Journal of Applied Finance & Banking, vol. 2, no. 6, 2012, 73-81
ISSN: 1792-6580 (print version), 1792-6599 (online)
Scienpress Ltd, 2012

Using Ants to Detect Fraudulent Financial Statements

Christos D. Katsis[1], Yorgos Goletsis[2], Paraskevi V. Boufounou[3], George Stylios[4] and Evangelos Koumanakos[5]

Abstract

Fraudulent financial reporting is a matter of great social and economic concern. Managers may distort financial statements so as to present their companies more favorably to investors or creditors. On the other hand, auditors are the ones who are expected to identify fraudulent financial statements; however several limitations on auditors' performance are reported. Therefore, decision support systems that can support auditor's characterization and identify fraudulent statements can be of great benefit for the auditor himself but also for all stakeholders. In this work, we investigate the use of a swarm intelligence technique, mimicking the functions of natural ants for fraudulent financial statements detection. More specific, an Ant miner classifier is applied on a real dataset arising from the Athens Stock Exchange firms. The method is compared against commonly used statistical and machine learning classifiers. The obtained results are promising. Apart from the high classification accuracy, Ant miner offers interpretation ability, by extracting compact rules regarding the decisions made, which is a further advantage.

[1] Department of Applications of Information Technology in Administration and Economy, Technological Educational Institute of Ionian Islands, GR31100, Lefkada, Greece,
e-mail: ckatsis@teiion.gr

[2] Department of Economics, University of Ioannina, GR45110 Ioannina, Greece,
e-mail: ggoletsis@cc.uoi.gr

[3] Department of Business Administration, Technological Educational Institute of Ionian Islands, GR28100, Cephalonia, Greece,
e-mail: viv.boufounou@gmail.com

[4] Department of Applications of Information Technology in Administration and Economy, Technological Educational Institute of Ionian Islands, GR31100, Lefkada, Greece,
e-mail: gstylios@yahoo.gr

[5] Department of Economics, University of Ioannina, GR45110 Ioannina, Greece,
e-mail: ekoumanakos@cc.uoi.gr

Article Info: *Received :* June 18, 2012. *Revised :* September 21, 2012.
Published online : December 20, 2012

JEL classification numbers: C45, C53, C57, M42, H83
Keywords: Ant Miner, Artificial intelligence, Classification, Fraudulent financial statements, Machine learning

1 Introduction

Fraudulent financial reporting is a matter of grave social and economic concern. ‘’Fraud’’ can be defined as an intentional act to deceive or mislead another party [1]. Top management fraud usually takes the form of falsified financial statements aiming mostly to present their companies more favorably to investors or creditors. All related stakeholders are affected from such an action (investors, shareholders, creditors, employees). The identification of fraudulent financial statements (FFS) is a major task for auditors. Decision Support Systems that can support auditors characterization can significantly improve the identification of FFS.

Looking at the problem as a classification problem, literature review reveals that financial ratios analysis along with statistical based methods (e.g. logistic regression) have been mostly applied. These include Naive Bayes [2], logistic regression ([3], [4], [5]) probit [6], multi-criteria regression [7]. Artificial intelligence methods and mainly Neural Networks [8], [9], [10], [5] have been applied providing high results, usually higher than purely statistical methods. Support Vector Machines (SVMs) have been also considered as promising methods [9]. Still, both Neural Networks and SVM do not provide an justification of the classification decision and operate as ‘black boxes’. Decision trees, that tackle this, have also been applied [10], with less good results.

More recently nature inspired techniques that mimic the behavior of natural species have demonstrated high results in other data classification problems. Their main concept relies on the evolution principles of real-life species. Ant Colony Systems (ACS) [11] imitate the behaviour of natural ants. They are usually applied for optimisation and feature selection prior to the application of classification algorithms. ACS have been applied, but still in a limited level, to other classification problems, such as medical decision making [12]. An advantage of Ants based classification is rule extraction. Such a feature can offer interpretation of the classification outcome.

In the financial domain, ants based classification has been recently applied to the bankruptcy prediction problem [13] and to credit scoring [14] with overall good results, compared to the ones obtained by other state-of-the-art classifiers. The goal of this paper is therefore to investigate the applicability ants based classification into the FFS problem. This is done over a real dataset containing data from Greek manufacturing firms listed on the Athens Stock Exchange.

2 Methodology

Biologically inspired techniques rely on the evolution principles of real-life species by mimicking the way biological entities communicate or compete to ensure their survival. An artificial Ant Colony System (ACS) [15] is an algorithm based on agents that simulate the natural behaviour of ants, developing mechanisms of cooperation and learning. Ants live in colonies and although individual ants can do only simple tasks on their own, it is the colony’s cooperative work that allows the colony to demonstrate an intelligent behaviour. Ants communicate with each other due to trails of a substance called pheromone. While ants move, they drop a certain amount of pheromone on the ground, leaving behind a trail of this substance that can be followed by other ants. The more ants follow a pheromone trail, the more attractive the trail becomes to be followed in the near future. This is a kind of autocatalytic behaviour, described by a loop of positive feedback, where the probability of an ant choosing a path increases directly with the number of ants that have passed in the path before, since ants prefer paths with a larger amount of pheromone. Consequently, shorter paths tend to have a high rate of growth in the amount of pheromone. Moreover, because of pheromone evaporation, pheromone on the longer paths vanishes with time. In order to simulate this process, an ACS algorithm performs a loop applying two basic procedures: (i) a procedure specifying how ants construct or modify a solution for the problem in hand and (ii) a procedure for updating the pheromone trail. The construction or modification of a solution is performed in a probabilistic way. The probability of adding a new item to the solution under construction is, in turn, a function of a problem-dependent heuristic and the amount of pheromone previously deposited in this trail. The pheromone trails are updated considering the evaporation rate and the quality of the current solution. Therefore, a practical implementation of an ACS includes the following definitions [16]:

- An appropriate representation for the problem with which ants can incrementally construct or modify solutions, by means of a probabilistic transition rule based on the amount of pheromone in the trail and on a local heuristic;
- A heuristic function that measures the quality of the items that can be added to the current partial solution;
- A method to enforce the construction of valid solutions (in the real-world search space);
- A rule that specifies how a pheromone trail should be updated;
- A probabilistic transition rule that uses the current value of the heuristic function and the current amount of pheromone in the trail.

In the rule extraction case the ACS is used to construct / discover classification rules in the form '*IF (term1 AND term2 AND ...) THEN (predicted class)*', where each term in the form <attribute=value>. In this framework, Ant-Miner has been developed as an ACS to extract classification rules with promising results in benchmark datasets [12, 17].

Ant-Miner discovers an ordered list of classification rules based on a heuristic function involving information gain – as a heuristic function– and positive feedback involving artificial pheromone. Each ant starts with a rule with no term in its antecedent (empty rule) and adds one term at a time to its current partial rule. The current partial rule constructed by an ant corresponds to the current partial path followed by that ant. Similarly, the choice of a term to be added to the current partial rule corresponds to the choice of direction for which the current path will be extended, among all the possible directions (all terms that could be added to the current partial rule). The choice of a term (attribute-value pair) to be added depends on both, a problem-dependent heuristic function and on the amount of pheromone associated with each term. For each iteration of the Repeat-Until loop, an ant attempts to discover a rule by selecting terms in a probabilistic manner, until all the attributes have been used to make the current rule, or adding any other available term would make the rule coverage less than the *min_cases_per_rule* parameter. The discovered rule is then pruned in an attempt to reduce

overfitting to the training data and increase rule quality. Afterwards, the pheromone values for the terms in the current rule are increased, in order to increase the probability that other ants will select those terms, and then the pheromone values for all terms are normalised. The While loop iterates until the number of training examples remaining in the dataset becomes less than or equal to *Max_uncovered_cases* parameter. The rule discovered in the Repeat-Until loop that has the highest quality is then added to the list of discovered rules, and the training examples correctly covered by that rule are removed from the training dataset. An example is correctly covered by a rule if the example satisfies the rule antecedent and has the class predicted by the rule.
The algorithm employees four user specified parameters, as provided in Table 1:

Table 1: User specified parameters employed in the Ant Miner

Parameter	Parameter description
No_of_ants	The maximum number of iterations the Repeat-Until loop of the algorithm will perform
Min_instances_per_rule	During the construction phase, a term cannot be added to the current partial rule if this would make the number of covered cases fall below this number. The algorithm thereby excludes possible solutions which cover very few cases.
Max_uncovered_instances	The point at which rule discovery will stop. Each iteration of the While Loop reduces the size of the training set and the number of cases will finally fall below this number.
No_rules_converg	The alternative stopping condition for the Repeat-Until loop is if this number of identical rules is produced in sequence by the ants.

For the interested reader, a more detailed description of the Ant Miner algorithm can be found in [13, 15].

3 Dataset

Our initial dataset derives from an earlier study [18] which consists of financial variables arriving from 277 Greek manufacturing firms listed on the Athens Stock Exchange. All the firms in the sample were checked by auditors. For 50 of these firms, indication or proof of involvement in issuing false financial statement (FFS) was published. These FFS firms were matched with 227 non-FFS firms. The variables used in the sample were extracted from formal financial statements, such as balance sheets and income statements. The selection of variables to be used as for the automated financial statement characterization was based upon prior research work, linked to the topic of FFS [18-21]. Financial statements were classified as false according to the following parameters:

- Inclusion in the auditors' report of serious doubts on to the accuracy of the accounts, observations by the tax authorities regarding serious taxation intransigencies which significantly altered the company's annual balance sheet and income statement.
- The application of Greek legislation regarding negative net worth.
- The inclusion of the company in the Athens Stock Exchange categories of "under

observation and “negotiation suspended” for reasons associated with the falsification of the company’s financial data.

- The existence of court proceedings pending with respect to FFS or serious taxation contraventions.

The financial variables used in the chapter are shown in Table 2.

Table 2: Financial variables extracted from formal financial statements used as FFS predictors

financial variable #	financial variables
var.1	Return on Long -term capital / Return on Capital and Reserves
var.2	Accounts Receivable/Total Assets –previous year
var.3	Total liabilities/Total Assets
var.4	Accounts Receivable/Total Assets
var.5	Working capital/Total Assets
var.6	Deposits and cash/Current Assets
var.7	Net Fixed Assets/Total Assets
var.8	Number of days accounts payable
var.8	Long term debt/Total Capital and Reserves
var.9	Sales/Total Assets
var.10	Results carried forward/Total Assets
var.11	Number of days accounts receivable
var.12	Change Accounts Receivable/Total Assets
var.13	Working capital leveraged
var.14	Inventory turnover
var.15	Total Assets/Capital and Reserves
var.16	Earnings before interest and tax/Total Assets
var.17	Cash flows from operations
var.18	Cash flows from operations –previous year
var.19	Current assets to current liabilities
var.20	Growth of Operational Cash Flow
var.21	Change Accounts Receivable/Net Sales
var.22	Altman z-score
var.23	Capital and Reserves/total liabilities

To deal with the class imbalance problem of our dataset, the Synthetic Minority Over-Sampling TEchnique (SMOTE) [22] is applied. SMOTE generates synthetic minority examples by interpolating between several minority class examples that lie together to over-sample the minority class. For every minority example, its *k* nearest neighbors of the same class are calculated and some examples are randomly selected from them according to the over-sampling rate. After that, new synthetic examples are generated along the line between the minority example and its selected nearest neighbors. Thus, the overfitting problem is avoided and causes the decision boundaries for the minority class to spread further into the majority class space. SMOTE improves the accuracy of the classifiers for the minority class as it forces focused learning and introduces a bias towards the minority class. A commonly used alternative is oversampling by replication of the minority class. However, in that case, the decision region for the minority class becomes smaller, which is the opposite of the desired effect [23]. After applying the SMOTE technique, a modified, balanced dataset arises, containing 225 firms with FFS out of a total of 452 firms.

In the following section, the applicability and the performance of the Ant Miner classifier is evaluated on the identification of fraudulent financial statements problem. Also, the classification accuracy is compared with commonly applied machine learning and statistical techniques.

4 Results

The Ant-miner classifier has been applied in the dataset described in the previous section. The parameters of the classifier are set heuristically. More specifically, the number of ants is set to five, the number of consecutive rules required for convergence to five, the minimum instances per rule (support) to ten instances and the maximum number of uncovered cases to five. Since the Ant Miner algorithm can handle only nominal data, a disretisation step is applied. More specifically, equal internal binning (# of bins is set to five) is applied.

The obtained accuracy of the Ant Miner classifier is compared to statistical classifiers, namely Quadratic Discriminant Analysis (QDA), logistic regression and Naïve Bayes, and common machine learning classifiers as Multilayer Perceptron Artificial Neural Networks (MLP-ANNs) and C4.5 decision trees algorithm. The experimental results in terms of accuracy are provided in Table 3.

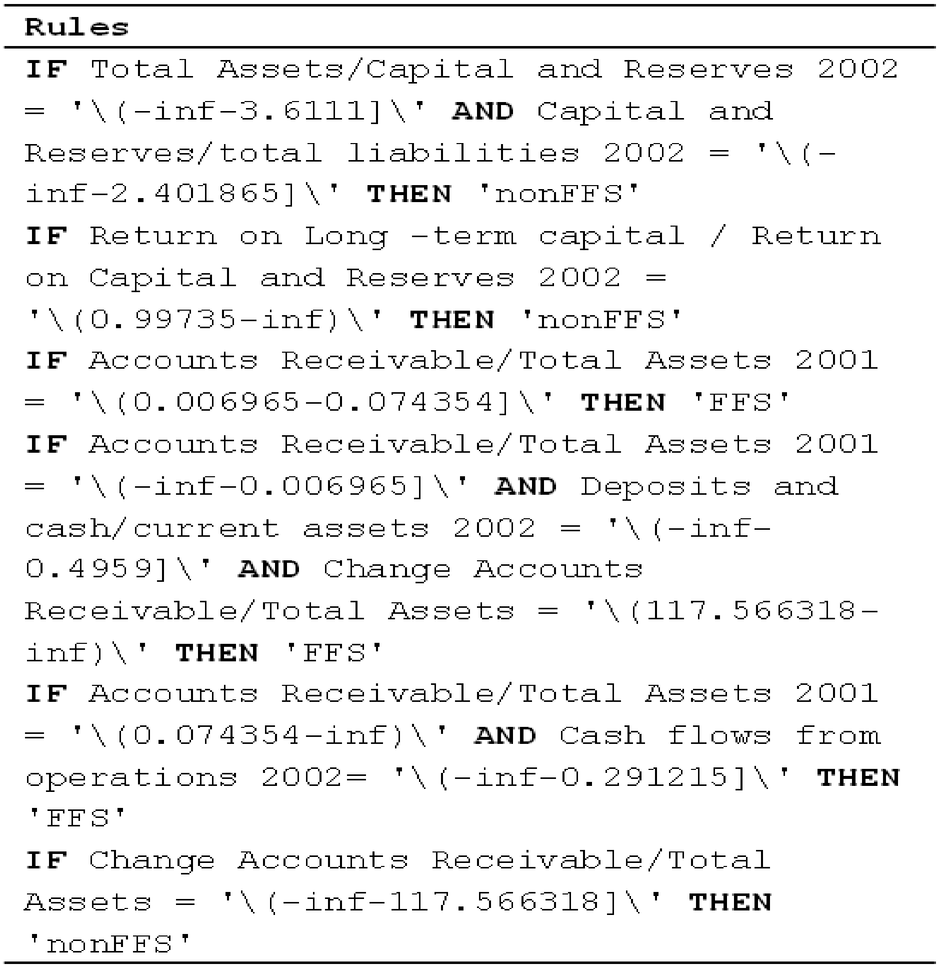
Table 3. Comparative results in terms of accuracy

Classifier	Accuracy (%)
Ant miner	77.97
QDA	72.56
Logistic regression	74.28
Naïve Bayes	65.31
MLP-ANN	74.77
C4.5	70.35

To minimize the bias associated with the random sampling of the training and testing data samples, ten-fold cross-validation is applied. As we can notice from Table 2 the overall classification accuracy achieved by the Ant Miner is superior to the classic machine learning and statistical techniques.

Apart from the high classification accuracy, the rule extraction ability is an advantage since it provides interpretation regarding the classification decision. More specific, the Ant-Miner classifier produces an average of seven rules. The fact that the extracted rule sets are rather compact is an additional appealing characteristic. Such a rule set can be more comprehensible to the user/auditor. Indicate rule sets generated using Ant-Miner algorithm in the full featured dataset are depicted in Figure 1.

Figure 1. Rules extracted after applying the Ant Miner.



Rules
IF Total Assets/Capital and Reserves 2002 = '\(-inf-3.6111]\' **AND** Capital and Reserves/total liabilities 2002 = '\(-inf-2.401865]\' **THEN** 'nonFFS'
IF Return on Long -term capital / Return on Capital and Reserves 2002 = '\(0.99735-inf)\' **THEN** 'nonFFS'
IF Accounts Receivable/Total Assets 2001 = '\(0.006965-0.074354]\' **THEN** 'FFS'
IF Accounts Receivable/Total Assets 2001 = '\(-inf-0.006965]\' **AND** Deposits and cash/current assets 2002 = '\(-inf-0.4959]\' **AND** Change Accounts Receivable/Total Assets = '\(117.566318-inf)\' **THEN** 'FFS'
IF Accounts Receivable/Total Assets 2001 = '\(0.074354-inf)\' **AND** Cash flows from operations 2002= '\(-inf-0.291215]\' **THEN** 'FFS'
IF Change Accounts Receivable/Total Assets = '\(-inf-117.566318]\' **THEN** 'nonFFS'

.

Still, it must be noted that these rules correspond to the specific dataset; the creation of a universal rule-set would require large training datasets.

4 Discussion

The applicability of the Ant Miner classifier to the fraudulent financial statement detection problem has been examined. The proposed algorithm seems to perform satisfactorily in terms of accuracy. Moreover, apart from the quite high classification accuracy, the rule extraction ability and the fact that the extracted rule sets are rather compact constitute an additional appealing characteristic. Thus, the Ant Miner should be considered as promising classifier for the domain and could become a part of a decision support system for auditors.

References

[1] A. A. Arens and J. K. Loebbecke, *Auditing: An Integrated Approach*, 1997.

[2] D. Qingshan, "Detection of fraudulent financial statements based on Na&#x00EF;ve Bayes classifier," in *Computer Science and Education (ICCSE), 2010 5th International Conference on*, 2010, 1032-1035.

[3] Y.-C. Tseng and R.-D. Chang, "Effects Of A Decision Aid For The Assessment Of Fraudulent Financial Reporting: An Application Of SAS No. 82," *Journal of Business and Economic Resesarch*, **4**, 2006.

[4] O. S. Persons, "Using financial statement data to identify factors associated with fraudulent financial reporting," *Journal of Applied Business Research*, **11**, 38-46, 1995.

[5] T. B. Bell and J. V. Carcello, "A decision aid for assessing the likelihood of fraudulent financial reporting," *Auditing*, **19**, 168-184, 2000.

[6] M. D. Beneish, "Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance," *Journal of Accounting and Public Policy*, **16**, 271-309, 1997.

[7] C. Spathis*, et al.*, "Detecting falsified financial statements: A comparative study using multicriteria analysis and multivariate statistical techniques," *The European Accounting Review*, **11**, 509-535, 2002.

[8] M. J. Cerullo and V. Cerullo, "Using neural networks to predict financial reporting fraud: Part 1," *Computer Fraud and Security*, **1999**, 14-17, 1999.

[9] H. Öğüt*, et al.*, "Prediction of financial information manipulation by using support vector machine and probabilistic neural network," *Expert Systems with Applications*, **36**, 5419-5423, 2009.

[10] E. Kirkos*, et al.*, "Data Mining techniques for the detection of fraudulent financial statements," *Expert Systems with Applications*, **32**, 995-1003, 2007.

[11] M. Dorigo*, et al.*, "Ant system: Optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, **26**, 29-41, 1996.

[12] R. Parpinelli*, et al.*, "An Ant Colony Based System for Data Mining: Applications to Medical Data," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, 2001, 791-797.

[13] Y. Goletsis*, et al.*, "Can Ants predict bankruptcy? A comparison of Ant Colony Systems to other state-of-the-art computational methods," *New Mathematics and Natural Computation*, **5**, 571-588, 2009.

[14] Y. Goletsis*, et al.*, "Credit scoring using an Ant mining approach," *Human Systems Management*, **29**, 79-88, 2010.

[15] M. Dorigo*, et al.*, "Ant system: Optimization by a colony of cooperating agents," *Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics*, **26**, 29-41, Feb 1996.

[16] E. Bonabeau*, et al.*, *Swarm Intelligence: From Natural to Artificial Systems*. New York: New York: Oxford Univ. Press, 1999.

[17] R. S. Parpinelli*, et al.*, "Data mining with an ant colony optimization algorithm," *Ieee Transactions on Evolutionary Computation*, **6**, 321-332, Aug 2002.

[18] S. Kotsiantis*, et al.*, "Predicting fraudulent financial statements with machine learning techniques," *Advances in Artificial Intelligence, Proceedings*, **3955**, 538-542, 2006.

[19] C. Spathis, " Detecting false financial statements using published data: some evidence from Greece," *Managerial Auditing Journal* **17**, 13, 2002.

[20] C. Spathis, *et al.*, "Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques," *The European Accounting Review* **11**, 26, 2002.

[21] K. Fanning and K. Cogger, "Neural Network Detection of Management Fraud Using Published Financial Data," *International Journal of Intelligent Systems in Accounting, Finance & Management*, .**7**, 4, 1998.

[22] N. V. Chawla, *et al.*, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, **16**, 321-357, 2002.

[23] H. Han, *et al.*, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," *Advances in Intelligent Computing, Pt 1, Proceedings,* **3644**, 878-887, 2005.