BABEŞ-BOLYAI UNIVERSITY

PHD SCHOOL OF THE FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

PHD THESIS

– SUMMARY –

Cost-effectiveness of the (semi-)automated fraud detection systems in motor insurance in Romania

Supervisor:

Prof. Cristina CIUMAŞ, PhD

PhD student:

Botond **BENEDEK**

2020

ABSTRACT

Studies since 1990 attempt to define and quantify the motor insurance fraud, identify key fraud indicators and suggest fraud detection approaches. However, as of today, there is no widely accepted list of fraud indicators or a general model. Therefore, the aim of this study is three-fold. First of all, to propose a systematic and comprehensive literature review on the subject. After thoroughly presenting the motor insurance fraud detection and cost sensitive decision making literature (including the performance measurement difficulties in the case of imbalanced datasets), we suggest an improved cost saving calculation method. Finally, we compared the cost saving ability of 77 (40 respectively) fraud detection methods using Romanian corporate data. The findings clearly show that data mining techniques are becoming more and more popular but also reveal the missing of cost-sensitive approaches and the lack of reliable up to date datasets in academic research. Taking into accounting the current Romanian context, the results show that only a small percent of the currently available motor insurance fraud detection methods can be applied in a truly lucrative manner.

Keywords: *motor insurance; insurance fraud; fraud detection; literature review; data mining; cost-sensitive decision making*

TABLE OF CONTENTS

Abstract..5
Acknowledgements 6
Abbreviations 9
List of Tables 10
List of Figures..11
INTRODUCTION...12
CHAPTER 1 – The Motor Insurance Industry 15
1.1 The emergence of motor insurance 15
1.2 Motor insurance policy types and available coverages… 16
1.2.1 General types of motor insurance policies...17
1.2.2 Available coverages in the developed markets 17
1.3 Liabilities and legislation 19
1.3.1 General legal liabilities in motor insurance 19
1.3.2 Duties after an accident – in general..20
1.3.3 Legislation in the EU 21
1.3.4 Motor liability insurance regulation in Romania 22
1.4 The motor insurance market..23
1.4.1 The motor insurance market in the USA, Canada, Japan and China 23
1.4.2 The European motor insurance market 26
1.4.3 The Romanian motor insurance market 29
CHAPTER 2 – Financial, Insurance and Motor Insurance Fraud......................................31
2.1 Financial fraud...31
2.1.1 Presentation of the major financial fraud types 32
2.1.2 Classification of data mining methods and techniques used to detect and prevent financial fraud...34
2.2 Insurance fraud 36
2.2.1 Insurance fraud schemes...38
2.2.2 Cost of insurance fraud 41
2.2.3 Consequences of insurance fraud illustrated with examples 46
2.2.4 Combatting insurance fraud 48
2.2.5 Brief presentation of the special investigation units 54

2.3 The motor insurance fraud 59
2.3.1 Motor insurance fraud schemes 59
2.3.2 The prevalence and the costs of motor insurance fraud....................................61
CHAPTER 3 – Motor Insurance Fraud Detection 63
3.1 The related literature 63
3.2 The general process of motor insurance fraud detection...71
3.3 Data mining methods used in the motor insurance fraud detection 72
3.4 Motor insurance fraud indicators 83
3.3.2 Motor insurance fraud indicators in the literature...83
3.4.2 Ranking of the non-bodily injury related fraud indicators................................88
3.4.3 Aggregated ranking of the non-bodily injury related fraud indicators 94
3.5 Imbalanced datasets and performance measurement 97
3.5.1 The imbalanced dataset problem 98
3.5.2 Measuring the classification performance 102
3.6 Systematic presentation of the studied researches 105
CHAPTER 4 – Cost-effectiveness of the Motor Insurance Fraud Detection Methods in Romania 114
4.1 The proposed cost saving calculation method...114
4.1.1 Cost saving calculation methods in the literature 116
4.1.2 The proposed cost saving calculation method 119
4.2 A case study on a Romanian sample..120
4.3 Heatmaps with the cost saving ability of the detection methods 124
CONCLUSIONS 127
Bibliography 133
Electronic references 145
Appendix...146

Abbreviations

ALFA,	Agency Against Insurance Fraud
CCI,	Comprehensive Car Insurance
DT,	Decision Tree
FBI,	Federal Bureau of Investigation
FCM,	Fuzzy C-Means Clustering
FFD,	Financial Fraud Detection
FN,	False Negatives
FP,	False Positives
FSA,	Financial Supervisory Authority
GA,	Genetic Algorithm
GMDH,	Group Method of Data Handling
IAATI,	International Association of Auto Theft Investigators
IFB,	Insurance Fraud Bureau
IFED,	Insurance Fraud Enforcement Department
IFR,	Insurance Fraud Register
IPD,	Insurance & Pension Denmark
KNN,	K-Nearest Neighbors
MTPL,	Mandatory Third Party Liability Insurance
MLP,	Multilayer Perceptron
RF,	Random Forest
SIU,	Special Investigation Unit
SVM,	Support Vector Machine
TP,	True Positives
TN,	True Negatives
UNESPS,	Spanish Union of Insurance and Reinsurance Companies
m,	million
bn,	billion

List of Tables

Table 1: Data mining methods (with their strengths and limitations) used in the motor insurance fraud detection 81

Table 2: Non-bodily injury related fraud indicators, their occurrence frequency and significance 85

Table 3: Bodily injury related fraud indicators and their occurrence frequency and significance 87

Table 4: Ranking of the non-bodily injury related fraud indicators 89

Table 5: The aggregated ranking of the non bodily injury related fraud indicators 94

Table 6: The Kendall rank correlation coefficients with their p-values 97

Table 7: Confusion matrix of a binary classifier 102

Table 8: Most common performance metrics derived from the confusion matrix 104

Table 9: Accuracy, sensitivity and specificity of the applied techniques 105

Table 10: Motor insurance fraud detection practices based upon the detection/classification algorithm 108

Table 11: Datasets used in motor insurance fraud detection researches 109

Table 12: Annual distribution of the motor insurance fraud detection researches 111

Table 13: Distribution of the motor insurance fraud detection articles by journal title 112

Table 14: Confusion matrix of a binary classifier 114

Table 15: Cost matrix of a fraud detection method 118

Table 16: Cost saving potential of the detection methods 123

Table 17: Cost saving potential of detection methods in case of two Romanian insurance companies 124

List of Figures

Figure 1: The general motor insurance fraud detection process 71
Figure 2: Example graphical representation of a Bayesian belief network showing the causal relationship between hypotheses H6 and H3 73
Figure 3: A simple neural network with three input variables, two outputs, and a single hidden layer 74
Figure 4: Two dimensional example showing the margin of separation between support vectors and hyperplane 76
Figure 5: Standard genetic algorithm process 77
Figure 6: Binary decision tree where the input variables X_i will be a list of input predicates and the outputs Y_i will classify the results into one of multiple classes 78

INTRODUCTION

Motor insurance fraud is an issue that has significant consequences in both the insurance industry and daily life. Fraud can reduce confidence in the industry, destabilize economies, and affect people's cost of living. Surveys and studies since 1990 attempt to quantify the size of these kind of activities, identify key fraud indicators, suggest methods and models that can indicate or forecast the potential fraudulent claims in the motor insurance industry. However, as of today, there is no widely accepted list of fraud indicators or a general model. In most countries the actual size of this kind of activity is also unknown. Moreover, most of the authors in the academic literature used their own datasets with different potential fraud indicators and applied them on a self-developed model whose performances are measured by an arbitrary metric.

The relevance and importance of the topic is underlined by the many reports published related to the topic, which state, that insurance fraud affects 10-20% of all the contracts even in the most developed countries. Therefore, insurance fraud is the second most widespread and large-scale financial fraud, after tax fraud. A report from the FBI states that an average US family has $400-$700 extra expenses each year caused by the insurance fraud (FBI, 2011). If we consider the motor insurance only, in the USA and Western Europe, 7-10% of policies are affected, while in Central and Eastern Europe regions the number is 10-20%. In Romania, there is no official statistics regarding motor insurance fraud, however, Thomas Brinkmann, country manager of Friss, declared in 2017, at the International Insurance-Reinsurance Forum (FIAR), that at least 15% of motor insurance policies are affected in Romania. This seems to be backed up by the 2016 FSA report as well, which estimates 16%. From a financial point of view, this means that the scale of the fraud is 1.7-1.8 billion RON yearly, which is equivalent to around 400 million euros.

Therefore, the aim of this study is three-fold. First of all, similarly to (Ngai, et al., 2011), (West & Bhattacharya, 2016), and (Abdallah, et al., 2016), this study attempts to present a systematic and comprehensive academic literature review strictly focusing on the motor insurance fraud detection (contrary to the aforementioned publications whose focus were the general financial fraud detection). 44 journal articles and 8 conference proceedings on the subject published between 1990 and 2019 will be analyzed based on key aspects such as the dataset, the fraud indicators, the detection algorithm or the performance of the detection methods.

After thoroughly presenting the motor insurance fraud detection literature (including the cost sensitive decision making literature and the performance measurement difficulties in the case of imbalanced datasets, such as the motor insurance fraud datasets), we suggest an improved cost-saving calculation method (which takes into account the real claim screening process steps with their costs) in order to offer an easily usable metric to compare the different approaches from the literature.

Finally, we compare the cost saving ability of 77 (respectively 40) motor insurance fraud detection methods using Romanian corporate data. By doing this, we will be able to answer the main question of the research, namely that the currently available motor insurance fraud detection methods can be applied in a truly lucrative manner.

Of course, during the analysis of the aims and questions enumerated above, more and more topic-related questions arise. Some of the questions are: which are the most important fraud indicators or the most efficient fraud detection methods in specific market and economic circumstances? We could also list here the questions related to the literature, such as: to what extent are the fraud detection datasets presented in the literature representative, considering the current market environment?

In order to answer these secondary research questions, we complemented the main study with more subchapters, which are not strictly necessary (in order to answer the main research question), however, they are connected to the topic of the study and offer great help in finding these answers. For instance, the chapter in which we discuss fraud indicators and build an aggregated raking based on different points of view or the chapter presenting heatmaps, which aids in comprehending the performance of specific fraud detection methods in different market and financial circumstances, thus easing the future work of any insurance company or scholar.

In addition to the topics listed so far, the research seeks to provide the reader with a general overview of financial and insurance fraud, its types, related legislation and last but not least their economic implications.

The rest of the study is organized as follows: in the first chapter, after a brief introduction about the emergence of motor insurance, the laws currently regulating the motor insurance, with emphasis put on EU and Romanian laws and the various types of motor insurance policies and available coverages will be presented. The chapter also offers insight into the current trends of the motor insurance market.

In the second chapter different insurance fraud techniques will be presented, along with various data mining methods used for prevention and detection. The chapter dedicates special attention to insurance fraud. Beside these, the techniques used by the antifraud departments to combat insurance fraud will also be presented. Finally, in the last part of the chapter a detailed presentation of the motor insurance fraud schemes along with the size and expense of such activities in certain countries will be offered.

In the third chapter we focus especially on motor insurance fraud and their detection methods. Besides presenting the related literature, we will also analyze the most important fraud indicators, the most frequently used data mining and artificial intelligence-base fraud detection methods, accentuating both their pros and cons. Additionally, we will review the datasets present in the literature and the specific measuring metrics used for measuring performance.

In the fourth chapter we will analyze the studied fraud detection methods' cost saving ability in the current Romanian context. By considering the performance of the fraud detection methods depending on the economic context and the potentially differing industrial requirements, we also build up heatmaps to simulate all the detection methods' performances.

Finally, the last chapter summarizes the results and presents the limitations of the study.

SUMMARY: CHAPTER 1 – THE MOTOR INSURANCE INDUSTRY

When the motor vehicle was invented – at the end of the 19th century – the need for insurance did not appear instantly. In some countries, like the United Kingdom, there was a safety law in use: a person had to walk in front of a moving vehicle with a red flag to warn its drivers of hitting the pedestrians. In 1896 they revoked this regulation and the British motor industry slowly started to develop. In 1898 the first motor insurance policy was issued, even though comprehensive insurance appeared only a few years later (Benetton, 1994).

Since then, the motor insurance industry became one of the biggest insurance industry. For example, in 2018, the number of registered vehicles reached more than 273 million in the USA, while in Europe were more than 215 million. Furthermore, only in the US, the private passenger automobile insurance premium exceeded the 148 billion dollars in 2019.

Because of the above mentioned rapid development of the motor insurance industry, in the first chapter we present some of the key numbers of the motor insurance market, such as the number of policies, accidents, fatalities or the average insurance premium, with special emphasis on the European and Romanian markets.

In addition to presenting the key numbers of the insurance markets, in the first chapter the various types of motor insurance policies and the available coverages are also presented. Based on (Sims, 2019) the following coverages are the most commonly used in order to protect the insured automobile owners, the third parties and their properties: medical payment coverage, liability coverage, physical damage coverage and uninsured motorists coverage. Each of these coverages with their specifications are presented in the first chapter.

After presenting the diverse types of motor insurance policies and their specifications, it is essential to clarify the responsibilities resulting from insurance contracts and its associated legal regulations, at the very least in the case of the European Union and Romania. Obviously, the regulations themselves or the methods which are used to determine compensations, differ in each country, however a few common main principles can be found. This is especially true in case of the European market, where the regulations are harmonized at the level of the European Union.

SUMMARY: CHAPTER 2 – FINANCIAL, INSURANCE AND MOTOR INSURANCE FRAUD

Before all, the second chapter of the thesis offers a clear and detailed definition of concepts, such as: fraud, fraudulent claim, financial fraud, insurance fraud and fraud detection. After that, in the first part of the second chapter, different financial and insurance fraud techniques are presented, along with various data mining methods used for prevention and detection. Based on the FBI – Financial Crimes Report, financial fraud can be split into the following seven categories: credit card fraud, securities and commodities fraud, financial statement fraud, insurance fraud, mortgage fraud, money laundering (FBI, 2011).

In the second part of this chapter we analyze those fraud detection and prevention methods, which are used the most by the financial institutions. (Ngai, 2011) splits these methods in the following six categories: classification, clustering, prediction, outlier detection, regression and visualization. The general operating principles of these six categories, their advantages and disadvantages are also presented in this section.

By focusing on the goals of the current thesis, we dedicated special attention to insurance fraud, by presenting how they operate, their characteristics and the related statistics and trends. Besides, we concisely presented the techniques which are used by the antifraud departments to combat insurance fraud.

Finally, in the last part of the chapter, we offered a detailed presentation of the motor insurance fraud, along with the size and expense of such activities in certain countries. Based on the literature, on corporate surveys and on the reports of different investigative bodies, we identified 14 different motor insurance fraud schemes. These are the following: ditching, past posting, vehicle repair, vehicle smuggling, phantom vehicles, staged accidents, inflated damages, vehicle identification number (VIN) switch, rental car fraud, property schemes, inflated inventory, phony or inflated thefts, paper boats and arson for profit. In addition to the identification of motor insurance fraud schemes type, the prevalence and the costs of motor insurance fraud are presented in detail.

SUMMARY: CHAPTER 3 – MOTOR INSURANCE FRAUD DETECTION

In the third chapter we focus especially on motor insurance fraud and its detection methods. First of all, the related literature is presented. We managed to identify 52 scientific articles and conference proceedings (since 1990) which are indexed by the Web of Science and are related to the motor insurance fraud domain. A part of these studies approached the motor insurance fraud domain theoretically (such as (Weisberg & Derrig, 1991); (Derrig & Ostaszewski, 1995); (Derrig, 2002); (Tennyson & Salsas-Forn, 2002); (Warren & Schweitzer, 2018)), while others offer a specific fraud detection system (for example (Belhadji & Dionne, 1998); (Artís, et al., 1999); (Artís, et al., 2002); (Viaene, et al., 2002); (Viaene, et al., 2005); (Viaene, et al., 2007); (Pinquet, et al., 2007); (Phua, et al., 2004); (Bermúdez, et al., 2008); (Wilson, 2009); (Sundarkumar, et al., 2015); (Sundarkumar & Ravi, 2015); (Hassan & Abraham, 2016); (Subudhi & Panigrahi, 2017); (Wang & Xu, 2018); (Majhi, et al., 2019); (Zelenkov, 2019)).

After the presentation of the literature on motor insurance fraud detection, the fraud detection process is presented. At this point, it is important to mention, that none of the currently existing anti-fraud technologies can by themselves prove fraud. High level investigations are crucial in tackling fraud, and currently technology cannot replace an experienced and skilled investigator. However, technology is a great aid in the so-called early claim screening process, which can be done with a (semi-)automated indicator-based fraud screening model. This model can implement various different techniques, from the simplest if-else rule system, to the most complex, hybrid statistical and computational methods. Therefore, the next part of the third chapter offers an in-depth presentation of the currently existing methods with their advantages and disadvantages.

During the development of any (semi-)automated fraud detection model, it is of utmost importance to define the potential fraud indicators correctly. Because of these, in the next part of the chapter we presented the most important bodily and non-bodily injury related fraud indicators. However, we discussed only the non-bodily injury related indicators in-depth. The reason being that these types of indicators are available during the early claim screening process. The indicators related to bodily injuries or hospital treatment and recovery are not available in this phase. Regarding non-bodily injury indicators, we compiled a list of 92 fraud indicators. While creating this list, we took into consideration the related literature, as well as the opinion of industry experts from Romania. Based on the rankings created from different points of view, we created an aggregated list as well, which could be used as a starting point for further studies or works for any scholar or industry expert.

After presenting the various fraud detection methods and fraud indicators, the performance measurement of the motor insurance fraud detection methods is presented. The performance measurement of fraud detection methods (including, but not limited to motor insurance frauds) is a binary classification problem. However, the performance measurement of motor insurance fraud detection methods is special, since the imbalanced (skewed) dataset problem and its impact on performance measurement has to be considered as well.

Finally, in order to summarize the knowledge with respect to the above presented challenges from the 52 studies regarding the motor insurance fraud detection, in the last part of the third chapter, we present different classifications and summary tables. With this compartmentalization we are able to showcase trends in research methods, highlighting the ones that have been successful and any possible factors yet undiscussed. Moreover, the performance of the detection methods, the different datasets and the most important researches and the journals that publish them become easy to review.

SUMMARY: CHAPTER 4 – COST-EFFECTIVENESS OF THE MOTOR INSURANCE FRAUD DETECTION METHODS IN ROMANIA

In the fourth chapter, based on the challenges related to motor insurance fraud detection presented before, we analyzed the cost-effectiveness of the currently known detection methods. It is obvious, that from a corporate point of view there is a need to measure the benefit (cost saving) of any detection methods. In order to calculate this benefit generated by the usage of certain detection methods, our start point was the method proposed by (Phua et al., 2004) and (Viaene, et al., 2007).

It is important to note, that in both cases the authors are calculating with the assumption that true negative cases (when an honest claim is labelled as honest) do not have extra expenses for insurance companies. However, after interviewing several Romanian industry experts and anti-fraud department managers, we came to the conclusion that in practice, these true negative cases have an extra cost as well. The expense can be either the salary of an employee for fulfilling certain extra tasks related to the verification process of a claim's legitimacy, or simply the cost of a service used. Even though these expenses are only a few euros, considering the great amount of claims, they can amass a huge value.

Similar to the true negative cases, the cost calculation for true positive cases (when a fraudulent claim is labelled as fraudulent) is different in practice than it is in the literature. The reason being, that in the literature, in case of a true positive case the insurance company does not pay to the insured and the only expense is the one incurred during the investigation. However, in the industrial practice, the situation is different. Many studies including (Derrig & Ostaszewski, 1995), (Weisberg & Derrig, 1998) have proven that, motor insurance fraud is mainly composed of overvalued contracts (build-up). When a build-up fraud is detected the insurance company usually pays a reduced compensation. Our interviewees have confirmed that it is really rare for the insurance company to fully deny paying compensation (they usually offer a smaller amount, which they consider correct). This has many explanations, some of them are the lengthy and costly juridical process or the negative marketing.

Due to the problems presented above, in the fourth chapter we proposed a new cost calculation method, which takes into account all the expenses arising during the fraud detection process. The cost-effectiveness of the fraud detection methods presented in the earlier mentioned 52 scientific articles and conference proceedings which are related to the motor insurance fraud domain, were calculated using the proposed cost calculation method. It is important to note, that a part of these studies approached the motor insurance fraud domain theoretically and do not offer a specific fraud detection system. Because of these, we were able to identify only 77 different statistical, data mining or artificial intelligence-based fraud detection methods. However the 77 detection methods still contain some approaches where the performance under 100% real circumstances is questionable[1], thus we also analyzed a shrunk list made up of only 40 models (models for which the detection method's performance was tested on imbalanced real world datasets). Table 1 summarizes the 77 (respectively the 40) models' cost saving ability in three different cases.

Table 1: Cost saving potential of the detection methods in case of Romania

	Most likely case		Worst case		Best case	
	77 models	40 models	77 models	40 models	77 models	40 models
Fraud Rate	10%	10%	5%	5%	20%	20%
Average Cost Per Claim	2420	2420	2420	2420	2420	2420
Average Cost Per Investigation	145	145	193	193	97	97
Average Cost Of Normal Claim	12	12	12	12	12	12
Average Saving In Case Of Identified Fraudulent Claims	485	485	315	315	1213	1213
Nb. of unprofitable models	41	28	70	39	0	0
Nb. of models with cost saving:						
>25% of max possible	26	5	5	1	71	35
>50% of max possible	17	1	3	1	63	32
>75% of max possible	11	1	1	1	34	9
>90% of max possible	0	0	0	0	7	0

Source: Own editing

The proposed cost saving calculation method was also applied in the case of two Romanian insurance companies in order to verify the profitability of the 77 (respectively the 40) models in case of certain companies (see Table 2). Claims for third party liability insurance (MTPL) and comprehensive car insurance (CCI) were analyzed separately for both insurance companies. Table 2 summarizes the results.

[1] For instance, (Artís, et al., 1999) and (Artís, et al., 2002) used a real world set of data during the study, but the sample was not exactly random, as it was made up of 997 fraudulent and 998 legitimate claims. (Wilson, 2009) used 49 fraudulent and 49 lawful claims. (Wang & Xu, 2018) worked with 1660 fraudulent and 1660 legal claims. The true performance of the mentioned methods on imbalanced sets of data is uncertain.

Table 2: Cost saving potential of detection methods in case of two Romanian insurance companies

	Company A		Company B	
	MTPL	CCI	MTPL	CCI
Fraud rate	10%	3%	12%	8%
Average Cost Per Claim	1250	870	2360	1160
Average Cost Per Investigation	205	97	160	160
Average Cost Of Normal Claim	12	12	0	0
Average Saving In Case Of Identified Fraudulent Claims	160	242	675	338
No. of unprofitable models	40	33	22	36
No. of models with cost saving:				
>25% of max possible	0	4	7	1
>50% of max possible	0	1	2	1
>75% of max possible	0	1	1	1
>90% of max possible	0	0	0	0

Source: Own editing

The results in both cases clearly show that only a small percent of the currently available motor insurance fraud detection methods can be applied in a truly lucrative manner.

Of course, seeing the previous results, questions like what happens if we use other inputs in case of the calculations, or fraud detection methods are also unprofitable in case of other insurance companies can arise instantly. In order to answer these questions, the cost-effectiveness of the 40 methods were calculated in 3534 different situations covering all probable input combinations. The results of these scenarios are presented with the help of heatmaps.

CONCLUSIONS

An essential part of any study is the construction of a competent classification framework and an assembly of appropriate literature. The research of motor insurance fraud detection is no exception. Despite the recognized significance of data extraction methods in the detection of motor insurance fraud, the aforementioned framework or an organized review of their usage in motor insurance fraud detection studies is missing. Because of this, first of all, in this study, we carried out a review of academic articles and delivered an extensive bibliography and categorization framework for the usage of data mining and machine learning in motor insurance fraud detection. In this sense, our goal was to provide valuable information for both scholars and practitioners of the respective fields in which certain techniques can be used in motor insurance fraud detection, and to report and assemble an organized review of the flourishing publications regarding motor insurance fraud detection.

Regarding this part of the study (comprehensive literature review), we would like to emphasize certain general difficulties the authors faced and that are essential in understanding the literature. First of all, studies have shown that there is no generally accepted definition of automobile/motor insurance fraud. The most common definition (worldwide) is probably the Massachusetts Regulation (211 CMR 93.03) which defines fraudulent claims as "claims submitted with the intent of receiving a larger payment from the insurer than the amount, if any, to which the claimant is entitled under the policy, including claims for (i) non-existent losses; (ii) amounts in excess of actual losses; or (iii) incidents which the claimant has arranged in an effort to receive an insurance payment" (Massachusetts Regulation, 1993). Nevertheless, this definition does not contain other types of fraud, such as the so-called misrepresentation or the recklessness of the insured person (which is a consequence of the very existence of the insurance). As presented in (Wilson, 2009), the so-called "misrepresentation" is less deliberate. "The defrauders may rationalize that there is nothing wrong with their actions. An example is a parent who's child just received a driver's license and, knowing that the family's insurance rates will skyrocket, allows their child to operate a vehicle without notifying the insurance company that the 16 year old driver has not yet been rated" (Wilson, 2009, p. 2). Similarly, recklessness does not necessarily mean an intentional attempt to cause a loss. However, the insured person engages in actions that he/she would not normally engage in if the possible adverse outcome was their personal burden. An example is when people have full coverage (which covers not only their liability but damage to their car as well) and because of it they become less careful. In addition, different countries have different legislation which leads to different insurance products. The different products lead to different insured behavior and render certain motor insurance fraud types meaningless in some countries. An example could be the fraudulent behavior in case of bodily injury claims which are important on the US market, but seem less relevant in the European context (because of the compulsory health insurance) (Artís, et al., 1999). In summary, the lack of precise definition of motor insurance fraud could be explained by the variety of motor insurance products, by the diversity of legislation from country to country and by the insured person's behavior which can also be very diverse.

Another general problem in motor insurance fraud detection is the lack of reliable data. As described in (Brockett, et al., 1998) "the reliability of each dependent variable couldn't be verified in the real world, due to data limitations. Unlike other fraud detection problems, such as credit card fraud, most automobile bodily injury claims cannot ultimately be verified. It is either too costly or impossible to determine and classify without doubt a fraudulent bodily injury claim unless a reliable court decision is available. However, insurance companies tend not to resolve a claim in this manner because it is both risky and costly" (Brockett, et al., 1998, p. 247). As a result, datasets (not just the bodily injury claims, but all motor insurance claims) used in fraud detection in most cases contain subjective indicators and classifications. This problem was presented in detail by (Weisberg & Derrig, 1991) and by (Derrig & Ostaszewski, 1995). They found that professionals handling different bodily injury claims had ambiguous perceptions of which bodily injury claims constitute fraud (Weisberg & Derrig, 1991). Additionally, (Derrig & Ostaszewski, 1995) showed the lack of concordance among experts with respect to which claims were fraudulent. Nevertheless, these subjective indicators and classifications made by industry experts are the only available data sources. Continuing with data issues, we would also like to emphasize that only a few datasets were studied. From the 14 datasets in 4 cases the years of collection are unknown and only one dataset contains motor insurance claims from the past ten years, highlighting the issue of missing reliable data in academic research. Furthermore, the biggest part of these datasets cover only three markets, namely they were collected on the Spanish, Chinese and USA markets. Moreover, most of the datasets contain only a few thousand claims, while the real annual number of the motor insurance claims is far greater, therefore the representativeness of the datasets is also questionable.

After the literature review, the last important conclusion of this study can be drawn: the lack of cost-sensitive approaches. Despite the purpose of motor insurance fraud detection process being the decreasing of expenses of insurance companies caused by motor insurance fraud, only a few studies of this type exist. Most of the studies focused on minimizing the error rate (misclassification) rather than on the total costs or on the profitability of the investigation process. Only (Phua, et al., 2004), (Viaene, et al., 2007) and (Zelenkov, 2019) proposed cost-sensitive approaches.

After the comprehensive literature review, along with the different motor insurance fraud schemes, in this study we presented an aggregated ranking of the most important motor insurance fraud indicators. While creating this aggregated ranking, we took the indicators presented in the literature into consideration, as well as the indicators recommended by the industry experts during our interviews and the ones present in our survey. Thus, the aggregated ranking can serve as a really good starting point for any future scholar or insurance company who wants to look into any part of motor insurance fraud detection. It can be helpful when insurance companies are working on their own accident reporting statements available on their website, or on their online claim reporting. It could also aid in deciding which accident-related data should be registered (for data analysis performed later on). It offers help for any scholar who is gathering data for a research or during the development of a model, when the initial input parameters have to be decided.

The last part of the study is related to the performance measurement of fraud detection methods in case of imbalanced datasets and the cost saving ability of these methods. The study shows that the currently utilized performance indicators either cannot be used on imbalanced datasets (at least not in the ways presented in the literature so far) or they render the comparison of methods really hard, as it is impossible to describe the performance of a method with a single indicator. Moreover, these performance indicators do not take into consideration the expenses arising during the fraud detection process. Therefore, the current study offers a cost-based performance evaluation, instead of the minimization of misclassification errors. While there is an approach proposed by (Phua, et al., 2004) which considers the lowest number of expenses, during the industry interviews the following was noted: not only the fraud detection process is different in the industry practice and the literature, the related expenses vary as well. By taking into account the real industry procedure and the related expenses, the current study shows that applying most models recommended by the literature in the most likely economic and market environment is not profitable at all.

As the economic and market environment is changing constantly, we created more different simulations in order to analyze the cost saving ability of certain models. We paid special attention to the Romanian environment, furthermore, we used the data provided by two insurance companies and analyzed the cost saving ability of the models on the parameters specific to those insurance companies as well (examining the CCI and MTPL claims separately). The results clearly show that only a small percent of the studied motor insurance fraud detection approaches can be applied in a truly lucrative manner.

Beyond the Romanian case study, using heatmaps, we compared the cost saving ability of the studied motor insurance fraud detection methods, thus offering a possibility for insurance companies and scholars interested in the topic to simply and efficiently choose the most optimal method. We took into account the parameters specific to the given research environment as well (including the available fraud indicators, the investigation costs or the average cost of a claim).

Motor insurance fraud detection challenges and future directions

Motor insurance fraud detection is a developing field where it is best to outpace the perpetrators. Furthermore, it is obvious that many aspects of intelligent fraud detection are still to be looked into. In the following section we will showcase some of the most important open issues related to motor insurance fraud detection and also recommend areas for possible upcoming research.

Considering that it is a categorization issue, motor insurance fraud detection suffers from the same problems as other similar issues. The selection of features has a huge effect on the success rate of any categorization method. One of the greatest advantages of the computational intelligence and data extraction techniques is their capability of adapting to fit the domain of the problem. Only a small part of the current studies has used any kind of customization or fine-tuning for certain issues, even though fine-tuning is a significant factor regarding the performance of an algorithm. For instance, the number of nodes and internal layers within a neural network has a significant effect on accuracy as well as on computational performance. Likewise, the selected kernel function will significantly modify the success rate of a support vector machine and criteria like the fitness function, crossover technique, and the likelihood of mutation will affect the outcomes of a genetic programming algorithm. Studies on customizing or boosting computer-based techniques are necessary to achieve a true perspective of each method's ability.

Financial data being private, has led to organizations holding back fraudulent information. This affected the analyzed fraud types as well as the datasets in use. In the presented publications a significant part of the motor insurance fraud simulations was made up of less than a few thousand samples, usually with similar amounts of legal and fraudulent specimens. This is the exact opposite of the reality of the problematic field, where legitimate transactions outweigh fraud cases by far. Having few samples may induce biases in the data which do not precisely depict real world cases. Further studies are required with lifelike samples for a proper representation of each technique's performance.

Contrary to most classification problems, motor insurance fraud detection solutions need to have the capability of processing active attempts to avoid them. As the aforementioned techniques are constantly developing, fraudsters are also enhancing their methods. Thus, the capability of developing detection models faster than the fraudsters is a must for motor insurers. Researchers have looked into models for adaptive categorization, but in-depth studies are needed to fully enhance these to be used in real-world fraud detection issues.

As discussed earlier, the expense of fraud for businesses is immense and it also has associated costs: systems necessitate a fair amount of maintenance and computational power, along with experts who must be employed to constantly monitor them and look into identified cases with fraud suspicion. The cost of misclassification of a legitimate claim as fraudulent (false positive), is usually far less than the opposite (false negative). Only a few studies have been carried out on the disproportionate nature of the mentioned expenses. Taking into account the accuracy, sensitivity or specificity of each fraud detection technique, the main goal should be reaching an optimum depth for each method, resulting in the smallest cost. Studies explicitly directed to finding this balance would increase the practical worth of motor insurance fraud detection greatly.

Regarding the cost saving ability of the motor insurance fraud detection methods, it would be useful to examine the combination of our suggested method and the approach suggested by (Phua, et al., 2004); handling the total damage (where the amount of compensation is equal to the value of the insured vehicle) and the partial damage claims separately. Obviously, an industry dataset which separates these two types is necessary. Nevertheless, this is the most accurate way to determine the cost saving ability of the motor insurance fraud detection methods, which is a required condition of applicability in the industry.

Finally, it would be interesting to analyze and compare the cost saving ability of those detection methods, for which the authors did not communicate the performance in the way presented in this study, but for instance with a ROC curve or by the means of AUROC.

Limitations

At least three major restrictions can be found in this study. Firstly, only a handful of keywords were used when browsing the Web of Science database for conference proceedings and papers published between 1990 and 2020. Moreover, the publications we searched for were written in English only.

Secondly, the research was based on the assumption that it can always be clearly determined whether a claim is fraudulent or not. However, as we presented earlier certain researchers found that professionals handling different motor insurance claims had ambiguous perceptions of which claims constitute fraud.

Finally, the fact that the presented detection methods' detection performance (sensitivity and specificity) was tested on different datasets. It would be ideal to analyze the sensitivity and specificity of all the presented detection methods on the same real world, imbalanced dataset (possibly multiple datasets) and only afterwards determine the cost saving ability of different detection methods.

Bibliography

Abdallah, A., Maarof, M. A. & Zainal, A., 2016. Fraud detection system: A survey. *Journal of Network and Computer Applications,* Volume 68, pp. 90-113.

Achim, M. V. & Borlea, S. N., 2020. *Economic and Financial Crime: Corruption, Shadow Economy, and Money Laundering.* s.l.:Springer Nature.

Agresti, A., 1990. *Categorical Data Analysis, Wiley Series in Probability and Mathematical Statistics.* New York: Wiley.

Agyemang, M., Barker, K. & Alhajj, R., 2006. A comprehensive survey of numeric and symbolic outlier mining techniques. *Intelligent Data Analysis,* 10(6), pp. 521-538.

Ahmed, S. R., 2004. *Applications of data mining in retail business.* s.l., International Conference on Information Technology: Coding and Computing, pp. 455-459.

ALFA, 2020. *Insurance Fraud Agency*. [Online] Available at: https://www.alfa.asso.fr/ [Accessed 19 03 2020].

Artís, M., Ayuso, M. & Guillen, M., 1999. Modelling different types of automobile insurance fraud behaviour in the Spanish market. *Insurance: Mathematics and Economics*, 24(1-2), pp. 67-81.

Artís, M., Ayuso, M. & Guillén, M., 2002. Detection of automobile insurance fraud with discrete choice models and misclassified claims. *Journal of Risk and Insurance,* 69(3), pp. 325-340.

ASF norm no. 20, 2018. *Financial Supervisory Authority.* [Online] Available at: https://asfromania.ro/files/Asigurari/norme/2019/Norma%2020_2018_forma_aplicabila_dec2019.pdf [Accessed 10 10 2019].

Association of Certified Fraud Examiners, 2019. *Insurance Fraud Handbook*, Austin: ACFE.

Badriyah, T., Rahmaniah, L. & Syarif, I., 2018. *Nearest neighbour and statistics method based for detecting fraud in auto insurance..* s.l., IEEE, pp. 1-5.

Bahnsen, A. C., Aouada, D. & Ottersten, B., 2014. *Bahnsen, A. C., Aouada, D., & Ottersten, B. (2014, December). Example-dependent cost-sensitive logistic regression for credit scoring..* IEEE, In Proceedings of the thirteenth international conference on machine learning and applications (ICMLA).

Belhadji, E. B., Dionne, G. & Tarkhani, F., 2000. A model for the detection of insurance fraud. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 25(4), pp. 517-538.

Belhadji, E.-B. & Dionne, G., 1998. Development of an expert system for the automatic detection of automobile insurance fraud. *Available at SSRN 134768.*

Benetton, N., 1994. *A brief history of motor insurance.* [Online] Available at: https://insurance-times.net/article/brief-history-motor-insurance [Accessed 13 06 2020].

Bermúdez, L. et al., 2008. A Bayesian dichotomous model with asymmetric link for fraud in insurance. *Insurance: Mathematics and Economics,* 42(2), pp. 779-786.

Berry, M. J. & Linoff, G. S., 2004. *Data mining techniques: for marketing, sales, and customer relationship management.* Second ed. New York: Wiley.

Bhattacharyya, S., Jha, S., Tharakunnel, K. & Westland, J. C., 2011. Data mining for credit card fraud: A comparative study. *Decision Support Systems,* 50(3), pp. 602-613.

Bhowmik, R., 2011. Detecting auto insurance fraud by data mining techniques. *Journal of Emerging Trends in Computing and Information Sciences,* 2(4), pp. 156-162.

Bolton, R. J. & Hand, D. J., 2002. Statistical fraud detection: A review. *Statistical science,* 17(3), pp. 235-249.

Bornmann, L., Butz, A. & Wohlrabe, K., 2018. What are the top five journals in economics? A new meta-ranking.. *Applied Economics,* 50(6), pp. 659-675.

Brennan, P., 2012. *A comprehensive survey of methods for overcoming the class imbalance problem in fraud detection.* Dublin: Institute of technology Blanchardstown.

Brockett, P., Derrig, R. A., Levine, L. L. & Alpert, M., 2002. Fraud classification using principal component analysis of RIDITs. *Journal of Risk and Insurance,* 69(3), pp. 341-371.

Brockett, P. L., Xia, X. & Derrig, R. A., 1998. Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud. *Journal of Risk and Insurance,* 65(2), pp. 245-274.

Caudill, S. B., Ayuso, M. & Guillén, M., 2005. Fraud detection using a multinomial logit model with missing information. *Journal of Risk and Insurance,* 72(4), pp. 539-550.

Cecchini, M., Aytug, H., Koehler, G. J. & Pathak, P., 2010. Making words work: Using financial text as a predictor of financial events. *Decision Support Systems,* 50(1), pp. 164-175.

Coalition Against Insurance Fraud, 2019. *The State of Insurance Fraud Technology.* [Online] Available at: https://www.insurancefraud.org/downloads/Fraud_tech_study_2019.pdf [Accessed July 2020].

Decebal, B. M., 2005. *Controlul în asigurări.* Cluj-Napoca: Casa Cărţii de Ştiinţă.

Derrig, R. A., 2002. Insurance fraud. *Journal of Risk and Insurance,* 69(3), pp. 271-287.

Derrig, R. A. & Ostaszewski, K. M., 1995. Fuzzy techniques of pattern recognition in risk and claim classification. *Journal of Risk and Insurance,* 62(3), pp. 447-482.

Domingos, P., 1999. *A general method for making classifiers cost-sensitive.* s.l., Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining.

Dong, W. et al., 2014. *The Detection of Fraudulent Financial Statements: an Integrated Language Model.* s.l., Pacific Asia Conference on Information Systems (PACIS), pp. 383-400.

Duda, R. O., Hart, P. E. & Stock, E. G., 2001. *Pattern Classification.* New York: Wiley.

Duman, E. & Ozcelik, M. H., 2011. Detecting credit card fraud by genetic algorithm and scatter search. *Expert Systems with Applications,* 38(10), pp. 13057-13063.

Eick, S. G. & Fyock, D. E., 1996. Visualizing corporate data. *AT&T technical journal,* 75(1), pp. 74-86.

Ennas, G., Biggio, B. & Di Guardo, M. C., 2015. Data-driven journal meta-ranking in business and management. *Scientometrics,* 105(3), pp. 1911-1929.

Farquad, M. A. H., Ravi, V. & Raju, S. B., 2012. Analytical CRM in banking and finance using SVM: a modified active learning-based rule extraction approach. *International Journal of Electronic Customer Relationship Management,* 6(1), pp. 48-73.

FBI, 2011. *Financial Crimes Report 2010-2011.* [Online] Available at: https://www.fbi.gov/stats-services/publications/financial-crimes-report-2010-2011 [Accessed 27 10 2018].

FBI, 2012. *Insurance Fraud.* [Online] Available at: https://www.fbi.gov/stats-services/publications/insurance-fraud [Accessed 30 11 2019].

Frank, B., 2018. Element of motor insurance cover. In: J. U. Ntiedo, A. A. Ikechukwu & B. U. Uduak, eds. *Anatomy of Finance.* Uyo: Department of Banking and Finance University of Uyo, pp. 342-349.

Fraser, I. A., Hatherly, D. & Lin, K. Z., 1997. An empirical investigation of the use of analytical review by external auditors. *British Accounting Review,* Volume 29, pp. 35-48.

Gao, Z. & Mao, Y., 2007. A framework for data mining-based anti-money laundering research. *Journal of Money Laundering Control,* 10(2), p. 170–179.

Gepp, A., Wilson, H. J., Kumar, K. & Bhattacharya, S., 2012. A comparative analysis of decision trees vis-a-vis other computational data mining techniques in automotive insurance fraud detection. *Journal of Data Science,* 10(3), pp. 537-561.

Ghezzi, S. G., 1983. A private network of social control: Insurance investigation units. *Social Problems,* 30(5), pp. 521-530.

Glancy, F. & Yadav, S., 2011. A computational model for financial reporting fraud detection. *Decision Support Systems,* 50(3), pp. 595-601.

Halvaiee, N. S. & Akbari, M. K., 2014. A novel model for credit card fraud detection using Artificial Immune Systems. *Applied soft computing,* Volume 24, pp. 40-49.

Han, J., Pei, J. & Kamber, M., 2011. *Data mining: concepts and techniques.* Third ed. Waltham: Morgan Kaufmann Publishers.

Hassan, A. K. I. & Abraham, A., 2016. *Modeling insurance fraud detection using imbalanced data classification.* Cham, Springer, pp. 117-127.

He, H., Bai, Y., Garcia,, E. & Li, S., 2008. *ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning.* s.l., In 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), pp. 1322-1328.

He, H., Wang, J., Graco, W. & Hawkins, S., 1997. Application of neural networks to detection of medical fraud. *Expert Systems with Applications,* 13(4), pp. 329-336.

Holton, C., 2009. Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem.. *Decision Support Systems,* 46(4), pp. 853-864.

Hoogs, B., Kiehl, T., Lacomb, C. & Senturk, D., 2007. A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud. *Intelligent Systems in Accounting, Finance & Management: International Journal,* 15(1-2), pp. 41-56.

Huang, S. Y., 2013. Fraud detection model by using support vector machine techniques. *International Journal of Digital Content Technology and its Applications,* 7(2), pp. 32-42.

Humpherys, S. L. et al., 2011. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems,* 50(3), pp. 585-594.

IAATI, 2020. *International Association of Auto Theft Investigators.* [Online] Available at: https://www.iaati.org/ [Accessed 10 02 2020].

Insurance Europe, 2019. *Insurance fraud: not a victimless crime,* Brussels: Insurance Europe.

Insurance Fraud Bureau, 2015. *Cutting corners to get cheaper motor insurance backfiring on thousands of motorists warns the ABI.* [Online]
Available at: https://www.insurancefraudbureau.org/media-centre/news/2015/cutting-corners-to-get-cheaper-motor-insurance-backfiring-on-thousands-of-motorists-warns-the-abi/
[Accessed 01 09 2018].

IFB, 2020. *Insurance Fraud Bureau*. [Online] Available at: https://insurancefraudbureau.org/ [Accessed 17 01 2020].

IFED, 2020. *The Insurance Fraud Enforcement Department*. [Online] Available at: https://www.cityoflondon.police.uk/police-forces/city-of-london-police/areas/city-of-london/about-us/about-us/ifed/ [Accessed 23 01 2020].

IPD, 2020. *Insurance & Pension Denmark.* [Online] Available at: https://www.forsikringogpension.dk/en/ [Accessed 11 02 2020].

Insurance Sweden, 2020. *Insurance Sweden.* [Online] Available at: https://www.svenskforsakring.se/en/ [Accessed 27 02 2020].

Jans, M., Van Der Werf, J. M., Lybaert, N. & Vanhoof, K., 2011. A business process mining application for internal transaction fraud mitigation. *Expert Systems with Applications,* 38(10), pp. 13351-13359.

Japkowicz, N., 2000. *The class imbalance problem: Significance and strategies.* s.l., Proc. of the Int'l Conf. on Artificial Intelligence.

Japkowicz, N. & Stephen, S., 2002. The class imbalance problem: A systematic study. *Intelligent data analysis,* 6(5), pp. 429-449.

Kaminski, J. L., 2004. *Insurance Fraud, OLR Research Report.* [Online] Available at: http://www.cga.ct.gov/2005/ [Accessed 17 6 2019].

Karamizadeh, F. & Zolfagharifar, S. A., 2016. Using the clustering algorithms and rule-based of data mining to identify affecting factors in the profit and loss of third party insurance, insurance company auto. *Indian Journal of Science and Technology,* 9(7), pp. 1-9.

Kenyon, D. & Eloff, J. H., 2017. *Big data science for predicting insurance claims fraud.* Johannesburg, IEEE, pp. 40-47.

Kirkos, E., Spathis, C. & Manolopoulos, Y., 2007. Data mining techniques for the detection of fraudulent financial statements. *Expert systems with applications,* 32(4), pp. 995-1003.

Kou, Y., Lu, C.-T., Sirwongwattana, S. & Huang, Y.-P., 2004. *Survey of fraud detection techniques.* s.l., IEEE, pp. 749-754.

Kubat, M., Holte, R. C. & Matwin, S., 1998. Machine learning for the detection of oil spills in satellite radar images. *Machine learning,* 30(2-3), pp. 195-215.

Law no. 132, 2017. *Financial Supervisory Authority.* [Online] Available at: https://asfromania.ro/files/asigurari/legi/Lege%20132_2017_forma%20aplicabila_dec2019.pdf [Accessed 10 10 2018].

Law no. 136, 1995. *Portal Legislativ.* [Online] Available at: http://legislatie.just.ro/Public/DetaliiDocument/6381 [Accessed 19 01 2020].

Lichtenthaler, U., 2018. The world's most innovative companies: a meta-ranking.. *ournal of Strategy and Management,* 11(4), pp. 497-511.

Li, Y., Yan, C., Liu, W. & Li, M., 2018. A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification. *Applied Soft Computing,* Volume 70, pp. 1000-1009.

Long, B., 2005. *Automobile Insurance - A Brief History.* [Online] Available at: https://web.archive.org/web/20060618214632/http://www.drbilllong.com/Insurance/HistoryI.html [Accessed 07 01 2019].

Loughran, D. S., 2005. Deterring fraud: the role of general damage awards in automobile insurance settlements. *Journal of Risk and Insurance,* 72(4), pp. 551-575.

Lynn, B. D., 1974. *General insurance (Irwin series in insurance and economic security).* 9th edition ed. Homewood(Illinois): R. D. Irwin.

Maes, S., Tuyls, K., Vanschoenwinkel, B. & Manderick, B., 2002. *Credit card fraud detection using Bayesian and neural networks.* s.l., Proceedings of the 1st international naiso congress on neuro fuzzy technologies, pp. 261-270.

Majhi, S. K., Bhatachharya, S., Pradhan, R. & Biswal, S., 2019. Fuzzy clustering using salp swarm algorithm for automobile insurance fraud detection. *Journal of Intelligent & Fuzzy Systems,* 36(3), pp. 2333-2344.

Massachusetts Regulation, 1993. *Mass.gov - Division of Insurance Regulations.* [Online] Available at: https://www.mass.gov/service-details/division-of-insurance-regulations [Accessed 12 12 2019].

Moldoveanu, N., 2002. *Combaterea fraudei in asigurari.* Bucuresti: Bren.

Motor Insurance Directive, 2009. *The European Parliament and the Council of the European Union.* [Online] Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32009L0103&from=en [Accessed 15 10 2018].

Nagrecha, S., Johnson, R. A. & Chawla, V. N., 2018. FraudBuster: Reducing fraud in an auto insurance market. *Big Data,* 6(1), pp. 3-12.

National Health Care Anti-fraud Association, 2009. *Consumer Alert: The Impact of Health Care Fraud on You.* [Online] Available at: www.nhcaa.org [Accessed 17 2 2020].

Ngai, E. W. et al., 2011. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems,* 50(3), pp. 559-569.

Nian, K. et al., 2016. Auto insurance fraud detection using unsupervised spectral ranking for anomaly. *The Journal of Finance and Data Science,* 2(1), pp. 58-75.

OGU 54, 2016. *Financial Supervisory Authority.* [Online] Available at: https://asfromania.ro/files/asigurari/legi/Oug%2054_2016.pdf [Accessed 25 05 2019].

Olszewski, D., 2014. Olszewski, Dominik. "Fraud detection using self-organizing map visualizing the user profiles. *Knowledge-Based Systems* , Volume 70, pp. 324-334.

Oxford Languages, 1999. *Oxford Concise English Dictionary*. Tenth ed. Oxford: Oxford University Press.

Padmaja, T. M., Dhulipalla, N., Bapi, R. S. & Krishna, P. R., 2007. *Unbalanced data classification using extreme outlier elimination and sampling techniques for fraud detection.* s.l., IEEE, pp. 511-516.

Panigrahi, S., Kundu, A., Sural, S. & Majumdar, A. K., 2009. Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning. *Information Fusion,* 10(4), pp. 354-363.

Pathak, J., Vidyarthi, N. & Summers, S. L., 2005. A fuzzy-based algorithm for auditors to detect elements of fraud in settled insurance claims. *Managerial Auditing Journal,* 20(6), pp. 632-644.

Peng, Y. et al., 2006. *Application of clustering methods to health insurance fraud detection.* s.l., 2006 International Conference on Service Systems and Service Management, pp. 116-120.

Pérez, J. M. et al., 2005. *Consolidated tree classifier learning in a car insurance fraud detection domain with class imbalance.* Berlin, Springer, pp. 381-389.

Phua, C., Alahakoon, D. & Lee, V., 2004. Minority report in fraud detection: classification of skewed data. *Acm sigkdd explorations newsletter,* 6(1), pp. 50-59.

Pinquet, J., Ayuso, M. & Guillén, M., 2007. Selection bias and auditing policies for insurance claims. *Journal of Risk and Insurance,* 74(2), pp. 425-440.

Powers, D. M., 2010. *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.* s.l.:arXiv preprint arXiv:2010.16061.

Powers, D. M., 2012. *The problem with kappa.* s.l., Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics.

Priyanka, D. & Sri, H. B., 2019. *Analysis of Various Fraud Behaviors Using Soft Computing Techniques.* Singapore, Springer, pp. 601-614.

Quah, J. T. & Sriganesh, M., 2008. Quah, J. T., & Sriganesh, M. (2008). Real-time credit card fraud detection using computational intelligence. *Expert systems with applications,* 35(4), pp. 1721-1732.

Ravisankar, P., Ravi, V., Rao, R. G. & Bose, I., 2011. Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems,* 50(2), pp. 491-500.

Remi, O., 2006. *Elements of Insurance: Ofas Investment Limited.* Yaba: Lagos.

Road Traffic Act, 1930. *Legislation.gov.uk.* [Online] Available at: https://www.legislation.gov.uk/ukpga/1930/43/pdfs/ukpga_19300043_en.pdf [Accessed 05 07 2020].

Safta, I. L., Achim, M. V. & Borlea, S. N., 2020. Manipulation of Financial Statements Through the Use of Creative Accounting. Case of Romanian Companies. *Studia Universitatis „Vasile Goldis" Arad–Economics Series,* 30(3), pp. 90-107.

Sahin, Y., Bulkan, S. & Duman, E., 2013. A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications,* 40(15), pp. 5916-5923.

Sánchez, D., Vila, M. A., Cerda, L. & Serrano, J. M., 2009. Association rules applied to credit card fraud detection. *Expert systems with applications,* 36(2), pp. 3630-3640.

Sharma, S., 1996. *Applied Multivariate Techniques.* New York: Wiley.

Shaw, M. J., Subramaniam, C., Tan, G. W. & Welge, M. E., 2001. Knowledge management and data mining for marketing. *Decision support systems,* 31(1), pp. 127-137.

Sims, M. B., 2019. *The Different Types of Auto Insurance Coverage Explained.* [Online] Available at: https://www.autoinsurance.org/types-of-auto-insurance-coverage/ [Accessed: 27 10 2019].

Sohl, J. E. & Venkatachalam, A. R., 1995. A neural network approach to forecasting model selection. *Information & Management,* 29(6), pp. 297-303.

Sternberg, M. & Reynolds, R., 1997. Using cultural algorithms to support re-engineering of rule-based expert systems in dynamic performance environments: a case study in fraud detection.. *IEEE Transactions on Evolutionary Computation,* 1(4), pp. 225-243.

Šubelj, L., Furlan, Š. & Bajec, M., 2011. An expert system for detecting automobile insurance fraud using social network analysis. *Expert Systems with Applications,* 38(1), pp. 1039-1052.

Subudhi, S. & Panigrahi, S., 2017. Use of optimized Fuzzy C-Means clustering and supervised classifiers for automobile insurance fraud detection. *Journal of King Saud University-Computer and Information Sciences,* 32(5), pp. 568-575.

Subudhi, S. & Panigrahi, S., 2018. *Effect of Class Imbalanceness in Detecting Automobile Insurance Fraud.* s.l., IEEE, pp. 528-531.

Sundarkumar, G. G. & Ravi, V., 2015. A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. *Engineering Applications of Artificial Intelligence,* Volume 37, pp. 368-377.

Sundarkumar, G. G., Ravi, V. & Siddeshwar, V., 2015. *One-class support vector machine based undersampling: Application to churn prediction and insurance fraud detection.* s.l., IEEE, pp. 231-237.

Talanx Gruppe, 2014. *Germany's law on compulsory motor insurance marks its 75th anniversary.* [Online] Available at: https://www.talanx.com/newsroom/aktuelle-themen/2014/11_geb_kfz_pflichtversicherung.aspx?sc_lang=en [Accessed 12 03 2019].

Tan, P., Steinbach, M. & Kumar, V., 2005. *Introduction to data mining.* First ed. s.l.:Addison-Wesley Longman Publishing.

Tao, H., Zhixin, L. & Xiaodong, S., 2012. *Insurance fraud identification research based on fuzzy support vector machine with dual membership.* s.l., IEEE, pp. 457-460.

Tennyson, S. & Salsas-Forn, P., 2002. Claims auditing in automobile insurance: fraud detection and deterrence objectives. *Journal of Risk and Insurance,* 69(3), pp. 289-308.

Thomson Reuters, 2009. *Usa healthcare fraud report,* New York: Thomson Reuters.

Turban, E., Aronson, J. E., Liang, T. P. & Sharda, R., 2007. *Decision Support and Business Intelligence Systems.* Eighth ed. s.l.:Pearson Education.

UNESPA, 2019. *Spanish Union of Insurance and Reinsurance Companies.* [Online]
Available at: https://www.unespa.es/en/ [Accessed 15 11 2019].

Vasu, M. & Ravi, V., 2011. A hybrid under-sampling approach for mining unbalanced datasets: applications to banking and insurance. *International Journal of Data Mining, Modelling and Management,* 3(1), pp. 75-105.

Vatsa, V., Sural, S. & Majumdar, A. K., 2005. *Vatsa, Vishal, Shamik Sural, and Arun K. Majumdar. "A game-theoretic approach to credit card fraud detection.* Berlin, Springer, pp. 263-276.

Vaughan, E. J. & Vaughan, T., 2007. *Fundamentals of risk and insurance.* New York: John Wiley & Sons.

Viaene, S. et al., 2007. Strategies for detecting fraudulent claims in the automobile insurance industry. *European Journal of Operational Research,* 176(1), pp. 565-583.

Viaene, S., Dedene, G. & Derrig, R. A., 2005. Auto claim fraud detection using Bayesian learning neural networks. *Expert Systems with Applications,* 29(3), pp. 653-666.

Viaene, S., Derrig, D. A. & Dedene, G., 2004. A case study of applying boosting Naive Bayes to claim fraud diagnosis. *EEE Transactions on Knowledge and Data Engineering,* 16(5), pp. 612-620.

Viaene, S., Derrig, R. A., Baesens, B. & Dedene, G., 2002. A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection. *Journal of Risk and Insurance,* 69(3), pp. 373-421.

Wang, J.-H., Liao, Y. L., Tsai, T. M. & Hung, G., 2006. *Technology-based financial frauds in Taiwan: issues and approaches.* s.l., IEEE International Conference on Systems, Man and Cybernetics, pp. 1120-112.

Wang, Y. & Xu, W., 2018. Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems,* Volume 105, pp. 87-95.

Warren, D. E. & Schweitzer, M. E., 2018. When lying does not pay: How experts detect insurance fraud. *Journal of Business Ethics,* 150(3), pp. 711-726.

Webb, A., 2003. *Statistical Pattern Recognition.* Second ed. Chichester: John Wiley & Sons.

Weisberg, H. I. & Derrig, R. A., 1991. Fraud and Automobile Insurance: A Report on Bodily Injury Liability Claims in Massachusetts. *Journal of Insurance Regulation,* 9(4), pp. 497-541.

Weisberg, H. I. & Derrig, R. A., 1998. Quantitative methods for detecting fraudulent automobile bodily injury claims. *Risques,* 35(July–September), pp. 75-99.

Wenzel, T., 1995. *Analysis of national pay-as-you-drive insurance systems and other variable driving charge.* [Online] Available at: https://www.osti.gov/biblio/125357 [Accessed 10 12 2018].

West, J. & Bhattacharya, M., 2016. Intelligent financial fraud detection: a comprehensive review. *Computers & Security,* Volume 57, pp. 47-66.

Whitrow, C. et al., 2009. Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery,* 18(1), pp. 30-55.

Wilson, J. H., 2009. An analytical approach to detecting insurance fraud using logistic regression. *Journal of Finance and Accountancy,* 85(150), pp. 1-15.

Wong, N., Ray, P., Stephens, G. & Lewis, L., 2012. Artificial immune systems for the detection of credit card fraud: an architecture, prototype and preliminary results. *Information Systems Journal,* 22(1), pp. 53-76.

Wu, S. X. & Banzhaf, W., 2008. *Combatting financial fraud: a coevolutionary anomaly detection approach.* s.l., ACM, pp. 1673-1680.

Xu, L. & Chow, M. Y., 2006. A classification approach for power distribution systems fault cause identification.. *IEEE Transactions on Power Systems,* 21(1), pp. 53-60.

Xu, W., Wang, S., Zhang, D. & Yang, B., 2011. *Random rough subspace based neural network ensemble for insurance fraud detection.* s.l., IEEE, pp. 1276-1280.

Yamanishi, K., Takeuchi, J. I., Williams, G. & Milne, P., 2004. On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. *Data Mining and Knowledge Discovery,* 8(3), pp. 275-300.

Yan, C. & Li, Y., 2015. *The Identification Algorithm and Model Construction of Automobile Insurance Fraud Based on Data Mining.* s.l., IEEE, pp. 1922-1928.

Yue, D. et al., 2007. *A review of data mining-based financial fraud detection research.* s.l., IEEE, pp. 5519-5522.

Zaki, M. & Theodoulidis, B., 2013. Analyzing financial fraud cases using a linguistics-based text mining approach.

Zelenkov, Y., 2019. Example-dependent cost-sensitive adaptive boosting. *Expert Systems with Applications*, Volume 135, pp. 71-82.

Zhang, D. & Zhou, L., 2004. Discovering golden nuggets: data mining in financial application. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* 34(4), pp. 513-522.

Zhang, G., Patuwo, B. E. & Hu, M. J., 1998. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1), pp. 35-62.

Zhou, W. & Kapoor, G., 2011. Detecting evolutionary financial statement fraud. *Decision Support Systems*, 50(3), pp. 570-575.

Zhou, Z. H. & Liu, X. Y., 2005. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *EEE Transactions on knowledge and data engineering,* 18(1), pp. 63-77.

Zimmermann, C., 2013. Academic rankings with RePEc. *Econometrics*, 1(3), pp. 249-280.

Electronic references

www.abcasigurari.ro
www.alfa.asso.fr
www.allianztiriac.ro
www.asfromania.ro/
www.asirom.ro
www.baar.ro
www.cga.ct.gov
www.cityinsurance.ro
www.cityoflondon.police.uk
www.data.gov.ro
www.ergo.ro
www.euroins.ro
www.fbi.gov
www.forsikringogpension.dk/en/
www.generali.ro
www.gothaer.ro
www.groupama.ro
www.iaati.org/
www.iii.org
www.insurance-times.net
www.insuranceeurope.eu
www.insurancefraud.org
www.insurancefraudbureau.org
www.legislatie.just.ro
www.legislation.gov.uk
www.mass.gov
www.nhcaa.org
www.omniasig.ro
www.osti.gov
www.svenskforsakring.se/en/
www.talanx.com
www.unespa.es/en/
www.uniqa.ro
www.worldbank.org