent optimization environment that strongly abstracts the heuristic optimization

process, also with a detailed user interface (Wagner et al. 2014).

3.3 Set of ratios used

As mentioned in the Introduction, financial ratios were mainly used as discriminative and explanatory variables in the prediction/classification models. Financial ratios are defined as relative magnitudes of two selected numerical values taken from financial statements (balance sheet, income statement and cash flow statement) to gain meaningful information about a company. These ratios are related to, e.g., liquidity, leverage, growth, margins, profitability, rates of return and valuation.

The use of financial ratios is justified on the use and effectiveness that these financial indicators have shown in predicting business insolvency (du Jardin 2009). An ample initial set of fifty-nine ratios was used, a larger number of ratios compared to previous works in the prediction of business insolvency. This set of fifty-nine financial ratios was selected based on the popularity of each ratio in the accounting literature, as well as their significance in the relevant studies related to business insolvency.

The ratios are grouped into different financial measures. This grouping is done because each ratio, belonging to a financial measure, has a similar purpose to the other ratios belonging to the same group. Below is a brief description of each group, whereas Table 6 (in Appendix A) includes a brief definition of the financial ratios considered:

- Activity-related (ACT01-ACT05): these type of ratios are linked to the volume of operations carried out by a company.
- Leverage (LEV01-LEV04):
 - Operating leverage: it is the ability of companies to use costs fixed operations to optimize the effects of changes in sales, that is, it measures how a company's operating income will change in response to a change in sales.
 - Financial leverage: unlike operating leverage, financial leverage takes into account financial costs (taxes, fees, etc.).
- Debt (DEB01-DEB03): the debt ratios are responsible for measuring the volume of external financing that the company uses.
- Structure (STR01-STR09): these ratios are responsible for measuring the percentage structure of assets and liabilities.
- Liquidity (LIQ01-LIQ13): these provide an idea of whether a company will be able to pay its debts when due.
- Profitability (PRF01-PRF06): from them it is possible to have an idea of whether the company generates sufficient income to cover costs and remunerate its owners.
- Turnover (TUR01-TUR08): these quantify the performance of a company in a specified period of time.
- Solvency (SOL01-SOL09): these measure the ability to meet financial obligations (both short and long term).
- Treasury (TRS01-TRS02): these measure the ability to meet short-term financial obligations.

3.4 Dataset of companies

The selected companies correspond to small and medium companies (SMEs) located in the Autonomous Community of Galicia (Spain). A wide range of SMEs is used, although excluding financial, insurance, construction and real estate agencies, due to the particular characteristics of these activity sectors. The data of the companies were obtained from the Iberian Balance Sheet Analysis System (SABI) dataset [30]. This is a tool that provides balance sheets information of more than 2.5 million Spanish companies. The criterion used in this work for categorizing a company as failed was the legal declaration of suspension of payments, which is the concept most used in business bankruptcy studies.

The available business population was divided into two subsets:

- A first sample set (training set), which is made up of 136 failed (insolvent) companies, whose 1 year prior to insolvency was in the interval between 2007 and 2012 (both included). The same number (136) of non-failed companies was included in this set, these chosen from (non-failed) 2,298 companies in total, and matched with those failed by: 1) accounting year, 2) volume of asset and 3) activity sector.
- A test set consisting of another set of 136 failed companies and 2,389 non-failed companies. However, for the evaluation of the fitness of each encoded solution, a reduced test set was considered, with the same number (136) of failed and non-failed companies (and the same matching criteria as in the case of the training set). Thus, this reduced set has balanced data for an easier interpretation of the results of the classification using the selected ratios encoded in every genotype, in addition to the lower computational time required when the fitness is calculated for every encoded solution. Although other measures based on the confusion matrix could be considered, given the balanced data in this reduced test set, the accuracy or percentage of correctly classified companies is used as the fitness associated with each encoded solution since

it provides a balanced indicator of the classification performance in both categories of companies.

Moreover, with the whole test set, many more input patterns or records are considered. A “record” consists of the data of a company in a fiscal year (with its category of failure/non-failure 1 year in the future). In other words, the same company can provide different records corresponding to different years. In this way, the whole test set provides 18,360 records (136 belonging to companies that fail and 18,224 obtained from the 2,389 non-failed companies in different fiscal years). This complete test set is used for comparisons of results between different classification approaches.

The instances of the reduced test set are, therefore, a subset of the full test set and, specifically, the 136 observations of failed companies. These instances are always the same in the different evaluations performed with both test sets.

4 Results

4.1 Results with the DE/KNN approach

4.1.1 DE/KNN setup

With the DE/KNN automatic selection of ratios, different “test variants” are considered:

- i. T1-30FS: DE selects from the 30 most relevant ratios according to the Fisher Score.
- ii. T2-30TS: DE selects from the 30 most relevant ratios according to the T-statistic.
- iii. T3-59 Ratios: DE selects from the entire set of 59 ratios.

Thus, the first two variants initially select the 30 most relevant ratios according to the Fisher Score and T-statistic filters (filter definitions can be found, for example, in Zhao et al. (2018)). These 30 preselected ratios are the ones that the DE process will consider in the evolutionary selection process. Consequently, the genotypes encode the selection or not of these 30 ratios. The third test variant considers the entire set of 59 ratios without any prior selection, that is,

the EA selects the appropriate ratios from the full set. Moreover, these three variants were tested with two values of the parameter K in the classifier ($K = 3$ and $K = 15$).

The setup of DE was: population size=100, low crossover probability ($CR = 0.1$), whereas the F parameter takes a random value in the interval $[0,9]$ every time a candidate solution is defined (Sect. 3.1.1), and the DE process was run over 500 generations. These parameters were experimentally adjusted to provide the best results in most of the test variants, avoiding also premature convergence.

With each test variant, the evolutionary algorithm was run 30 times to select the ratios most appropriate for the classification process. The selected ratios in each test variant were those selected in the best solution and in the 30 independent runs.

4.1.2 Comparison of percentage of correct answers with the DE/KNN selected ratios

Table 1 shows a summary of the results with the three test variants considered. Table 1 specifies, for each test variant and value of K in the KNN classifier, the average value of the best result in the independent DE runs, as well as the best value of such independent runs. These values correspond to the fitness (accuracy) considering the (reduced) test set, that is, the accuracy or correct classified companies of the reduced (and balanced) test set.

Taking into account the results in Table 1 it is not possible to establish which is the best value of K since, using the 59 ratios (in the initial pool of ratios) and $K = 3$, slightly better values are obtained (with respect to $K = 15$). However, in the other two test variants, the best values (in each test variant) are obtained with $K = 15$.

Figure 1 shows the percentage of times (normalized in $[0,1]$) that the different ratios have been selected (in the best solution) in all the tests performed with the EA, that is, considering the six different variants included in Table 1. Therefore, Fig. 1 illustrates the number of times each ratio has been selected (in the best solution) in the EA runs and in the different test variants, regardless of whether it has been finally selected in each test variant. For comparison, Fig. 1 also includes the normalized values of the T-statistic and Fisher Score measures for all ratios.

Table 1 Classification accuracy (fitness) in the different test variants (using the reduced test set)

Test variant (starting pool of ratios)	Number de neighbors (KNN)			
	3NN		15NN	
	Average	Best	Average	Best
59 ratios (T3-59 Ratios)	88.48	90.77	88.17	89.63
30 best ratios with Fisher Score (T1-30FS)	86.23	88.6	88.40	89.70
30 best ratios with T-statistic (T2-30TS)	85.87	87.20	87.69	89.44

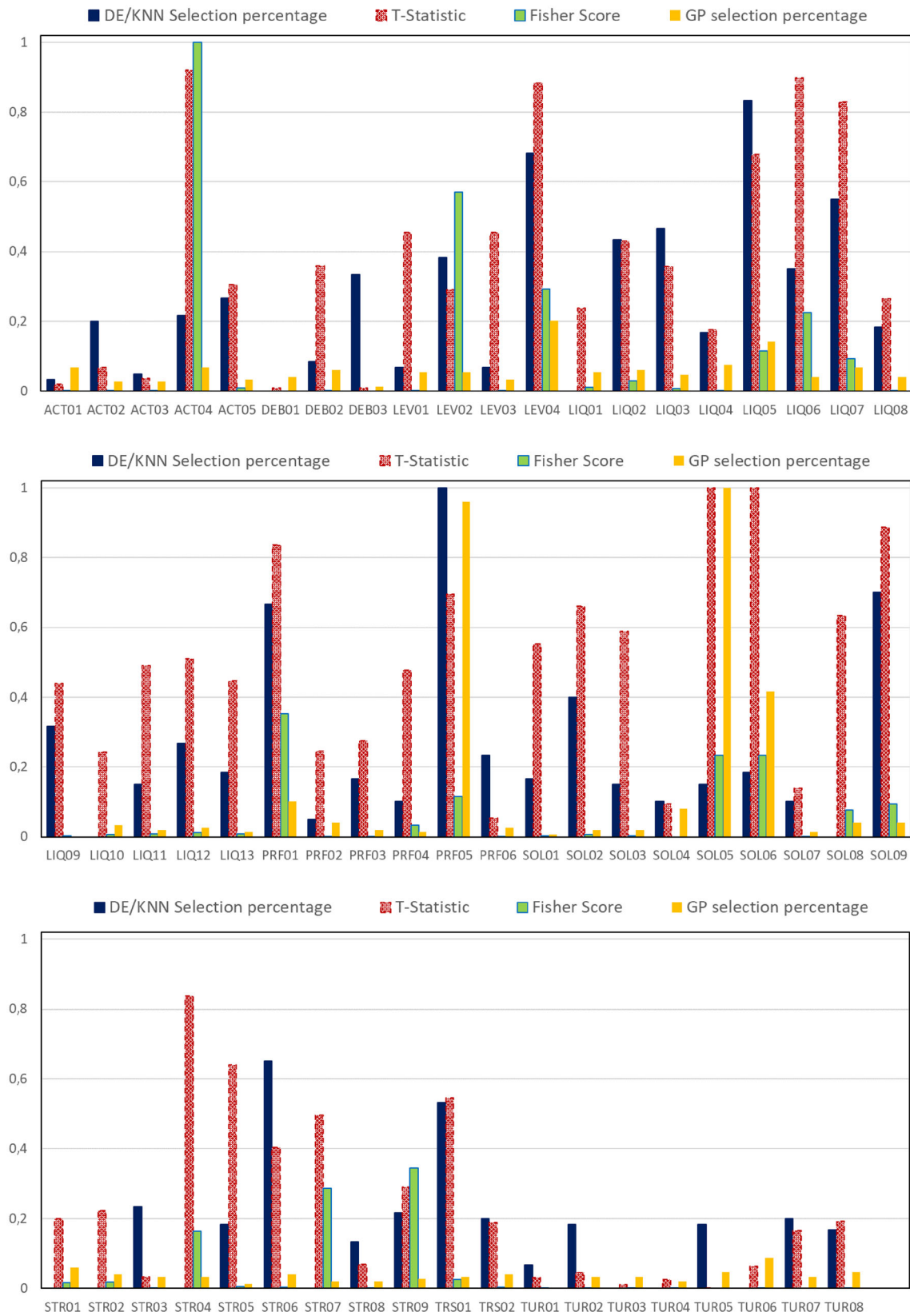


Fig. 1 Percentage of times that a ratio has been selected in the DE/KNN different tests, as well as with GP as feature selector. The T-statistic and Fisher Score values are also included for the 59

financial ratios. All values are normalized in [0,1], considering their maximum and minimum values

Table 2 Classification measures (in percentages) using the complete test set with the three test variants and with an ANN with selected inputs from the DE/KNN approach

Test variant	Accuracy	Sensitivity
T1-30FS (30 best ratios with Fisher Score), 3NN	87.45	80.88
T1-30FS (30 best ratios with Fisher Score), 15NN	92.67	88.23
T2-30TS (30 best ratios with T-statistic), 3NN	89.25	90.44
T2-30TS (30 best ratios with T-statistic), 15NN	91.32	91.92
T3-59 Ratios (starting with all the 59 ratios), 3NN	94.62	82.35
T3-59 Ratios (starting with all the 59 ratios), 15NN	95.72	82.35
ANN, 3 selected ratios	90.35	92.65
ANN, 10 selected ratios	90.41	93.38

Ratios PRF05, LIQ05, PRF01, LEV04 and SOL09 are the five most frequently selected by the EA runs considering all the test variants. In many cases, there is a correlation between the selection of a ratio and the value of the T-statistic or the Fisher Score value. For example, ratios such as ACT01, ACT03, DEB01, TURN03, TURN04 and TUR06 were never selected and also these ratios have low values with the two filtering measures. Few examples, such as STR04 and STR07 have never been selected despite the values in the filters. Finally, ratios such as DEB03, STR03 and TUR05 were sometimes selected but these ratios present very low values in the filtering methods. Therefore, in these latter ratios, although they may not be chosen as relevant by the filtering methods (by themselves), in fact they are and contribute to providing a high value of classification accuracy.

This graph (Fig. 1) obviously can be detailed in the different test variants, but this integration of the ratio selection with the different test variants shows also the capability of the different ratios to provide a correct classification in the KNN classifier. However, it must be taken into account that the graph does not give information regarding whether the ratios provide a high capability for the classification by themselves or in combination with other selected ratios.

4.1.3 Comparison of the different test variants

Table 2 show the results in the three test variants, with the classification results with the (complete) test set. The class “insolvency” corresponds to the class “positive” examples, whereas the class “non-insolvency” (healthy companies) corresponds to the class of “negative” examples. The “sensitivity” measure ($TP/(TP + FN)$, True positive rate) is emphasized here. This charges a value more important than the rest, because the most important

objective in the application is predicting the future insolvency of companies. This measure is improved lowering the number of false negatives (FN), that is, minimizing the number of companies that are predicted not to fail, but actually fail.

As shown in Table 2, as a noteworthy aspect, using the second test variant, when the evolutionary selection starts from the 30 best ratios according to the T-statistic filter, the sensitivity is higher or slightly higher compared to the other two test variants. Moreover, with $K = 15$, which implies a (rather) high value of the number of neighbors considered, the results are higher or slightly higher in all cases.

If no selection is applied to the input variables, the results are worse compared to those shown in Table 2. When the KNN uses the full set of 59 ratios, for example with $K = 3$, accuracy = 83.52% and sensitivity = 78.67%, i.e., worse results compared, for example, to those obtained using the evolutionary selection from the full set of 59 ratios and also using a KNN with $K = 3$ (Table 2, T3-59 Ratios-3NN).

Finally, to test the generalization of the selection process with DE/KNN, in Santos et al. (2022), the selected and relevant ratios were used with a more robust classifier, a classical multilayer perceptron. For this purpose, a large number (10,000) of ANNs (Artificial Neural Networks) with different combinations of neurons in hidden layers were trained. The trained ANNs with the best sensitivity (and accuracy larger than 90%) were selected. That is, the selected ANNs were the most accurate possible in predicting insolvent companies, minimizing false negatives.

Table 2 also shows the classification results with two selected ANNs, with 3 and 10 selected ratios as inputs. The ratios chosen for the inputs of the ANN are the 3/10 most frequently selected ratios considering all the test variants (Fig. 1). The sensitivity results are the highest obtained, but

Table 3 GP parameters

Explanatory variables	The ones corresponding to each selected subset, obtained by the different feature selection methods
Variable transformation	Normalization based on maximum and minimum values of financial ratios
Evaluator	Mean Squared Error (MSE of predicted values with respect to correct values in the training set)
Solution creator	Probabilistic Tree Creator
Symbolic expression tree grammar	Arithmetic functions (+, -, *, /)
Maximum depth	10 (maximum depth of the tree)
Maximum length	100 (maximum length of the symbolic classification model)
Population size	1,500
Maximum generations	100
Crossover	Subtree Swapping Crossover (crossover of subtrees at the crossover point)
Mutation	Multi Symbolic Expression Tree Manipulator (allows different types of mutation)
Mutation probability	15%
Selector	Tournament - Window size 8 (used in mutation and crossover)
Elites	1 (only best solution retained)
Model creator	Accuracy Maximizing Thresholds (the returned solution is the one that uses as classification threshold the one that maximizes the percentage of successes in the training set)

it should be noted that the ANNs were selected with this criterion from a large number of trained ANNs. Anyway, the fact that higher values (sensitivity) are obtained with an ANN using the variables selected from the DE/KNN process, means that the variables selected with DE/KNN are useful also with a different classifier, allowing to obtain better results with the more powerful ANN classifier. In other words, it can be considered as a case of learning transfer, in this case between features selected from different classifiers. Furthermore, these results with the ANNs are slightly better (in terms of sensitivity) compared to ANNs trained with inputs selected from an ANN sensitivity analysis (Yeung et al. 2010) and reported in Beade et al. (2016), also using the same ANN selection from a large number of trained ANNs.

4.2 Results with the GP approach

4.2.1 GP setup

The most relevant parameters (using - where applicable - HeuristicLab nomenclature) together with their values/selections used with GP runs are listed in Table 3. The same parameters are used when GP is used as a feature selection method and when it is used exclusively as a classifier. Some parameters are the usual values set in HeuristicLab

(such as Solution Creator and Model Creator) (Wagner et al. 2014), while others (Tournament window size, Population Size, Generations, Mutation Probability, Maximum Depth and Maximum Length in evolved trees) were experimentally selected to provide solutions with high classification performance over the 1,000 GP independent runs. Likewise, the set of functions was experimentally chosen and with the basic arithmetic functions (addition, subtraction, multiplication and division) solutions with high classification power were obtained.

4.2.2 GP as feature selector and GP classification results with different selected inputs

GP is used both as a feature selection method and as a classifier. The results in Table 4 correspond to the classification performance (accuracy and sensitivity over the complete test set) when GP was used exclusively as a classifier. Of the 1,000 independent GP runs, the results shown in Table 4 correspond to the best solution provided by GP (as a classifier) in terms of AUC, as explained in Sect. 3.2.

This process with 1,000 independent GP runs is the most computationally time-consuming method. For example, using a platform with a 3.60 GHz i7-7700 processor and 16 GB of memory, the independent GP runs require an

Table 4 Classification measures (in percentages) using GP as classifier, with inputs selected from the DE/KNN test variants and GP as feature selection method

Selected inputs - GP as classifier	Accuracy	Sensitivity
Ratios from T1-30FS (30 best ratios with Fisher Score), 3NN	90.66	94.85
Ratios from T1-30FS (30 best ratios with Fisher Score), 15NN	92.61	92.65
Ratios from T2-30TS (30 best ratios with T-statistic), 3NN	91.48	92.65
Ratios from T2-30TS (30 best ratios with T-statistic), 15NN	92.34	92.65
Ratios from T3-59 Ratios, 3NN	93.04	92.65
Ratios from T3-59 Ratios, 15NN	92.30	93.38
Ratios used in ANN-3 Ratios	90.42	95.59
Ratios used in ANN-10 Ratios	91.87	92.65
Best ratios from GP as selector	93.15	94.85
59 ratios (without feature selection)	95.08	90.44

average of 14 h and 12 min (computing times are similar when using GP as a classifier and when using GP as a feature selection method).

The results of the different rows in Table 4 correspond to these measures when the GP classification process (1,000 independent runs) can use the selected inputs provided by different alternatives: the above variants with DE/KNN as well as the use of GP as a feature selection approach. That is, in the latter case, taking into account the relative frequencies of each variable in the evolved trees and in the GP runs that generate the best solutions (5% out of a total of 1,000 independent GP runs), the selected inputs are those with p -value < 0.05 (Sect. 3.2). Figure 1 shows such relative frequencies of selection (normalized in $[0,1]$) of the different ratios with the GP feature selection approach. 10 ratios were finally selected (p -value < 0.05). The three ratios with the highest selection frequencies correspond to SOL05, PRF05 and SOL06, which were also selected by DE/KNN. However, considering the ratio selection frequencies shown in Fig. 1, there is no clear correlation between the GP and the DE/KNN selection processes, when the whole set of ratios is taken into account, and the same can be said with respect to the correlation between the GP selection and the two filtering methods considered.

It should be remembered that these 1,000 GP runs are for this previous feature selection process, and are not related to the subsequent GP runs when GP is used as a classifier. Moreover, Table 4 shows the classification results when the GP classifier uses the whole set of 59 ratios, i.e., without reducing the dimension of the input set.

All the variable selection methods analyzed lead to subsets of ratios that, to a greater or lesser extent, involve

the income statement. For example, the ratios Cash flow/Total assets (LIQ05) and Profit for the year/Total assets (PRF05) appear in a generalized way in the subsets of variables selected by the different methods.

From the classification results in Table 4, the most relevant aspects to be highlighted are the following:

- i. The use of GP as a classifier method improves the results obtained with the other classifier methods (Table 2), offering more balanced solutions. Comparing the results in Tables 4 and 2, in many cases (T1-30FS-3NN, T2-30TS-3NN, T2-30TS-15NN and ANN-3 ratios), both sensitivity and accuracy are superior when GP is used as a classifier. In a few cases, the sensitivity obtained with GP being higher than in the previous study, the accuracy is slightly lower (T1-30FS-15NN, T3-59 Ratios-3NN and T3-59 Ratios-15NN). In one case (ANN-10 ratios), with GP as classifier, the sensitivity is lower and the precision is higher than those obtained with KNN as classifier.

Note even that with very few selected ratios, GP can obtain quite good classifiers. For example, in the case of ANN-3 ratios, using only 3 previously selected ratios in the evolved trees (ratios PRF05, LIQ05 and SOL09), the final classifier selected has accuracy and sensitivity values higher than 90%, with the best value in sensitivity. Note that these 3 ratios relativize the annual evolution (all the numerators correspond to the income statement).

- ii. Using GP as the classifier method, when starting from the ratios selected with GP (as the selector method), the sensitivity and accuracy data are

Table 5 AUC values when GP is used as classifier, with inputs selected from the DE/KNN test variants and GP as feature selection method, as well as the entire set of 59 ratios

Inputs for GP as classifier	Maximum AUC	Average AUC
Ratios from T1-30FS, 3NN	97.03	96.90
Ratios from T1-30FS, 15NN	96.88	96.82
Ratios from T2-30TS, 3NN	96.69	96.68
Ratios from T2-30TS, 15NN	97.29	96.97
Ratios from T3-59 Ratios, 3NN	97.33	97.12
Ratios from T3-59 Ratios, 15NN	97.16	97.05
Ratios used in ANN-3 Ratios	97.25	97.12
Ratios used in ANN-10 Ratios	96.88	96.82
Best ratios from GP as selector	97.88	97.81
59 ratios (without feature selection)	97.80	97.38

The maximum AUC corresponds to the GP solution with the best AUC of 1,000 independent GP runs. The average AUC corresponds to the average AUC of the 5 best GP solutions

superior to those obtained with the other methods, except in the case of “ANN-3 ratios” in terms of sensitivity (but it should be remembered that the ANNs were selected taking into account sensitivity).

Finally, Table 5 includes the results of measuring the performance, obtained by GP as classifier and in terms of AUC, when GP uses the different selected subsets of variables with the previous approaches. Table 5 shows both the maximum AUC obtained from the 1,000 independent GP runs (Sect. 3.2) and the average AUC of the 5 best solutions of those independent GP runs. The most relevant points are the following:

- i. The AUC obtained with the subset defined with GP as selector is clearly superior to that obtained by the rest of the subsets, both in maximum AUC and average AUC.
- ii. The AUC obtained with the total of variables (59 ratios) is superior to that obtained by the rest of the subsets (except the subset of inputs ratios from GP as selector) in both maximum AUC and average AUC. However, it should be noted that previous selection methods performed the selection with KNN taking into account only the accuracy in the definition of fitness, while GP classifiers have been selected with AUC, as a more appropriate performance measure to evaluate evolved trees using their full set of

classification thresholds. This logically biases towards obtaining the best results in terms of AUC.

Figure 2 (upper part) includes an example with an evolved program, corresponding to the best case considering the AUC measure of Table 5, i.e., the best evolved GP solution when using GP as classifier and starting from the set of features obtained with GP as feature selector. The program is shown in the hierarchical format provided by HeuristicsLab. The input variables are denoted in HeuristicsLab as *r59_maxmin_XX*, where “XX” refers to the specific input variable (out of the 59 listed in Sect. 3.3 and Appendix A) and “maxmin” refers to the normalization of the ratios considering the maximum and minimum values of each ratio (taking into account all observations of the training and complete test set). The nodes use only the basic arithmetic functions specified in Sect. 4.2.1. This model uses 5 ratios: ACT04, PRF05, TUR06, SOL05 and SOL06.

HeuristicLab allows a pruning of the evolved programs. To do this, for each node in the tree, HeuristicLab calculates an evaluation of the impact of the node on the output, as well as a substitution value of the node if it is simplified (e.g., a node corresponding to a function is simplified with a constant). The simplification process removes those branches of the tree that do not negatively alter the AUC. Figure 2 (bottom part) shows this final solution after manual simplification.

As previously indicated in the 3-variable model (ANN-3 ratios), all the variable selection methods analyzed led to subsets of ratios that, to a greater or lesser extent, contain ratios involving the income statement. This model (in its two versions, evolved and manually simplified), based on the application of GP for variable selection, is no stranger to this trend. Thus, the numerators of the ratios ACT04, PRF05 and TUR06 refer to annual activity (consistent with the model’s prediction horizon, 1 year prior to failure). Additionally, it includes ratios focused on the level of total indebtedness (SOL05 and SOL06). It is interesting to note that these two ratios (SOL05 and SOL06) are complementary to each other, without any problem for GP, which

handles perfectly the multicollinearity that occurs when the model has highly correlated explanatory variables.

The following final section includes new comments that integrate and discuss all these results.

5 Discussion and conclusions

In this study, classification models were designed to predict the insolvency/non-insolvency of a company in the following year. This prediction could indicate, one year in advance, the insolvency of the company, and could serve as a warning to try to change the economic direction of the

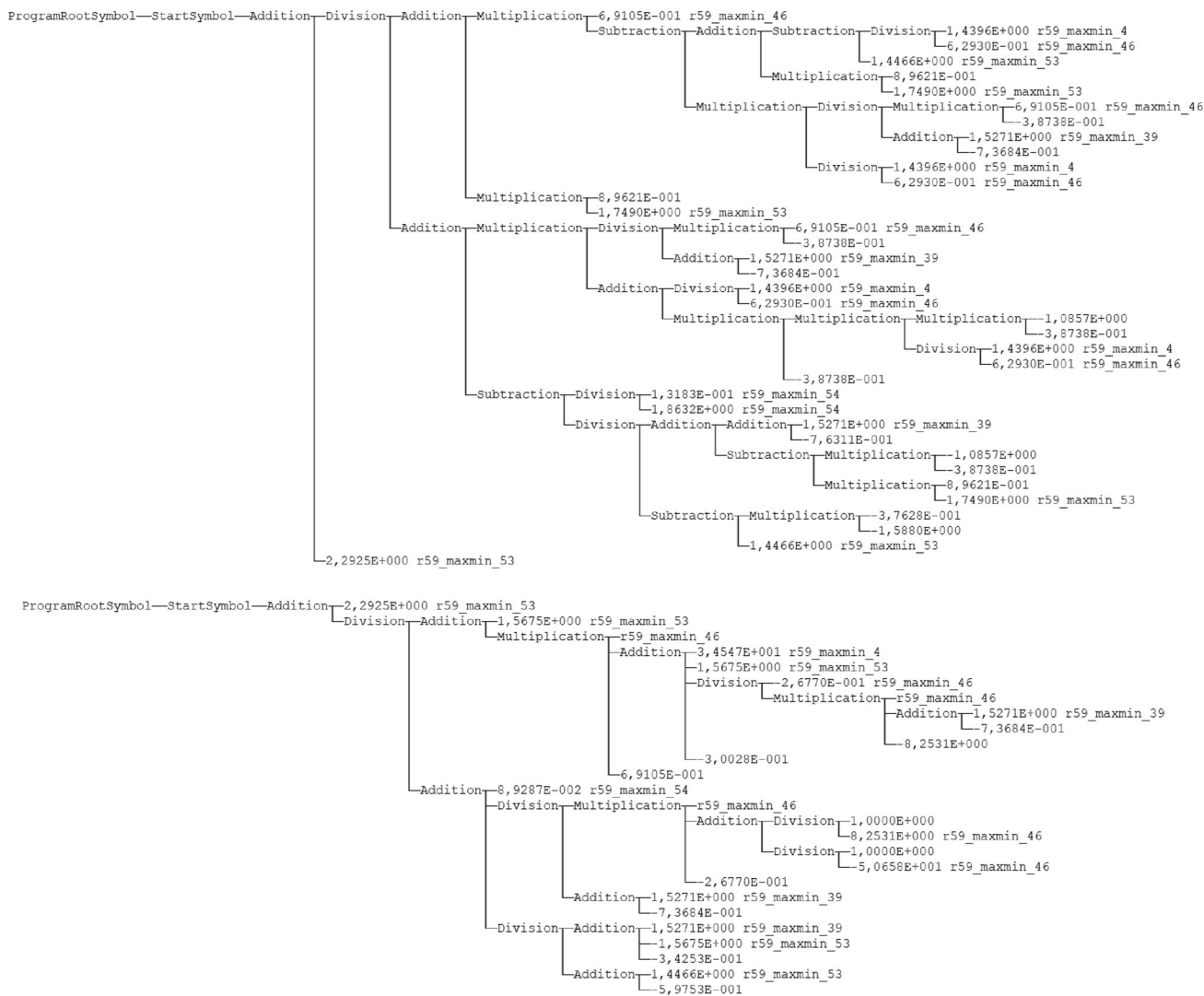


Fig. 2 Upper subfigure: evolved program corresponding to the best solution in terms of AUC, when GP is used as classifier and starting from the set of ratios selected by GP. Bottom subfigure: simplified pruned program

company, before reaching a critical point in the financial situation.

Our study tested genetic programming as an appropriate classification method in this field of business insolvency prediction, together with the use of feature selection methods in the prediction/classification models. In the latter case, the objective was to obtain a significant reduction of the input parameters in such predictive models, without losing the quality of the results obtained. An important point focused on reducing the error of the classifier/predictor in companies classified as non-failed companies that, in reality, end up being insolvent companies (sensitivity measure). That is, the conservative point of view of an investor or a loan officer is one of the most important aspects to consider.

For the feature selection process, two evolutionary approaches were considered. The first is a wrapper method that combines the global DE search, in the huge space of possible subsets of ratios, with a simple KNN classifier for a low computational cost associated with the fitness calculation. In addition, with this first alternative, a prior filtering of ratios based on classical univariate analysis (Fisher Score and T-statistic) was also considered. The second approach uses genetic programming as a novel feature selection method in the field of business insolvency prediction. This second approach is based on the relative frequency of variables in the trees evolved during the GP process.

There is no clear correlation between the sets of ratios selected from these different feature selection strategies. The maximum Pearson correlation is 0.50 between the 2 filters (Fisher Score and T-statistic) when calculating the correlation between each pair of feature selection methods and taking into account the values shown in Fig. 1. This means that there is no subset of ratios with high predictive power by itself, but different subsets can provide high decisional power in the considered classifier method.

With the first approach, the results obtained show a high percentage of classification success in predicting the insolvency one year in advance. Most of the tests performed have a success rate (accuracy) close to 90%. The results achieved by a basic KNN classifier, using the complete set of 59 ratios, are worse than those achieved by using the KNN with the selected ratios by the hybrid DE/KNN feature selection process.

Since the ratios were selected with a simple classifier, the selected ratios were tested using a more powerful

classifier such as a connectionist model. The trained ANNs, with the features selected by the DE/KNN process, were chosen to achieve the best possible results, not so much in terms of the percentage of successes in the global prediction, but in minimizing the failures in which the predicted result of a failed company is contrary to reality (false negatives).

With the second approach of feature selection with GP, as well as GP as a classifier, there are some relevant aspects that can be inferred from the results of this study:

- i. The use of GP as a classifier method improves on the results obtained with KNN and ANNs as classifier methods. Even with performance evaluations at a single classification threshold, the results in Tables 2 and 4 clearly point to that superiority.
- ii. When GP is used as the classifier method, the results obtained with the subset defined with GP as the feature selection method are superior (in most cases) to those obtained by other subsets. This has been shown when the different selection approaches are evaluated by measures such as sensitivity and accuracy (Table 4).
- iii. If GP is used as classifier and AUC as the performance measure (evaluating the totality of classification thresholds, instead of just one), the manifest superiority of the results obtained by the subset defined with GP as feature selection method is observed.

One of our objectives was to test the ability of GP as a classification method in the field of business insolvency prediction, rather than comparing a large number of feature selection methods. Since GP is intended to be used as a classifier, we also introduced the possibility of using GP as a feature selection process. From all the results presented, as a final conclusion, it is worth noting that the performance of the proposed selection method with GP stands out over the other methods analyzed in this study when GP is used as a classifier method. In addition, there is an improvement in the results obtained with the use of GP as a classifier method with respect to the other classifier methods. Therefore, the results of this study are an added value for the use of GP as a classifier method in predicting corporate insolvency, in addition to the interpretability of the prediction models developed with GP.

8 Appendix A

Table 6 Definition of the ratios used

Ratio	Assigned number	Definition
ACT01	1	Financial expenses/Added value
ACT02	2	Personal expenses/Fixed assets
ACT03	3	(Personal expenses + Fixed asset depreciation provision)/Added value
ACT04	4	Operating income/Operating expenses
ACT05	5	Added value/Net sales figure
LEV01	6	Financial expenses/EBIT (Earnings Before Interest and Taxes)
LEV02	7	Financial expenses/Total debt
LEV03	8	Financial expenses/Operating results
LEV04	9	Profit for the year/Total debt
DEB01	10	Total debt/Equity
DEB02	11	(Equity-Profit for the year)/Short-term debt
DEB03	12	Long-term debt/Total debt
STR01	13	Subscribed capital/Equity
STR02	14	Current assets/Total assets
STR03	15	Fixed asset depreciation provision/Fixed assets
STR04	16	Working capital/Total assets
STR05	17	Working capital/Total debt
STR06	18	Working capital/Net sales figure
STR07	19	Other liquid assets/Total assets
STR08	20	Profit for the year/Working capital
STR09	21	Asset decomposition measure
LIQ01	22	Operating cash flow/Total assets
LIQ02	23	Operating cash flow/Total debt
LIQ03	24	Operating cash flow/Short-term debt
LIQ04	25	Operating cash flow/Net sales figure
LIQ05	26	Cash flow/Total assets
LIQ06	27	Cash flow/Total debt
LIQ07	28	Cash flow/Short-term debt
LIQ08	29	Cash flow/Net sales figure
LIQ09	30	Other liquid assets/Short-term debt
LIQ10	31	Inventory/Short-term debt
LIQ11	32	(Inventory+Debtors)/Short-term debt
LIQ12	33	No credit interval
LIQ13	34	Debtors/Short-term debt
PRF01	35	EBIT/Total assets
PRF02	36	EBIT/Net sales figure
PRF03	37	Profit for the year/Net sales figure
PRF04	38	(Profit for the year - Debtors - Inventory)/Net sales figure
PRF05	39	Profit for the year/Total assets
PRF06	40	Profit for the year/Equity
TUR01	41	(Current assets - Inventory)/Net sales figure
TUR02	42	Inventory/Sales
TUR03	43	Net sales figure/Operating income
TUR04	44	Net sales figure/Current assets

Table 6 (continued)

Ratio	Assigned number	Definition
TUR05	45	Net sales figure/Fixed assets
TUR06	46	Net sales figure/Total assets
TUR07	47	Net sales figure/Working capital
TUR08	48	Other liquid assets/Net sales figure
SOL01	49	(Current assets - Inventory)/Short-term debt
SOL02	50	(Current assets)/Total debt
SOL03	51	(Current assets)/Short-term debt
SOL04	52	(Fixed assets)/Equity
SOL05	53	(Total debt)/Total assets
SOL06	54	Equity/Total assets
SOL07	55	Equity/Fixed assets
SOL08	56	Short-term debt/Total assets
SOL09	57	Ordinary results before taxes/Short-term debt
TRS01	58	Treasury/Short-term debt
TRS02	59	Treasury/Net sales figure

Acknowledgements This study was funded by the Xunta de Galicia and the European Union (European Regional Development Fund—Galicia 2014–2020 Program), with Grants CITIC (ED431G 2019/01) and GPC ED431B 2022/33, and by the Spanish Ministry of Science and Innovation (project PID2020-116201GB-I00).

Author's contributions All authors contributed to the conception and design of the study. Material preparation and data collection were carried out mainly by AB. Analysis of the results was performed by all authors. All authors contributed equally to the writing of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This study was funded by the Xunta de Galicia and the European Union (European Regional Development Fund—Galicia 2014–2020 Program), with Grants CITIC (ED431G 2019/01) and GPC ED431B 2022/33, and by the Spanish Ministry of Science and Innovation (project PID2020-116201GB-I00).

Availability of data and materials The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethical approval Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not

included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alaka HA, Oyedele LO, Owolabi HA, Kumar V, Ajayi SO, Akinade OO, Bilal M (2018) Systematic review of bankruptcy prediction models: towards a framework for tool selection. *Exp Syst Appl* 94:164–184
- Alfaro-Cid E, Mora AM, Merelo JJ, Esparcia-Alcázar AI, Sharman K (2009) Finding relevant variables in a financial distress prediction problem using genetic programming and self-organizing maps. In: Anthony B, Michael O (eds) *Natural computing in computational finance: volume 2*, pp 31–49. Springer, Berlin
- Altman EI, Marco G, Varetto F (1994) Corporate distress diagnosis: comparisons using discriminant analysis and neural networks (the Italian experience). *J Bank Finance* 18:505–529
- Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat* 46(3):175–185
- Barnes PA (1987) The analysis and use of financial ratios: a review article. *J Bus Finance Account* 14(4):449–461
- Beade A, Santos J, Rodríguez M (2016) Selección automática de ratios financieros significativos en modelos conexionistas de predicción de la insolvencia empresarial mediante análisis de sensibilidad. XVII Encuentro AECA (Asociación Española de Contabilidad y Administración de Empresas) (138b):1–28
- Brabazon A, Kampouridis M, O'Neill M (2020) Applications of genetic programming to finance and economics: past, present, future. *Genetic Program Evol Mach* 21:33–53
- Chakravarty K, Das D, Sinha A, Konar A (2013) Feature selection by differential evolution algorithm—a case study in personnel identification. In: *Proceedings IEEE congress on evolutionary computation*, pp 892–899

- Chandrashekar G, Sahin F (2014) A survey on feature selection methods. *Comput Electrical Eng* 40(1):16–28
- Das S, Suganthan PN (2011) Differential evolution: a survey of the state-of-the-art. *IEEE Trans Evol Comput* 15(1):4–31
- Doerr B, Zheng W (2020) Working principles of binary differential evolution. *Theor Comput Sci* 801(1):110–142
- Dokeroglu T, Deniz A, Kiziloz HE (2022) A comprehensive survey on recent metaheuristics for feature selection. *Neurocomputing* 494:269–296
- du Jardin P (2009) Bankruptcy prediction models: How to choose the most relevant variables? MPRA paper, University Library of Munich, Germany
- Eltaeib T, Mahmood A (2018) Differential evolution: a survey and analysis. *Appl Sci* 8(10):1945
- Etemadi H, Rostamy AA, Dehkordi HF (2009) A genetic programming model for bankruptcy prediction: empirical evidence from Iran. *Exp Syst Appl* 36(2, Part 2):3199–3207
- Feoktistov V (2006) *Differential evolution: in search of solutions*. Springer, New York
- García-Almanza AL, Alexandrova-Kabadjova B, Martínez-Jaramillo S (2010) Understanding bank failure: a close examination of rules created by genetic programming. In: 2010 IEEE electronics, robotics and automotive mechanics conference, pp 34–39
- HeuristicLab. <https://dev.heuristiclab.com/trac.fcgi/>
- Jayasekera R (2018) Prediction of company failure: past, present and promising directions for the future. *Int Rev Financial Anal* 55:196–208
- Jolliffe IT (2002) *Principal component analysis*, 2nd edn. Springer-Verlag, Berlin
- Koza J (1992) *Genetic programming: on the programming of computers by means of natural selection*. The MIT Press, Cambridge
- Lensberg T, Eilifsen A, McKee TE (2006) Bankruptcy theory development and classification via genetic programming. *Eur J Oper Res*, 169(2):677–697. Feature Cluster on Scatter Search Methods for Optimization
- Liang D, Tsai C-F, Wu H-T (2015) The effect of feature selection on financial distress prediction. *Knowl Based Syst* 73:289–297
- Lin W-C, Lu Y-H, Tsai C-F (2019) Feature selection in single and ensemble learning-based bankruptcy prediction models. *Exp Syst* 36:02
- McKee TE, Lensberg T (2002) Genetic programming and rough sets: a hybrid approach to bankruptcy classification. *Eur J Oper Res* 138(2):436–451. MCDA Methodologies for Classification and Sorting
- Neshatian K, Zhang M (2011) Using genetic programming for context-sensitive feature scoring in classification problems. *Connection Sci* 23(3):183–207
- Papíková L, Papík M (2022) Effects of classification, feature selection, and resampling methods on bankruptcy prediction of small and medium-sized enterprises. *Intell Syst Account Finance Manage*
- Poli R, Langdon WB, McPhee NF (2008) *A field guide to genetic programming*. Lulu Press, Morrisville
- Price KV, Storn RM, Lampinen JA (2005) *Differential evolution. A practical approach to global optimization*. Springer - Natural Computing Series, Berlin
- Iberian Balance Sheet Analysis System (SABI). <https://www.informas.es/en/business-risk/sabi>
- Salcedo-Sanz S, Deprado-Cumplido M, Segovia-Vargas MJ, Pérez-Cruz F, Bousoño-Calzón C (2004) Feature selection methods involving support vector machines for prediction of insolvency in non-life insurance companies. *Intell Syst Account Finance Manage* 12:261–281
- Salcedo-Sanz S, Fernández-Villacañas JL, Segovia-Vargas MJ, Bousoño-Calzón C (2005) Genetic programming for the prediction of insolvency in non-life insurance companies. *Comput Oper Res* 32:749–765
- Santos J, Sestayo O, Beade A, Rodríguez M (2022) Automatic selection of financial ratios by means of differential evolution and for predicting business insolvency. In: *Proceedings of the IWINAC 2022, Lecture Notes in Computer Science*, pp 534–544
- Serrano-Silva YO, Villuendas-Rey Y, Yáñez-Márquez C (2018) Automatic feature weighting for improving financial decision support systems. *Decis Support Syst* 107:78–87
- Tsai C-F, Sue K-L, Hu Y-H, Chiu A (2021) Combining feature selection, instance selection, and ensemble classification techniques for improved financial distress prediction. *J Bus Res* 130:200–209
- Viegas F, Rocha L, Gonçalves M, Mourão F, Sá G, Salles T, Andrade G, Sandin I (2018) A genetic programming approach for feature selection in highly dimensional skewed data. *Neurocomputing* 273:554–569
- Volkov A, Benoit DF, Van den Poel D (2017) Incorporating sequential information in bankruptcy prediction with predictors based on Markov for discrimination. *Decis Support Syst* 98:59–68
- Wagner S, Kronberger G, Beham A, Kommenda M, Scheibenpflug A, Pitzer E, Vonolfen S, Kofler M, Winkler S, Dorfer V, Affenzeller M (2014) Architecture and design of the heuristiclab optimization environment. In: Klempous Ryszard, Nikodem Jan, Jacak Witold, Chaczko Zenon (eds) *Advanced methods and applications in computational intelligence*. Springer International Publishing, Heidelberg, pp 197–261
- Xue B, Zhang M, Browne WN, Yao X (2016) A survey on evolutionary computation approaches to feature selection. *IEEE Trans Evol Comput* 20(4):606–626
- Yeung DS, Cloete I, Shi D, Ng WWY (2010) Sensitivity analysis for neural networks. *Natural Comput Ser* 2010th Edn
- Zhao XS, Bao LL, Ning Q, Ji JC, Zhao XW (2018) An improved binary differential evolution algorithm for feature selection in molecular signatures. *Mol Inf* 37(4):e1700081

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.