Artificial Bee Colony Algorithm for Feature Selection in fraud detection process*

Gabriel Covello Furlanetto[1][0000−0003−2917−9182], Vitória Zanon Gomes[2][0000−0003−4176−566X], and Fabrício Aparecido Breve[1][0000−0002−1123−9784]

[1] Department of Statistics, Applied Mathematics and Computer Science / Universidade Estadual Paulista (UNESP) / Avenida 24A, 1515 - Jardim Bela Vista / Rio Claro-SP, 13506-900, Brazil
[2] Department of Computer Science and Statistics / Universidade Estadual Paulista (UNESP) / Rua Cristóvão Colombo, 2265 - Jardim Nazareth / São José do Rio Preto-SP, 15054-000,Brazil
{gabriel.furlanetto,vitoria.zanon,fabricio.breve}@unesp.br

Abstract. More and more, nowadays, better performance and quality of current classifiers are required when the topic is fraud detection. In this context, processes such as feature selection help to increase the quality of the results obtained by the existing classifiers in the literature, since the high dimensionality of current datasets and redundant information significantly affect the performance of these techniques. This work proposes a wrapper method of feature selection using the ABC algorithm combined with Logistic Regression classification, seeking to obtain better results for fraud detection. The method proposed also contributes to defining the optimal parameters of other feature selection algorithms. Through the tests performed and the results obtained, it is possible to confirm the quality of the method, achieving the proposed objective.

Keywords: Artificial Bee Colony · Feature Selection · Fraud detection · Machine Learning.

1 Introduction

Fraud detection is no longer an option for companies nowadays [15]. With the growing increase in banking operations via smartphones and the internet, also driven by the pandemic reported in the year 2019, fraud multiplied, making the investment, in efficient fraud detection systems, vital for the survival of companies and, principally for financial institutions, since these malicious activities not only cause harm but also cause distrust on the part of customers.

At the same time, the use of Machine Learning techniques for this purpose has been growing gradually in the literature, and it is possible to find several

* Vitoria Zanon Gomes has received research support from Coordination for the Improvement of Higher Education Personnel (CAPES) under grant number 88887.686064/2022-00.

approaches, ranging from supervised learning through classical algorithms to the hybridization of techniques seeking better performance in unbalanced datasets [4].

However, despite the remarkable role of these techniques, the information itself, to be classified, can affect the performance of the algorithm. Datasets that gather information about financial transactions, tend to have a huge amount of attributes, often irrelevant, redundant, or highly correlated [17], making the classifier expend time looking for patterns and correlations that will not bring significant gain, only consuming time and computational resources. To solve this problem, the process known as Feature Selection is used.

The present work explores the Feature Selection process in order, not only, to reduce the complexity of the supplied attributes delivered to the classifier, but also to improve classifier performance by reducing attributes. For this, the Artificial Bee Colony (ABC) algorithm or Bee Swarm Algorithm (as found in the literature) was used, to perform the attribute selection, taking into account its simplicity, jointly with the classifier through Logistic Regression in a scenario of detection of frauds.

Furthermore, as another contribution, this work doesn't use the ABC algorithm only as a tool for feature selection, but also to find the optimal parameter and feed other algorithms whose goal is the same, such as K-best and RFE, for example. Unlike other works found in the literature that evaluate results using only one classifier algorithm this work also proposed test scenarios in which the performance of three different classification algorithms (Logistic Regression, Random Forest and Gradient Boosting) are evaluated with the optimal features. As a result, it is clear that ABC is stable as a feature selection tool and as a good option to define the optimal parameter for K-best.

For this purpose, this article was divided as follows: in section 2 the attribute selection method is detailed. In section 3, a generic approach to particle swarm algorithms is made, followed by the explanation of the ABC algorithm, which is the subject of this article. Works related to the proposed one are presented in section 4. In section 5, the implementation of the proposed method is described. After implementation, the tests scenarios and their results are described in section 6, and finally, in section 7 final considerations are made about the work.

2 Feature Selection

Feature Selection consists of the selection of features of a dataset, seeking to maximize the performance of classifiers by using only selected attributes for the classification process, leading to a reduction in complexity of the dataset, in addition to optimizing the accuracy of the method [17]. This selection can be made through several methods, which are divided into two categories:

– **Filter methods:** The methods of this category work without considering the classifier to be used. In them, the attributes of the dataset are analyzed individually and collectively, and the statistical data extracted from these

analysis help the method to define which are the most relevant characteristics for a good result of classification. This type of method is widely used due to its simplicity of implementation and little use of computational resources [17] [7].
- **Wrapper methods:** In this class of methods, the classifier acts as a kind of black box, being part of an objective function used to evaluate the various possible combinations of attributes. The main problem of this approach is in dealing with sets of data with high dimensionality, since the computational performance decreases significantly, although it does not make its use unfeasible [7].

Machine learning algorithms and optimization algorithms are widely used for this purpose. In this work, the optimization algorithms based on population will be highlighted, with the selection of attributes by wrapping methods.

3 Population-based Algorithms

Population-based algorithms are a subgroup within the class of bioinspired algorithms. As the name suggests, its performance is based on the behavior of species that live in society. Examples include the Genetic Algorithm [14] and algorithms based on swarm intelligence, such as Particle Swarm Optimization (PSO) [6], Ant Colony Optimization (ACO) [5], and Artificial Bee Colony Algorithm (ABC) [11]. Among those mentioned, ABC stands out for its simplicity, robustness, and ease of implementation. Besides, it still needs fewer input parameters to execute, in comparison with the others and it can be easily combined with other algorithms to obtain better performance.

3.1 Artificial Bee Colony Algorithm (ABC)

The ABC algorithm simulates the work of honey bees throughout the foraging process, that is, the search for food for the members of the hive [12].

Biological Behavior: The worker bees are responsible for all the maintenance of the hive, including the search for food. To this end, the workers are divided into three groups: Employees, Onlookers, and Scouts [12].

Scout bees, as the name suggests, are responsible for randomly looking for food sources in the vicinity of the hive. When they meet, they become employed bees. After choosing the food source, the bees assess its quality through factors such as the distance from the hive and the difficulty in extracting the nectar, and return to the hive with a sample of the food found [12]. Upon returning, the employed bees pass the information about the source and the sample of the nectar to the onlooker bees, who will summarize and evaluate the information brought by the employees, in a way to decide which source should be exploited [2].

The Algorithm: In the Bee Swarm Algorithm, food sources represent possible solutions to the problem to be solved, and their several characteristics to be analyzed by the bees are replaced by only one: an objective function. The bee's work is performed by routines executed iteratively until a stopping condition is reached, and then the best solution found is returned. This work can be described by the sequence of steps of the algorithm:

1. **Population initialization:** The population of n scout bees is randomly initialized, making these bees employed;
2. **Phase 1 - Employed bees:** The score for each of the current food sources is calculated, as well as the probability of them being chosen by an onlooker bee, considering their quality in relation to the others ;
3. **Phase 2 - Onlooker bees:** The n onlooker bees will choose, taking into account the previously calculated probability, food sources to be exploited, that is, which will undergo small modifications to try to improve your score. If it is possible to improve any of the solutions, this new solution is saved;
4. **Phase 3 - Verification of Stagnation (scout bees):** If any food source has reached its stagnation limit, that is, it is at a pre-defined of unimproved iterations, this source is abandoned, and the bee responsible for it becomes an explorer again, randomly choosing a new food source to evaluate;
5. At the end of the iteration, the best score is saved;
6. If the stop condition is met, the execution ends. Otherwise, it continues from item 2.

The sequence of steps described can also be seen, in the form of a flowchart, in Figure 1. In it, the order of execution of the algorithm can be seen.

4 Related Works

When searching the literature, other works with the same application of attribute selection with ABC can be find. Among them, Pavithra and Thangadurai [17], implement the ABC together with the Support Vector Machine (SVM) classifier to perform feature selection in the fraud detection scenario. In this work, the tests were performed with a dataset with a relatively low number of attributes, but positive results were obtained in relation to the classification without selection of attributes. Despite the little dimensionality of the test cases, this work supports the efficiency of ABC in fraud detection cases.

Meanwhile, in the work of Hancer, Xue, Karaboga, and Zhang [9], ABC is used as a filtering method. The objective function of the algorithm seeks to evaluate the proximity relationship between the attributes and, therefore, the objective of the algorithm becomes to determine a set of attributes where their proximity is as close as possible using the kNN classifier. In this work, the tests were performed with several datasets available for free in the machine learning repository of the University of California (UCI), and the results obtained validated the effectiveness of ABC as a filter method, which results in a better performance in the selection process.

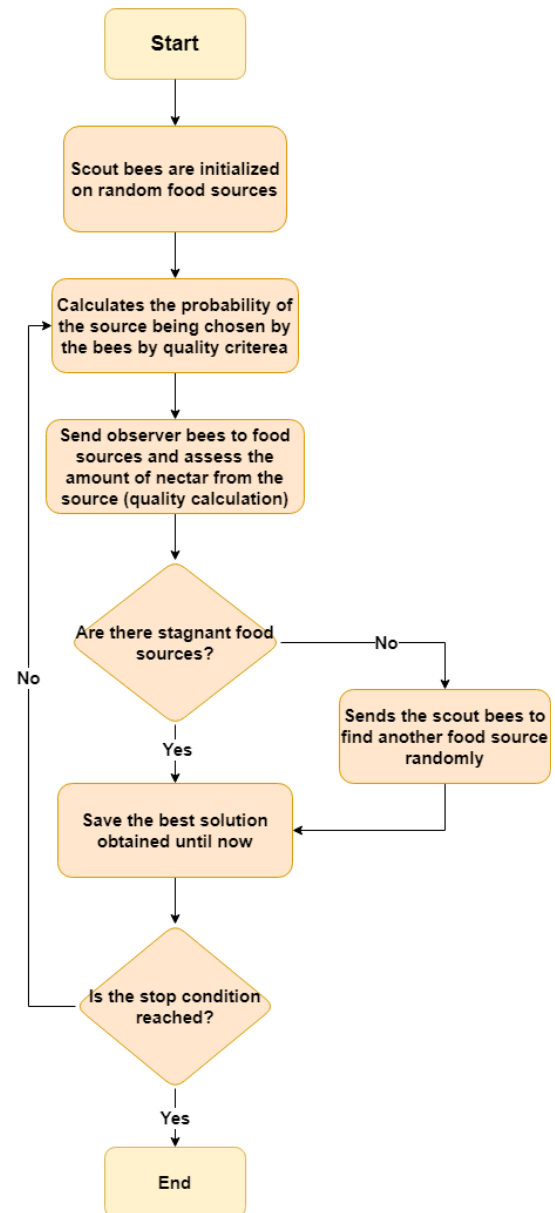

Fig. 1. Flowchart exemplifying the ABC algorithm execution.

Palanisamy and Kanmani [16], in turn, use ABC with a different approach in the onlooker bee phase. Food sources are represented by the attributes of the dataset, and therefore each employed bee becomes responsible for an attribute. Onlooker bees, then, are responsible for selecting the best sources, joining them together, and then evaluating the possible solution through a classifier. Good results were obtained when compared to other bioinspired algorithms, such as ACO, for feature selection. This approach is valid when used in datasets with high data dimensionality since through the work of the onlooker bees, a large reduction in the cost involved in the classification could be observed.

Finally, Agrawal and Chandra [1] use ABC for feature selection in medical image classification processes. Considering the number of factors to be observed in an exam to perform a diagnosis, a high range of attributes is involved in the classification, detracting from the performance of the classifier. For the tests, the kNN and SVM classifiers (with linear and Gaussian kernel) were used. The authors concluded that ABC was successful in its objective, even though biases and unbalanced data interfered remarkably in the final solution.

5 Methodology

In the proposed work, to improve the classification of possible frauds, through Machine Learning, the first step performed was the definition of the dataset that would be used to validate the model.

Through a search on the Kaggle website, the Credit Card Transactions Fraud Detection Dataset was found [3], a complete dataset with a good description of its attributes, generated through simulation, in a synthetic way [4], bringing data from fictitious customers and with good documentation. Thus, it was chosen for this work, since its objective is not only to validate the proposed method, but also to use it in the fraud detection application.

With the definition of the dataset, ABC was implemented using the Python programming language, since it is widely used in the research segment, moreover it has a large number of libraries aimed for data processing, Machine Learning, and statistics. For this, a class called Bee was created, where its attributes are the information related to each bee, such as which food source is under its responsibility, the quality of the source, and stagnation of the same.

The solution, in this case, is in the form of a list of indexes, whose length can vary from 1 to N, where N is the number of attributes of the dataset that you want to work with. The indexes in a solution are not repeated, to avoid information redundancy. Still in ABC, to evaluate the quality of the food source of each bee, the objective function was developed.

5.1 Objective Function

The objective function used performs the training and testing process, through Logistic Regression, Gradient Boosting, or Random Forest, of the dataset con-

[3] https://www.kaggle.com/kartik2112/fraud-detection
[4] https://www.github.com/namebrandon/Sparkov_Data_Generation

sidering the selected attributes as a possible solution. After the classification, the F1-score accuracy measure is used as a way to attest the quality of the solution that was evaluated. The objective of ABC, becomes, then, to search for a set of attributes that reduce the final error and maximize the F1-score of the classifier.

After the implementation of the ABC algorithm, which allows the selection of attributes, the Logistic Regression, Gradient Boosting, and Random Forest algorithms were used to carry out the classification.

5.2 Logistic Regression

Logistic Regression [13] is a statistical method that allows predicting the behavior of a variable, usually binary, based on the value of others, whether discrete or countinuous. As a classifier, it allows determining the class value of an item, according to the values of the input attributes, through a set of weights, determined in the training phase of the algorithm. These weights make up the so-called logistic function, which acts as a kind of objective function within the classifier.

In equation 1,an example of a logistic function can be seen:

$$f(x) = e^{(b0+b1x1+...+bnxn)}/(1+e^{(b0+b1x1+...+bnxn)}) \tag{1}$$

, where $b0, b1, ...bn$ would be the n+1 weights defined through the analysis of training data.

5.3 *Gradient Boosting*

The Gradient Boosting algorithm [8] is part of a class of machine learning algorithms that can be used for problems of predictive modeling of classification or regression.

Its development is based on decision tree models. For this, such trees are added one by one to the set and adjusted to correct the prediction errors of the previous models, a technique known as boosting. In this way, the "gradient reinforcement" occurs, since its loss is minimized as the model is adjusted, similarly to a neural network.

5.4 *Random Forest*

Random Forest [10] was built on the decision tree algorithm and seeks to increase its accuracy and solve its limitations. It consists of a set of decision trees and can also be used in classification and regression problems.

Thus, when we talk about classification, which is the subject of this work, the prediction employing the random trees algorithm is made based on the class label selected by most of the trees of decision generated to compose the algorithm in highlight.

It can be seen, therefore, that the use of these three algorithms as a classification method is due to their simplicity and because they are better suited to the analysis scenario, since the inputs can be classified only in two ways as possible cheats or as a normal operation, being then a scenario with a binary output variable.

Finally, to increase the performance of the proposed algorithm, considering that datasets referring to fraud are usually unbalanced because malicious activities occur less frequently in relation to normal transactions, the dataset balancing treatment was realized.

In this case, since it is a very large set and is difficult to process in a common computer, it was decided to use the balancing edge by undersampling. For this, from the number of records in the training dataset in which the known classification was fraud, that is, a minority dataset within the base, the number of records of common operations can be reduced, without any indication of malicious activity, so that the base of training and tests of the algorithm kept balanced.

After the method was implemented, and any adjustments were made, tests were carried out in an attempt to demonstrate its effectiveness in preventing and combating fraud. Thus, the results obtained are presented in section 6.

5.5 Algorithm Complexity

Considering the implemented algorithm, through its serial execution, its computational complexity order can be obtained from the time of each execution in relation to the input size data, as shown in Figure 2.

Through it, we can perceive that the context to be solved by the algorithm consists on a optimization problem of variable, of NP-complete, with a complexity of the order O(n) = (2^n).

6 Results and Discussion

As mentioned before, a set of Kaggle with synthetic fraud data was used for this article. This dataset is composed of the attributes presented in Table 1. Thus, after processing such attributes, to transform them into input for the developed algorithm, as described in 5, 3 scenarios of tests were proposed. In all of them, 100 repetitions of the algorithm were performed to remove the averages and deviations from the results. It should also be noted that all tests were performed on a notebook with a 64-bit Windows 10 operating system, 16.0 GB RAM, Intel(R) Core(TM) i7-1165G7, 2.80GHz, of 11th generation.

Each of the proposed scenarios are described below:

- **Scenario 1:** In the first test scenario, the Bee Swarm Algorithm is used in an isolated way to perform the selection of atributtes (Feature Selection). Then, Logistic Regression, Random Forest, and Gradient Boosting algorithms are applied to perform the classification on the data set and extract the results for analysis.

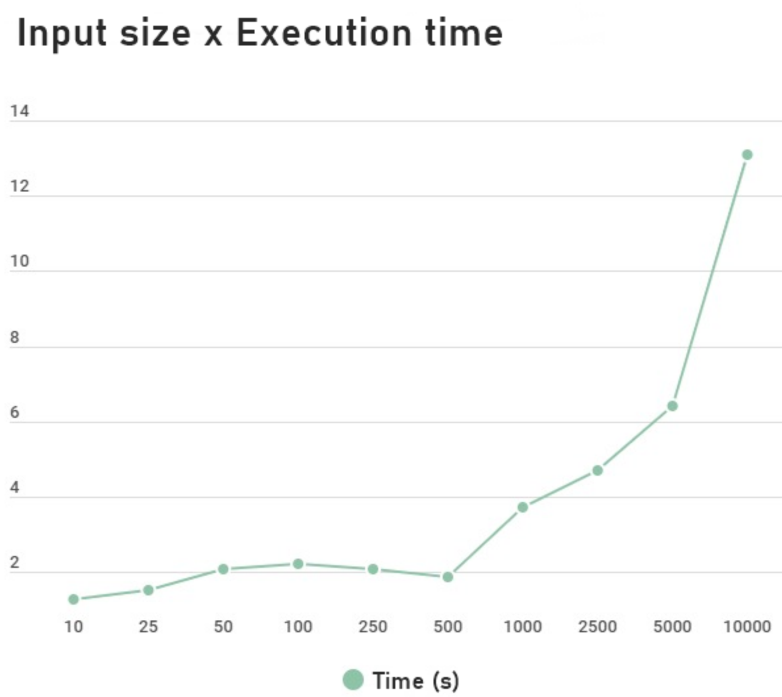

Fig. 2. Complexity analysis of implemented algorithm using runtime for different input sizes

Table 1. Dataset attibutes description for the algorithm tests

Attribute Name	**Description**
trans_date_trans_time	Transaction date and hour
cc_num	Checking Account Number
merchant	Merchant
category	Category
amt	Amount
first	First Name
last	Last Name
gender	Gender
street	Street Address
city	City Address
state	State Address
zip	Zip Code
lat	Latitude
long	Longitude
city_pop	City Population
job	Occupation
dob	Birthday Date
trans_num	Transaction Number
unix_time	System time
merch_lat	Latitude Merchant
merch_long	Longitude Merchant
is_fraud	Fraud Flag

- **Scenario 2:** In the second test scenario, the Logistic Regression, Random Forest, and Gradient Boosting algorithms are applied over the complete data set, without going through a selection of attributes, and then the execution metrics are extracted.
- **Scenario 3:** In the third scenario, ABC is applied to the dataset, to extract the amount of best attributes. This quantity of best attributes will be used as input to the K-best [5], RFE, [6] and Feature Importance [7], all feature selection per wrap algorithms. They in turn will perform the selection of attributes in

[5] K-best: Algorithm that classifies resources by their ranking scores and then selects the top k resources with the highest score. In this scenario, ABC is used to provide the input number k for the algorithm [18].

[6] Recursive Feature Elimination (RFE): Recursive Feature Elimination (RFE): Algorithm that performs selection by recursively removing the attributes and building a model on those that remain. It uses the precision of the model to identify which attributes (or a combination of them) to keep (strong attributes) and which to discard (weak attributes) [3].

[7] Feature Importance: As a third selection method, it was chosen to use the N most relevant attributes for the algorithm, whose importance is calculated using the Extra Trees-Classifier algorithm, that is, the most important attributes are calculated and, in the sequence, from the bee algorithm it is obtained how many will remain for training and testing of the model.

the dataset. After this procedure, for each of the attribute selection methods, the three classification models (Logistic Regression, Random Forest, and Gradient Boosting) are applied for extraction of the evaluation metrics.

In addition to the 100 executions for each of the proposed scenarios, it was used as parameters for the ABC algorithm, the maximum number of 100 iterations, 10 bees, and a stagnation limit equals 5 food sources in all runs.

From the proposed scenarios, from this point onwards, the results obtained and their analysis will be displayed from now on.

In Tables 2 and 3, it is possible to observe the data collected from the training scenarios with 100 iterations of the ABC, when pertinent (Scenarios 1 and 3), and from the tests with the same amount. In the results, two different situations can be observed:

- In the first, for the Gradient Boosting and Random Forest classifiers, the use of the ABC algorithm practically does not change the accuracy and the F1-Score (measures related to the algorithm's hit rates) of the results. The biggest difference is presented when there is a direct application of the algorithm to select attributes (Scenario 1), in which the results end up getting worse by approximately 1%.
- In the second, with the Logistic Regression classification model, there is an increase in performance, both for the accuracy and the F1-Score, for Scenarios 1 and 3 with respect to Scenario 2 (no attribute selection).

The same information, referring to the test data, is presented in the graphs of Figures 3 and 4.

Table 2. 100 iterations training

Test Scenario	Classifier	Feature Selection Algorithm	Accuracy Average	F1-Score Average	Accuracy Standard Deviation	F1-Score Standard Deviation
Scenario 1	Gradient Boosting	ABC	97,2425%	97,1355%	3,7057%	4,1919%
Scenario 2	Gradient Boosting	N/A	98,6867%	98,6744%	0,0000%	0,0000%
Scenario 3	Gradient Boosting	Feature Importance	98,4318%	98,4169%	0,1578%	0,1562%
Scenario 3	Gradient Boosting	Kbest	98,1872%	98,1643%	0,5567%	0,5793%
Scenario 3	Gradient Boosting	RFE	98,6458%	98,6321%	0,0990%	0,1008%
Scenario 1	Random Forest	ABC	99,8902%	99,8930%	0,8925%	0,8634%
Scenario 2	Random Forest	N/A	99,9996%	99,9996%	0,0019%	0,0019%
Scenario 3	Random Forest	Feature Importance	99,9998%	99,9998%	0,0013%	0,0013%
Scenario 3	Random Forest	Kbest	99,9997%	99,9997%	0,0016%	0,0016%
Scenario 3	Random Forest	RFE	99,9996%	99,9996%	0,0019%	0,0019%
Scenario 1	Logistic Regression	ABC	93,4635%	93,4943%	7,6677%	6,1883%
Scenario 2	Logistic Regression	N/A	83,9646%	82,2239%	0,0000%	0,0000%
Scenario 3	Logistic Regression	Feature Importance	85,3035%	83,7628%	3,7050%	4,1998%
Scenario 3	Logistic Regression	Kbest	95,8987%	95,7574%	0,0747%	0,0778%
Scenario 3	Logistic Regression	RFE	89,8953%	88,7512%	0,0803%	0,1794%

Table 3. 100 iterations test

Test Scenario	Classifier	Feature Selection Algorithm	Accuracy Average	F1-Score Average	Accuracy Standard Deviation	F1-Score Standard Deviation
Scenario 1	Gradient Boosting	ABC	97,1201%	97,0496%	3,8576%	4,3000%
Scenario 2	Gradient Boosting	N/A	98,4680%	98,4704%	0,0000%	0,0000%
Scenario 3	Gradient Boosting	Feature Importance	98,0604%	98,0645%	0,2122%	0,2091%
Scenario 3	Gradient Boosting	Kbest	97,8501%	97,8493%	0,4374%	0,4475%
Scenario 3	Gradient Boosting	RFE	98,4130%	98,4150%	0,0978%	0,0976%
Scenario 1	Random Forest	ABC	98,1452%	98,1729%	2,3717%	2,2374%
Scenario 2	Random Forest	N/A	99,0548%	99,0592%	0,0431%	0,0430%
Scenario 3	Random Forest	Feature Importance	98,8308%	98,8373%	0,2011%	0,1993%
Scenario 3	Random Forest	Kbest	98,6661%	98,6737%	0,2880%	0,2872%
Scenario 3	Random Forest	RFE	98,9898%	98,9954%	0,1456%	0,1436%
Scenario 1	Logistic Regression	ABC	93,8819%	94,0002%	7,6516%	6,1605%
Scenario 2	Logistic Regression	N/A	84,1252%	82,6835%	0,0000%	0,0000%
Scenario 3	Logistic Regression	Feature Importance	85,5011%	84,2290%	3,7847%	4,2124%
Scenario 3	Logistic Regression	Kbest	96,2842%	96,2183%	0,0874%	0,0890%
Scenario 3	Logistic Regression	RFE	90,2016%	89,2765%	0,1649%	0,2679%

In addition, it is noted that, in Scenario 3, when applying the bees algorithm to generate the number of attributes reverted to feed both the K-best and the RFE, there is greater stability of results, with a lower standard deviation, again for both metrics (Accuracy and F1-Score), if compared with Scenario 1 and with the use to feed the Feature Importance algorithm, although both the latter also presented better performance than the exclusive execution of the classifiers.

Fig. 3. Accuracy comparison in different scenarios.

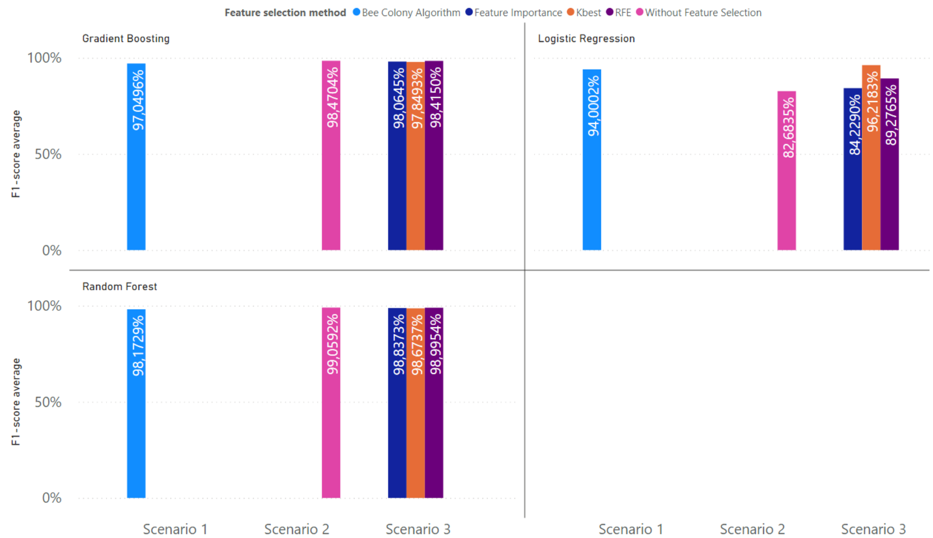

Fig. 4. F1-Score comparison in different scenarios.

In Tables 4 and 5, the results related to the training and testing confusion matrices are presented, respectively for the executions of the scenarios with 100 iterations. In them, the indexes of the matrices are designated in the column headers (Ex: **Avg 00** is referring to the confusion matrix in the position where the returned result should be 0 (expected result) and the predicted result actually materialized as 0). From these two tables, one can notice the low rate of false positives and false negatives, when compared with true positives and true negatives, mainly for the Gradient Boosting and Random Forest classification algorithms.

Finally, for the scenario in which there was the selection of attributes through the bee algorithm, the execution times of the algorithm and the average of the best score returned by the objective function of the same. These results can be seen in Table 6.

7 Conclusion

This article presented an attribute selection technique based on the ABC algorithm. The results show that the reduction in the number of attributes can not only reduce the complexity of the base, for further training and testing of the classifier, but it can also achieve a higher accuracy of classification than that obtained when using the complete set of data. The results obtained corroborate the quality of ABC as a Wrapper method for Feature Selection, validating the hypothesis that the algorithm could reduce the cost and increase the quality of the results obtained in a fraud detection process, so significant nowadays.

Table 4. Confusion Matrix for training with 100 iterations

Test Scenario	Classifier	Feature Selection Method	Avg 00	Avg 01	Avg 10	Avg 11
Scenario 1	Gradient Boosting	ABC	5201	217	73	5017
Scenario 2	Gradient Boosting	N/A	5234	98	40	5136
Scenario 3	Gradient Boosting	Feature Importance	5222	113	52	5121
Scenario 3	Gradient Boosting	Kbest	5214	131	60	5103
Scenario 3	Gradient Boosting	RFE	5235	103	39	5131
Scenario 1	Random Forest	ABC	5266	4	8	5230
Scenario 2	Random Forest	N/A	5274	0	0	5234
Scenario 3	Random Forest	Feature Importance	5274	0	0	5234
Scenario 3	Random Forest	Kbest	5274	0	0	5234
Scenario 3	Random Forest	RFE	5274	0	0	5234
Scenario 1	Logistic Regression	ABC	5062	475	212	4759
Scenario 2	Logistic Regression	N/A	4926	1337	348	3897
Scenario 3	Logistic Regression	Feature Importance	4954	1224	320	4010
Scenario 3	Logistic Regression	Kbest	5214	371	60	4863
Scenario 3	Logistic Regression	RFE	5257	1045	17	4189

Table 5. Confusion Matrix for training with 100 iterations

Test Scenario	Classifier	Feature Selection Method	Avg 00	Avg 01	Avg 10	Avg 11
Scenario 1	Gradient Boosting	ABC	2197	95	35	2177
Scenario 2	Gradient Boosting	N/A	2214	51	18	2221
Scenario 3	Gradient Boosting	Feature Importance	2204	59	28	2213
Scenario 3	Gradient Boosting	Kbest	2203	68	29	2204
Scenario 3	Gradient Boosting	RFE	2213	53	19	2219
Scenario 1	Random Forest	ABC	2194	45	38	2227
Scenario 2	Random Forest	N/A	2220	31	12	2241
Scenario 3	Random Forest	Feature Importance	2213	34	19	2238
Scenario 3	Random Forest	Kbest	2209	37	23	2235
Scenario 3	Random Forest	RFE	2217	30	15	2242
Scenario 1	Logistic Regression	ABC	2146	190	86	2082
Scenario 2	Logistic Regression	N/A	2082	565	150	1707
Scenario 3	Logistic Regression	Feature Importance	2095	516	137	1756
Scenario 3	Logistic Regression	Kbest	2208	143	24	2129
Scenario 3	Logistic Regression	RFE	2225	435	7	1838

Table 6. 100 iterations ABC test results

Test Scenario	ABC Iterations	Best Score Average	Average Time	Best Score Standard Deviation	Time Standard Deviation
Scenario 1	100	0,0351	14,0794	0,0061	3,7044

References

1. Agrawal, V., Chandra, S.: Feature selection using artificial bee colony algorithm for medical image classification. In: 2015 Eighth International Conference on Contemporary Computing (IC3). pp. 171–176. IEEE (2015)
2. Bansal, J.C., Sharma, H., Jadon, S.S.: Artificial bee colony algorithm: a survey. International Journal of Advanced Intelligence Paradigms **5**(1-2), 123–159 (2013)
3. Chen, X.w., Jeong, J.C.: Enhanced recursive feature elimination. In: Sixth International Conference on Machine Learning and Applications (ICMLA 2007). pp. 429–435 (2007). https://doi.org/10.1109/ICMLA.2007.35
4. Darwish, S.M.: An intelligent credit card fraud detection approach based on semantic fusion of two classifiers. Soft Computing **24**(2), 1243–1253 (2020)
5. Dorigo, M., Di Caro, G., Gambardella, L.M.: Ant algorithms for discrete optimization. Artificial life **5**(2), 137–172 (1999)
6. Eberhart, R., Kennedy, J.: Particle swarm optimization. In: Proceedings of the IEEE international conference on neural networks. vol. 4, pp. 1942–1948. Citeseer (1995)
7. El Aboudi, N., Benhlima, L.: Review on wrapper feature selection approaches. In: 2016 International Conference on Engineering & MIS (ICEMIS). pp. 1–5. IEEE (2016)
8. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Annals of statistics pp. 1189–1232 (2001)
9. Hancer, E., Xue, B., Karaboga, D., Zhang, M.: A binary abc algorithm based on advanced similarity scheme for feature selection. Applied Soft Computing **36**, 334–348 (2015)
10. Ho, T.K.: Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition. vol. 1, pp. 278–282. IEEE (1995)
11. Karaboga, D.: An idea based on honey bee swarm for numerical optimization. Tech. rep., Technical report-tr06, Erciyes university, engineering faculty, computer engineering department (2005)
12. Karaboga, D., Akay, B.: A comparative study of artificial bee colony algorithm. Applied mathematics and computation **214**(1), 108–132 (2009)
13. Kleinbaum, D.G., Dietz, K., Gail, M., Klein, M., Klein, M.: Logistic regression. Springer, New York (2002)
14. Lanzi, P.L.: Fast feature selection with genetic algorithms: a filter approach. In: Proceedings of 1997 IEEE International Conference on Evolutionary Computation (ICEC'97). pp. 537–540. IEEE (1997)
15. Liu, O., Ma, J., Poon, P.L., Zhang, J.: On an ant colony-based approach for business fraud detection. In: International Conference on Intelligent Computing. pp. 1104–1111. Springer (2009)
16. Palanisamy, S., Kanmani, S.: Artificial bee colony approach for optimizing feature selection. International Journal of Computer Science Issues (IJCSI) **9**(3), 432 (2012)
17. Pavithra, T., Thangadurai, D.K.: Fraud detection of credit cards using abc methodology based on svm algorithm (2019)
18. Yang, Y., Pedersen, J.O.: A comparative study on feature selection in text categorization. In: Icml. vol. 97, pp. 412–420. Nashville, TN, USA (1997)