Feature Selection Approaches to Fraud Detection in e-Payment Systems

Conference Paper *in* Lecture Notes in Business Information Processing · February 2017

DOI: 10.1007/978-3-319-53676-7_9

2 authors:

Rafael Lima
Federal University of Minas Gerais
5 PUBLICATIONS **23** CITATIONS
SEE PROFILE

Adriano C. M. Pereira
Federal University of Minas Gerais
150 PUBLICATIONS **1,230** CITATIONS
SEE PROFILE

Feature Selection Approaches to Fraud Detection in e-Payment Systems

Rafael Franca Lima and Adriano C. M. Pereira

Department of Computer Science (DCC),
Federal University of Minas Gerais (UFMG),
Belo Horizonte, Minas Gerais, Brazil 31270-901
rafaelfrancalima@dcc.ufmg.br,adrianoc@dcc.ufmg.br

Abstract. Due to the large amount of data generated in electronic transactions, to find the best set of features is an essential task to identify frauds. Fraud detection is a specific application of anomaly detection, characterized by a large imbalance between the classes, which can be a detrimental factor for feature selection techniques. In this work we evaluate the behavior and impact of feature selection techniques to detect fraud in a Web Transaction scenario. To measure the effectiveness of the feature selection approach we use some state-of-the-art classification techniques to identify frauds, using a real application data. Our results show that the imbalance between the classes reduces the effectiveness of feature selection and that resampling strategy applied in this task improves the final results. We achieve a very good performance, reducing the number of features and presenting financial gains of up to 57.5% compared to the actual scenario of the company.

Key words: e-Commerce, Web, Fraud Detection, e-Payment systems, Feature Selection, Anomaly Detection, Resampling

1 Introduction

People tend to use e-commerce systems because it is easy and convenient. However, this popularity generate a huge increase in the number of online frauds, resulting in billions of dollars losses each year worldwide. The works that deal with this topic usually try to find anomaly patterns that can be considered as a fraudulent behavior, such as a fraud in a Web transaction scenario [1, 2].

Although we find several works in this research topic [3], there are still some points of improvement to detect frauds in electronic transactions. One of these points is related to how to filter the large amount of data generated in electronic transactions and transform this data in useful information. The main task in this context is to find a subset of features that are able to identify the anomalous behavior, which is called Feature Selection [4].

In general, the feature selection techniques perform a selection of features based on the class value. However, in anomaly detection scenario, as fraud detection, the high imbalanced distribution between classes (e.g., distribution of fraud and non-fraud classes) generates new challenges to the feature selection techniques, which tend to select attributes in favor of the dominant class [5].

Thus, in this work we investigated how the imbalance between the classes affect the selection of attributes and, consequently, the models to detect fraud. To guide this study we enumerate some research hypotheses:

1. The high class-imbalanced reduces the effectiveness of feature selection to detect anomalies in electronic transactions;
2. Traditional methods of Feature Selection are not suitable to detect anomalies;
3. The reduction of class-imbalance, using some resampling approaches, before the feature selection step, can improve the effectiveness of the selection of attributes.

To check these hypotheses, in this work we conduct experiments using three popular traditional feature selection techniques, i.e., techniques that do not use different strategies for the imbalance between the classes. We applied these techniques in two distinct distribution of the same dataset. The first one keeps the real proportion of fraud, while the second one reduces the imbalance between classes using some resampling methods. Thus, it is possible to measure how the imbalance between classes affects the selection of attributes.

To perform a deep investigation in resampling approaches, we evaluate 7 methods to reduce the class-imbalance before feature selection step, including a resampling method created by us, specifically for this function. To build effectiveness models to identify frauds and evaluate the attributes selected by each approach, we used four classification techniques.

In order to test and validate these approaches, we conducted our experiments in an actual dataset from one of the most popular electronic payment systems in Latin America. In Section 5.1 we present a better characterization of this dataset. To evaluate the models we use three performance metrics related to classification accuracy (F1 and AUC) and Economic Efficiency (EE).

The main contributions of this work are: (i) The analysis of the effectiveness of traditional methods for feature selection in the detection of anomalies; (ii) A deep investigation in effectiveness of resampling methods before feature selection techniques; (iii) The creation of a resampling method to be used before feature selection step in anomalous scenario; (iv) The construction of a model for fraud detection, that combine strategies of resampling before feature selection and classification techniques; and (v) A complete validation of the proposed method using an actual data from an electronic payment system.

2 Related Work

Methods for detecting fraud in web-transactions are extensively discussed in the scientific community [3]. Although the works have focus on different techniques to detect frauds, none of these works focus on feature selection to fraud detection. In those studies they were not using different approaches for feature selection to suit an anomaly detection application.

To confirm this lack we performed a systematic literature review among fraud detection works, following the methodology described in [6]. In this revision we

evaluate 30 most cited works and 20 most relevant researches in fraud detection. We do not identify any use of appropriate *feature selection* approaches or investigation in resampling methods before feature selection to identify frauds.

We found some researches in feature selection for high imbalance dataset [7, 5, 8, 9, 10]. However, theses works do not conduct fraud detection experiments. The main scenarios discussed in these researches are the biological genes (micro-array analysis) and diseases. Although, these scenarios also contains high imbalance between classes, the bases used differ in the ratio between classes, attributes, numbers and nature of the data.

There are works that discuss methods of resampling to reduce the high imbalance between classes on training for classification [11], but few studies have analyzed the application of these methods before the feature selection step. Resampling methods before feature selection step was used in [12] and [13]. However, these studies only exploit micro-array datasets and not focus on the investigation of different methods and ratios of resampling before feature selection.

The researches presented in this section suggest the lack of study, as well as the creation of mechanisms that are suitable for feature selection in fraud detection. Therefore, in this work we investigated different resampling methods before feature selection step in order to develop more effective methods to identify frauds in e-commerce, which is the main contribution of this work.

3 Conceptual Fundamentals

In this section we refer to minority class (fraud) as positive and majority as negative. In this work we report three supervised feature selection techniques, which are briefly described, as follows:

- **CFS** (Correlation-based Feature Subset Selection) aims to find sets of attributes that are highly correlated with the class, and they are not correlated with the other attributes in the set . Several metrics can be used to calculate the correlation matrix, in general the symmetric uncertainty is used [14].
- **Gain Ratio** is a method based on the concept of entropy of a sample space. The entropy of this space is characterized by the impurity of the data. Thus, the entropy returns a value from 0 to 1, depending on homogeneity of the each attribute for classification [15].
- **Relief** considers that good attributes have equal values for instances of the same class and different values for instances of different classes. Relief as Gain Ratio perform the individual merit analysis for each attribute. It starts by randomly choosing of an instance and find the closest instance of a different class (nearest miss) and nearest instance of the same class (nearest hit). After this, it defines the weight of each attribute as P(different values of A to an instance and their nearest miss) - P(different values of A to an instance and their nearest hit) [16].

The Gain Ratio and Relief feature selection techniques generate a ranking with merit of attributes. In this work, we decide to not fix the number of attributes and we cut in the ranking through a threshold. This threshold was determined finding the elbow in the curve formed by the feature merits ranking.

To reduce the class-imbalance to apply this feature selection methods we used resampling methods. There are two strategies to perform this task, *undersampling* and *oversampling*. In undersampling (US) the instances from the majority class are removed. The possible problem of this method is that important information can be lost. While, *oversampling* (OS) strategy duplicates instances of the minority class. This method can cause overfitting on the data.

The random methods, random undersampling (**RUS**) and random oversampling (**ROS**) do not care about the weaknesses of the strategies and randomly choose the instances that will be replicated or removed to deal with the imbalance between classes. However, there are smart methods that address these weaknesses. In this work, besides ROS and RUS, we used four smart resampling methods and create another one, specifically to apply before feature selection step to anomaly detection. These methods are briefly described in the Table 1.

Table 1. Resampling methods

Method	Label	Type	Description	Reference
NearMiss-1	**NM-1**	US	Remove negatives examples whose average distances to three closest positive examples are the smallest.	[17]
NearMiss-2	**NM-2**	US	Remove negative examples based on their average distances to three farthest positive examples.	[17]
NearMiss-3	**NM-3**	US	Remove negative examples to guarantees every positive example is surrounded by some negatives examples.	[17]
Smote	**Smote**	OS	Adds new artificial minority examples by interpolating between pre-existing minority instances.	[18]
Sampling Outlier	**SO**	Mixed	Mixed undersampling and oversampling, following the pseudocode 1.	Created by us in this work

The pseudocode 1 shows the resampling method that we created specifically to reduce the class imbalance on feature selection step in anomaly scenarios. The main idea of this method is to remove the rare instances of negative class and replicate the rare instances of positive class, using the Smote method. Unlike the other methods, in this method is not necessary to inform the new ratio between the classes. In SO method, this ratio is determined by the number of instances that satisfy the conditions.

To compare the efficiency of the feature selection techniques and identify the frauds, we use some state-of-the-art supervised classification techniques, which are **Bayesian Networks** (see in [19]), **Logistic Regression** (see in [20]) and **Decision Tree - J48 Implementation** (see in [21]). Moreover, we use tree metrics to evaluate the results, which are briefly described:

1. **AUC** is the area under the curve formed by false positive (FPR) rate and true positive rate (TPR), varying the classification threshold.
2. **Avg_F1** is a weighted average of the precision and recall. For each class (0 and 1) we get a $F1$, therefore we use the Average F-Measure (Avg_F1) to facilitate comparison of results.
3. **EE** evaluate the financial gains after application of the fraud detection models, using the Equation 1. To facilitate the comparison of the models, our results are

Algorithm 1: Sampling Outlier Resampling Method

Give: i_{pos} and i_{neg} instances of positive and negative class
F_{pos} vector of size len(i_{pos})
F_{neg} vector of size len(i_{neg})
for $i \in i_{pos}$ **do**
  knn[i][]= Calculate the t KNN of i
  for $k \in Knn[i][]$ **do**
    $F_{pos}[K]$+=1
  end
end
for $j \in i_{neg}$ **do**
  knn[j][]= Calculate the z KNN of j
  for $k \in Knn[j][]$ **do**
    $F_{neg}[K]$+=1
  end
end
M_F_{pos} = Mean(F_{pos})
DP_F_{pos} = Stantard deviation of (F_{pos})
M_F_{neg} = Mean(F_{neg})
Dp_F_{neg} = Stantard deviation of (F_{neg})
$Cut_{pos} = (M\_F_{pos} - Dp\_F_{pos})$
$Cut_{neg} = (M\_F_{neg} - Dp\_F_{pos})$
for $i \in i_{pos}$ **do**
  if $F_{pos}[i] < Cut_{pos}$ **then**
    Replicate i using SMOTE
  end
end
for $j \in i_{neg}$ **do**
  if $F_{neg}[j] < Cut_{neg}$ **then**
    Remove j
  end
end

based in a Relative Economic Efficiency (*Relative_EE*), which measures the percentage gain on the current scenario of the company [22].

$$EE = k \cdot TN_{Value} - ((1-k) \cdot FN_{Value} + p \cdot FP_{Value}), \quad (1)$$

where k is a gain of the company for each transaction; p is the penalty for false positive transaction;TN_Value, FN_{Value} and FP_{Value} are the sum of the transaction values for true negative, false negative and false positive transactions.

4 Methodology

In this section we present the methodology used in this research to detect fraud in a Web transaction scenario. Figure 1 presents the process to investigate the research hypothesis. The main difference of this work is in steps 2 and 3, which are investigated and applied here.

The first step performs the **normalization of the database** to make easier the comparison and analysis of the techniques used herein. The continuous attributes were normalized between 0 and 1 and the discrete attributes were all considered as non-ordinal categories. After this, we separate a subset of training data, called **data for selection**, which will be used to select attributes for all feature selection techniques. This subset contains the same features, but different instances of the data used to validation, guaranteeing the generality of selection.

In the step 2 we started the validation of our first hypothesis, the high imbalance between classes reduces the effectiveness of selection attributes to detect fraud. To perform this **we used, before feature selection step, two distinct strategy about the proportion between the classes**. The first one generate a subset to feature selection with the real proportion between the classes, while in the second strategy we generate a subset to feature selection reducing the class-imbalance through distinct resampling approaches.

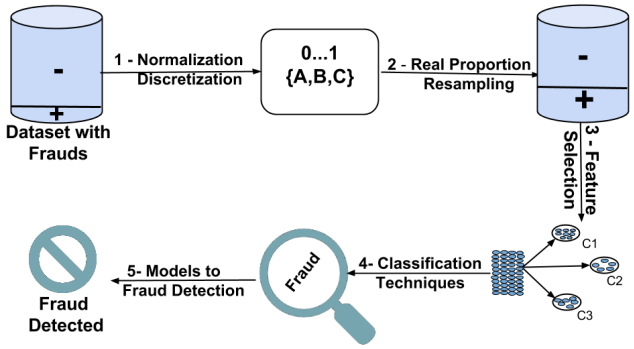

Fig. 1. Methodology to construct the fraud detection models

To confirm this hypothesis would be enough to use just one resampling method, but one of the differences of this work is the comparison of resampling strategies before feature selection to identify frauds. Therefore, **we used 6 state of the art techniques and created a resampling technique**, called *Sampling Outlier* (*SO*) with concepts that we considered important to feature selection in anomaly detection. These techniques were describes in section 3.

The resampling methods, except *SO*, require the information about the new ratio between the classes. Thus, for each resampling method we create 8 subsets with distinct fraud proportion (5%,10%,15%,20%,25%,30%,40% e 50%). Therefore, we created 50 subsets to feature selection, formed through: 6 resampling melhods with 8 disctint proportion of frauds, 1 resampling method that does not need to inform the fraud ratio and 1 subset with the real proportion of fraud. Using this settings, we can determine the best fraud proportion for each technique and verify if increasing the fraud proportion the performance of the feature selection also increases or not.

In the step 3 **we applied the feature selection techniques**, presented in the Section 3, on each subset to feature selection. Besides that, we create a set of features with all features available (NO_FS) and 1 subset of attributes generated through the combination of the best attributes, selected in each feature selection techniques ($Merge$). In $Merge$ the merit of a attributes is the frequency of this attribute on the subset of features with the best performances.

In the next step, **we used the classification techniques**, as described in Section 3, to identify fraud and evaluate the subsets of attributes for each approach. We use a training subset different from the feature selection training subset. To ensure the generality of the solutions we use the 8-fold Cross-Validation.

We do not use any resampling approach on training to classification, preserving the original fraud ratio. Although, we believe that the use of resampling also in this step could achieve better results, if we had used we could be favoring any of the resampling methods applied on the feature selection step. To evaluate the results found in this step we used the metrics described in Section 3 and we performed the *Friedman* ([23]) and *Wilcoxon* ([24]) statistical test.

After the analysis, in step 5 **we constructed the fraud detection models**, which consists of a classification technique and a feature selection approach.

5 Case Study

This section presents our case study where we apply our models to detect fraud in a Web payment system.

5.1 Data Set Description

We used a real dataset from an electronic payment system called PagSeguro[1] to evaluate our methods. PagSeguro is one of the most used electronic payment systems in Latin America, mainly in Brazil. It is evident the need for efficient techniques that are able to identify frauds in this kind of system.

Table 2 shows the main information about this dataset.

Number of features	380
Number of continuous features	248
Number of categorical features	133
Fraud proportion	≅ 1.2%
Period of analysis	2012/2013

Table 2. Dataset - General Overview

5.2 Experimental Results

In this subsection, we present and compare the results obtained after the use of the methodology explained in Section 4 applied in the real dataset from PagSeguro electronic payment system, from Subsection 5.1.

Table 3. Feature Selection Approaches - Number of Features. Legend: □:CFS; △:GainRatio; ○:Relief

	NM-1			**NM-2**			**NM-3**			**ROS**			**RUS**			**SMOTE**			**SO**			**Real**		
	□	△	○	□	△	○	□	△	○	□	△	○	□	△	○	□	△	○	□	△	○	□	△	○
																			28	86	33	11	77	29
5%	34	325	26	46	192	31	27	44	32	22	81	26	20	74	30	28	67	26	-	-		-	-	
10%	19	334	19	43	191	28	27	44	32	24	77	31	27	52	23	27	69	32	-	-		-	-	
15%	20	340	26	32	21	22	27	44	32	29	45	32	32	52	37	22	81	31	-	-		-	-	
20%	19	340	23	28	23	25	27	44	32	28	57	31	38	63	30	23	78	34	-	-		-	-	
25%	18	341	26	28	26	17	27	44	32	27	56	30	28	70	29	22	84	31	-	-		-	-	
30%	19	341	26	26	31	27	27	44	32	27	34	31	25	42	25	21	80	38	-	-		-	-	
40%	18	45	20	22	41	20	26	37	30	23	37	27	23	51	26	21	37	27	-	-		-	-	
50%	14	56	17	22	44	18	26	41	22	26	35	27	28	12	22	20	54	28	-	-		-	-	

[1] http://pagseguro.uol.com.br

Table 3 contains the number of features of each subset generated by the feature selection techniques, using the resampling or real proportion. After using the feature selection techniques, we obtain subsets of data with different numbers of features for each fraud proportion in all resampling methods, showing that the proportion of fraud directly influence on feature selection. The resampling method $NM-3$, due to its nature has generated the same instances, regardless of reported fraud ratio.

One important contribution of this work is the comparison between the different fraud ratio for each resampling method applied before feature selection step. Thus the graphic of Figure 2 shows the AUC metric obtained after the classification on several subsets of attributes. This subsets were generated by each feature selection techniques applied in the dataset with distinct fraud proportion in each resampling method and the real proportion (*Real*). Due to limited space we present this analysis only for the AUC metric, however we emphasize that we obtained similar behavior to analyze the $F1$ and EE metrics.

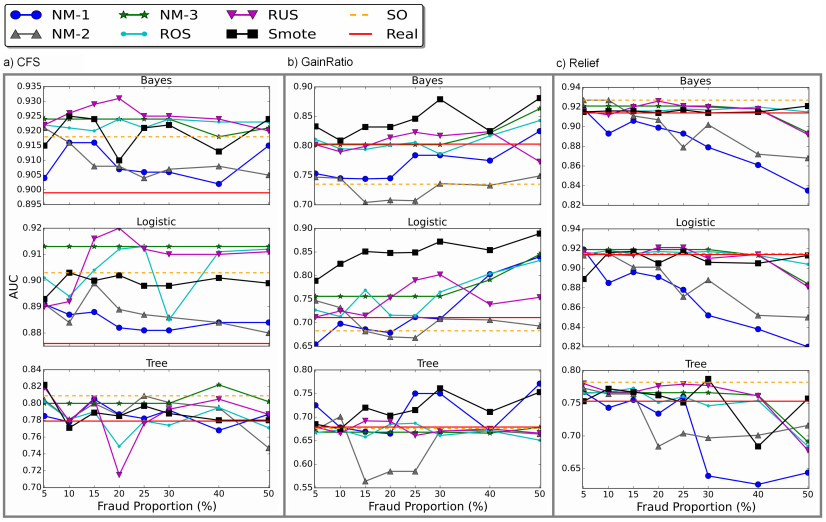

Fig. 2. Performance of classification techniques on subset of attributes selected by feature selection on dataset using resampling methods in distinct fraud ratio.

The graphd a) in Figure 2 show the behavior of resampling methods before the feature selection technique CFS. We can note that when we set a adequate fraud ratio, the RUS is a good alternative to the CFS technique. The SO method was better than 3 methods in all frauds ratio. All approaches that used resampling methods, before CFS feature selection, achieved better AUC than the real proportion (*Real*). Thus, we can prove that the high-class imbalance reduce the effectiveness of CFS feature selection technique and resampling methods before feature selection is a good strategy to improve the performance to identify frauds.

The graphs *b*) show the same analysis to GainRatio. The Smote method was the resampling more adequate for this technique. Contrary to CFS technique, the *SO* was not a good resampling method before GainRatio feature selection. When a good fraud ratio was used, all approaches using resampling methods, except *SO*, outperformed the approach using the *Real* before feature selection.

Lastly, the graphs *c*) present analysis to Relief technique. We note that Relief was less sensitive to variation in fraud proportion. However, it selected more effective attributes when used resampling before Relief Feature Selection. The exception were the approaches with resampling methods $NM-1$ and $NM-2$, which not outperformed the *Real*. The method *SO* was very effective to reduce the class-imbalance before feature selection with Relief.

In CFS and Relief techniques the subsets that used lower fraud ratios obtained better results. This behavior rejects the hypothesis that greater reduction in imbalance between classes implies in better performance in the feature selection. We can note by Figure 2 that the effectiveness of feature selection is not directly correlated with the increase of fraud ratio in resampling method.

In general, we can note for all feature selection techniques that the random undersampling (RUS) beats the random oversampling (ROS). This behavior agrees with [25], however in that work resampling is used before classification.

We can not found an ideal fraud proportion for all resampling methods. We realize that each method has achieved better performance using a fraud ratio. Then, the method *SO*, created for us, can be a good alternative when the cost of performing a classification to choose the best fraud ratio is high. We choose the ideal fraud ratio for each resampling method, according the graphic of Figure 2, to reduce class-imbalance before each feature selection technique.

Table 4 presents this fraud ratio for each resampling method and the percentual gain obtained in comparison with the same combination of feature selection and classification techniques, but using the real proportion before feature selection. Due to space constraints, we omit the results to the decision tree technique in this table. However, it follows the same behavior of the other techniques.

We have highlighted in **bold** the best results for each classification technique in table 4. We may note that using resampling for selecting attributes we achieved significant gains in terms of AUC, $F1$ and EE. The biggest gains were achieved when used EE. That happens, because the cost of a True Positive (TP) and False Positive (FP) on this metric is not the same. Simulating what happens in real scenarios, the cost is 97% (TP) to 3% (FP). While in the other metrics used in this work the cost of a TP and FP is the same.

After the analysis of ideal fraud ratio for each resampling methods we created a feature selection approach that combines the features more frequent in the best subset selected from dataset with resampling methods, presented in the table 4. This model, was called $Merge$. In addition, we create a approach that use all features to classification, that is, without the use of any feature selection in the dataset, this approach was called NO_FS.

Thus, we have 4 types of model to identify frauds. The models consists of the same classification techniques and distinct approach to feature selection or

			Bayes			Logistic		
FS	Resampling	%	AUC	F1	EE	AUC	F1	EE
CFS	NM-1	10	2.11	2.08	2.80	1.26	2.86	16.43
	NM-2	5	2.45	**3.61**	10.11	4.22	2.57	21.90
	NM-3	5:30	2.78	2.78	20.43	4.68	**4.86**	**59.08**
	ROS	20	2.78	2.78	10.97	4.11	3.43	12.68
	RUS	20	**3.56**	**3.61**	**21.29**	**5.02**	4.57	36.89
	Smote	10	2.89	3.06	11.18	3.08	2.86	3.46
	SO	X	2.11	2.08	2.80	3.08	4.14	13.26
Gain Ratio	NM-1	50	2.74	-5.86	23.02	18.14	9.84	**77.43**
	NM-2	5	-6.97	-1.85	3.77	5.06	0.00	25.66
	NM-3	50	7.47	**4.47**	33.96	18.85	5.24	31.42
	ROS	40	4.98	0.77	4.91	19.27	7.14	59.73
	RUS	25	1.74	-0.77	8.68	12.80	6.83	18.58
	Smote	30	**9.71**	-5.39	**45.66**	**25.04**	**12.70**	68.14
	SO	X	-8.47	0.00	36.98	-3.94	-0.63	26.99
Relief	NM-1	5	0.44	1.34	8.82	0.55	0.41	**5.23**
	NM-2	5	-5.03	-7.51	-28.36	-0.11	-4.38	-12.85
	NM-3	5:30	0.77	**3.49**	**19.96**	0.55	0.41	-2.18
	ROS	20	0.22	1.61	0.84	0.33	**0.68**	0.00
	RUS	20	1.31	2.82	17.65	**0.77**	0.27	4.36
	Smote	25	0.33	1.74	6.93	0.33	0.55	0.00
	SO	X	**1.42**	2.41	16.39	0.11	0.00	-8.28

Table 4. Percentage gain in fraud detection using resampling before feature selection over the same techniques using the real proportion before feature selection.

none feature selection. The model *Real* that do not use resampling before feature selection, the model *Resampling* that the use the best resampling method with the ideal fraud ratio for each feature selection technique, the model $Merge$ that combines the features most frequent in the best subset of features and the model No_FS that contains all features available on dataset.

Table 5 presents the comparison of the best fraud detection models for each classification techniques, using the three evaluation metrics. The models *Real* were excluded from this comparison, because in the previous analysis we demonstrated that when we use resampling before feature selection we achieved better results. Thus, it compares the models using resampling on feature selection step, without feature selection and with the $Merge$ feature selection strategy.

Table 5. Performance of models to fraud detection

	Bayes				**Logistic**				**Tree**			
		AUC	F1	EE		AUC	F1	EE		AUC	F1	EE
CFS	RUS	**0.931**	0.746	**0.564**	RUS	0.92	0.732	0.475	SO	**0.903**	**0.729**	0.393
GainRatio	Smote	0.881	0.614	0.386	Smote	0.889	0.71	0.38	Smote	0.753	0.689	0.395
Relief	RUS	0.926	**0.767**	0.56	RUS	0.917	0.737	**0.575**	RUS	0.776	0.713	**0.45**
Merge	-	0.918	0.743	0.539	-	**0.926**	**0.745**	0.435	-	0.774	0.689	0.39
No_FS	-	0.849	0.554	0.444	-	0.749	0.672	0.381	-	0.753	0.708	0.053

We can note in Table 5 that the Bayesian Network and Logistic Regression are more adequate techniques to detect fraud in this scenario than Decision Tree. The Feature Selection techniques Relief and CFS were more adequate to feature

selection for fraud detection. The resampling method Random Undersampling, when correctly calibrated the fraud ratio, is a good resampling method to reduce the class-imbalance on feature selection.

From Table 5 it is possible to infer that all models that use feature selection achieved significant gains over the model *No_FS*. Thus, this analysis confirms the importance of feature selection in fraud detection. The method $Merge$ can be a good strategy to select attributes to Logistic regression classification.

The model that combines Relief, using the random undersasmpling, and Logistic Regression, as classification technique, achieved the best economic gains . This model achieved economic gains of 57.5% in comparison to the real scenario of the Corporation, using just 8% of available features. In other words, if this model were used for fraud detection it would save 57.5% of fraud financial losses.

6 Conclusion

In this paper we analyze the use of feature selection to fraud detection in Web transactions. To perform this analysis we used three feature selection techniques and evaluate how the imbalance between classes can difficult this task. We used 7 resampling strategy in this step, including a resampling method created by us.

Through the results of this research, we can validate our hypothesis that imbalance between classes reduces the effectiveness of the feature selection techniques to fraud detection. As a possible solution to this problem we use distinct resampling strategy in the feature selection step. Our fraud detection model that can improve in 57.5% the financial results obtained by the corporation.

In adition, in this work we show some interesting behaviors about feature selection to anomaly detection and resampling on feature selection step. The main conclusions are:

- The feature selection technique Gain Ratio was more sensitive to the imbalance between classes, while the Relief technique proved less sensitive to imbalance.
- Increasing the fraud ratio in resampling methods do not imply in increasing linearly the effectiveness of feature selection.
- Each resampling method is best suited to a distinct fraud ratio.
- Our method Sampling Outlier is a good alternative when the cost of performing a classification to choose the best fraud ratio is high.

References

1. Bhattacharyya, S., Jha, S., Tharakunnel, K., Westland, J.C.: Data mining for creditcard fraud: A comparative study. Journal Decision Support Systems. **50**(3) (Fevereiro 2011) 602–613
2. Kim, K., Choi, Y., Park, J.: Pricing fraud detection in online shopping malls using a finite mixture model. Electronic Commerce Research and Applications **12**(3) (2013) 195 – 207
3. Richhariya, P., Singh, P.K.: Article: A survey on financial fraud detection methodologies. International Journal of Computer Applications **45**(22) (May 2012) 15–22

4. Ravisankar, P., Ravi, V., Rao, G.R., Bose, I.: Detection of financial statement fraud and feature selection using data mining techniques. Decision Support Systems **50**(2) (2011) 491 – 500
5. Kamal, A.H.M., Zhu, X., Pandya, A., Hsu, S., Narayanan, R.: Feature selection for datasets with imbalanced class distributions. International Journal of Software Engineering and Knowledge Engineering **20**(02) (2010) 113–137
6. Keele, S.: Guidelines for performing systematic literature reviews in software engineering. In: Technical report, Ver. 2.3 EBSE Technical Report. EBSE. (2007)
7. Chen, X.w., Wasikowski, M.: Fast: a roc-based feature selection metric for small samples and imbalanced data classification problems. In: Proc. of the 14th ACM SIGKDD Conf. on Knowledge discovery and data mining, ACM (2008) 124–132
8. Van Hulse, J., Khoshgoftaar, T.M., Napolitano, A., Wald, R.: Threshold-based feature selection techniques for high-dimensional bioinformatics data. Network modeling analysis in health informatics and bioinformatics **1**(1-2) (2012) 47–61
9. Cuaya, G., Munoz-Meléndez, A., Morales, E.F.: A minority class feature selection method. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Springer (2011) 417–424
10. Alibeigi, M., Hashemi, S., Hamzeh, A.: Dbfs: An effective density based feature selection scheme for small sample size and high dimensional imbalanced data sets. Data & Knowledge Engineering **81** (2012) 67–103
11. Chawla, N.V.: Data mining for imbalanced datasets: An overview. In: Data mining and knowledge discovery handbook. Springer (2005) 853–867
12. Van Hulse, J., Khoshgoftaar, T.M., Napolitano, A., Wald, R.: Feature selection with high-dimensional imbalanced data. In: Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on, IEEE (2009) 507–514
13. Maldonado, S., Weber, R., Famili, F.: Feature selection for high-dimensional class-imbalanced data sets using svm. Information Sciences **286** (2014) 228–246
14. Hall, M.A.: Correlation-based feature selection for discrete and numeric class machine learning. In: Proc. of the 17th Intl. Conference on Machine Learning. ICML '00, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (2000) 359–366
15. Kelleher, J., Namee, B.M.: Information based learning (2011)
16. Liu, H., Motoda, H., eds.: Computationall Methods of Feature Selection. Chapman and Hall (2008)
17. Mani, I., Zhang, I.: knn approach to unbalanced data distributions: a case study involving information extraction. In: Proceedings of workshop on learning from imbalanced datasets. (2003)
18. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. Journal of artificial intelligence (2002) 321–357
19. Maes, S., karl Tuyls, Vanschoenwinkel, B., Manderick, B.: Credit card fraud detection using bayesian and neural networks. Vrije Universiteir Brussel (2001)
20. Dobson, A.J.: An Introduction to Generalized Linear Models. London:Chapman and Hall (1990)
21. Salzberg, S.: C4.5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993. Machine Learning **16**(3) (1994) 235–240
22. Lima, R.A.F., Pereira, A.C.M.: Fraud detection in web transactions. In: WebMedia. (2012) 273–280
23. Friedman, M.: A comparison of alternative tests of significance for the problem of m rankings. The Annals of Mathematical Statistics **11**(1) (1940) 86–92
24. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) (1995) 289–300
25. Drummond, C., Holte, R.C., et al.: C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In: Workshop on learning from imbalanced datasets II. Volume 11., Citeseer (2003)