

Fraud Prevention in the Leasing Industry Using the Kohonen Self-Organising Maps¹

Mirjana PEJIĆ BACH, Nikola VLAHOVIĆ, Jasmina PIVAR

University of Zagreb, Faculty of Economics & Business, Trg J. F. Kennedy 6, Zagreb, Croatia, mpejic@efzg.hr, nvlahovic@efzg.hr, jpivar@efzg.hr

Background and Purpose: Data mining techniques are intensely used in various industries for the purpose of fraud prevention and detection. Research that focuses on the leasing industry is scarce, although frauds in the field of leasing occur rather often. First, we identify clusters of business clients in one leasing company by using the method of self-organising maps based on leasing contract attributes. Second, we compare clusters based on the presence of fraudulent clients, in order to develop fraudsters' profiles.

Methodology: For detecting characteristics of fraudulent clients, we use a client database containing leasing contract attributes of one Croatian leasing company. In order to develop profiles of fraudulent clients, we utilise a clustering procedure with the Kohonen Self-Organizing Maps supported by Viscovery SOMine software.

Results: Five clusters were identified and labelled according to the modal values of attributes describing the leasing object and the industry in which the client operates: (i) New cars / Trade; (ii) Used trucks or tugboats / Other services; (iii) New machinery / Construction; (iv) New motors / Trade; and (v) New machinery and tractors / Agriculture.

Conclusion: Self-organising maps have proved to be a useful methodology for developing profiles of fraudulent clients in leasing companies. Companies can use our results and make additional efforts in monitoring clients from the identified industries, buying specific leasing objects. In addition, companies can apply our methodology to their own databases, in order to develop fraudster profiles for their specific purposes, and implement fraud alert mechanisms in their client database.

Keywords: *fraud, leasing, self-organising maps, Viscovery SOMine, Ward algorithm, Croatia, data mining*

1 Introduction

Knowledge management consists of the processes of creating, storing/retrieving, transferring and applying knowledge (Alavi & Leidner, 2001). The process of knowledge discovery is an important subprocess in knowledge management (Wang & Wang, 2008). Some of the tasks solved by data mining are clustering and deviation detection (Folorunso & Ogunde, 2005), which also includes fraud detection. Numerous other applications are also focused to

rare events, such as bankruptcy (e.g. Moradi, Salehi, Ghorani & Yazdi, 2013). In this paper, the focus is on fraud in the leasing industry.

Frauds represent an issue for leasing companies and regulators, which should be able to predict fraudulent behaviour and take different actions to prevent losses caused by fraud. Defence against frauds includes the implementation of operational and technical solutions for fraud prevention and detection. Fraud detection systems are based on data mining techniques and methods that can discover and visualise patterns related to fraudulent behaviour, such

¹A preliminary version of this research (<http://doi.org/10.23919/MIPRO.2018.8400218>) was presented at 41st International Convention on Information and Communication Technology, Electronics and Microelectronics MIPRO 2018, Opatija, May 21-25, 2018.

as financial frauds (Sadgali, Sael, & Benabbou, 2019), credit card frauds (Carcillo et al., 2019), and frauds in the insurance sector (Leite, Gschwandtner, Miksch, Gstrein, & Kuntner, 2018). Cluster analyses and profiling of clients based on various behavioural, demographic and operational attributes contained in clients databases are essential tools in analysing transactions, and recognising client profiles, which have been used in various industries, such as banking (e.g. Pejić Bach, Juković, Dumičić, & Šarlija, 2014). Clients profiling based on the cluster analysis has also been used in various researches and has been proved as a useful tool in predicting fraudulent behaviour, which can help companies to develop appropriate fraud detection and response systems, e.g. financial statement fraud detection system (Chen, Liou, Chen & Wu, 2019). Current research on fraud detection and prevention in the leasing industry is scarce (Singleton & Singleton, 2007), with only a few examples that present the utilization of data mining techniques for that purpose. For example, Horvat, Pejić Bach and Merkač Skok (2014) used a decision tree modelling in order to discover fraud in leasing agreements.

Self organizing maps have been efficiently used to explain fraudulent behaviour in different contexts of the financial industry, including banking (e.g. Merkevicius, Garšva, & Simutis, 2004; Balasupramanian, Ephrem, & Al-Barwani, 2017) and insurance (e.g. Hainaut, 2019). However, to our best knowledge, previous works did not utilise self-organising maps for fraud profiling in leasing, although self-organising maps have been previously effectively deployed for fraud prevention and detection (Jian, Ruicheng, & Rongrong, 2016). The research question that emerges is whether self-organising maps are an appropriate method for identifying and describing clusters of clients in the context of the leasing industry, with the specific goal of detecting specific attributes that could explain the fraud in the leasing industry. In order to shed some light on this issue, we develop the methodology for developing fraudsters profiles using self-organising maps, based on the leasing contract attributes. We use the database of one leasing company with the rich data on client characteristics and behaviour, for the identification of fraudulent behaviour. First, we use self-organising maps in order to develop clusters of business clients in a leasing company based on leasing contract attributes. Second, we identify the characteristics of fraudulent clients among cluster members.

The paper is structured as follows. After the introduction, the literature review section describes frauds in the leasing industry and gives an overview of previous research related to fraud modelling. The second section explains the methodology of the research, including the self-organising maps, the sample description, and the statistical analysis. The fourth section provides results of the clustering procedure and the fraud analysis according to client and leasing characteristics. It also contains the interpretation of the clusters and profiles of fraudsters for each of the clusters based on all the attributes used for the

analysis. The last section is the discussion and conclusion section, which provides a response to the research question and describes the contributions of this research.

2 Literature review

2.1 Fraud in the leasing industry

Fraud causes material and immaterial losses to an organisation or a person. According to the Basel Committee (Basel Committee on Banking Supervision, 2002), frauds are loss events that are classified into internal and external frauds. Internal frauds are “losses due to acts of a type intended to defraud, misappropriate property or circumvent regulations, the law or company policy, excluding diversity and discrimination events, which involves at least one internal party” (Basel Committee on Banking Supervision, 2002, p.3), such as accounting administrators. External frauds are “losses due to acts of a type intended to defraud, misappropriate property or circumvent the law, by a third party” (Basel Committee on Banking Supervision, 2002, p.3), such as clients or partners. Fraud is often both internal and external.

European Commission (2011, p.3) defines a lease as “an agreement whereby the lessor conveys to the lessee in return for a payment or series of payments the right to use an asset for an agreed period”. In order to understand the concept of fraud in leasing, it is necessary to understand ownership rights in the context of the leasing contracts. During different stages of the leasing contract, difficulties in executing ownership rights can occur. Such differences can be the result of the complex leasing law framework (Flath, 1980). However, fraud in leasing, as in other financial industries, is often intentionally conducted by the client. In that case, leasing companies are usually not able to reach a client or locate a leasing object. For example, fraud happens when a client refuses to return a leasing object after a lease expires. In such a scenario, a leasing company can contact a client and it knows the location of a leasing object but regaining or repurchasing a leasing object is not possible without a complex law procedure.

This research focuses on frauds and defaults committed by clients (small and medium companies, and sole proprietorships) in the leasing industry. Defending leasing companies against leasing fraud brings challenging issues both operationally and technically. An efficient fraud defence system in the field of leasing has several prerequisites. A leasing organisation needs to create anti-fraud measures and introduce them to its employees, as well as to keep employees aware of the fact that frauds are a part of the leasing industry (Boobyer, 2003). Cross-departmental cooperation and communication, especially of sales, human resources, and accounting department, as well as cooperation with external experts are also needed. Addi-

tionally, an organisation should establish client verification procedures (Wang, Cheng, & Chen, 2019). In leasing, such procedures are used to verify leasing objects such as verification of client economic activity, verification of payments and so on. Upgrading information systems with data analytics and warning systems that would support decisions in relation to potentially fraudulent clients are crucial as well (Bănărescu, 2015).

2.2 Fraud modelling

Fraud and default modelling are based on various data mining methods. Ngai, Hu, Wong, Chen and Sun (2011) reviewed data mining techniques for the detection of financial fraud. They concluded that logistics models, neural networks, decision trees, and the Bayesian belief network are the primary data mining techniques for financial fraud detection. Sadgali, Sael and Benabbou (2019) reviewed the performance of various machine-learning techniques such as classification, clustering, and regression for fraud and prevention detection. In addition, visual analysis techniques are used for the identification of fraud detection. In identifying and preventing attempts of fraud, detection of suspicious events can be made by using visual analytics techniques (Leite, Gschwandtner, Miksch, Gstrein, & Kunter, 2018), who categorised, described and discussed current visualisation, interaction and analytical methods that can be used in fraud detection systems. Chen, Liou, Chen and Wu (2019) proposed the approach for detecting fraud in the financial statements in business groups by using data mining techniques.

However, current research does not conclude which method performs the best in fraud prevention and detection, although several authors identified that neural networks and clustering were the most efficient. Deep convolution neural networks (DCNN) were used to detect fraudsters in customer records of a mobile communication company (Chouiekh & Haj, 2018). The authors stated that DCNN outperforms support vector machines, random forest and a gradient boosting classifier in terms of accuracy and training duration.

Data mining methods have been implemented in various application areas related to fraud. Rousseeuw, Perrotta, Riani and Hubert (2019) combined the idea of the Fast LTS algorithm (least trimmed squares) for robust regression for the detection of unexpected events in time series. These unexpected events are often outliers and shifts that can represent suspicious transactions. An intuitionistic fuzzy set, one of the classification methods, and evidential reasoning were proposed for fraud detection in banking transactions by Eshghi and Kargari (2019), who modelled transactional behaviour by considering the trends of different variables. The method determines the originality of a newly arrived transaction.

Credit card fraud has been researched by several au-

thors. Lucas et al. (2020) used a hidden Markov model and a random forest classifier for credit card fraud detection. The hidden Markov model was used to associate a likelihood to a transaction given its sequence of previous transactions. Likelihoods are then used by a random forest classifier for fraud detection. Ryman-Tubb, Krause and Garn (2018) presented a survey of methods that use AI and machine learning for credit card fraud detection, with the conclusion that in terms of accuracy neural networks were on average better than other techniques. West and Bhattacharya (2016) analysed issues of credit card fraud mining related to the choice of detection techniques, problem representation, feature and performance analysis. Nami and Shajari (2018) proposed a two-stage method of detecting fraudulent payment card transactions. The method is based on k-nearest neighbours, the dynamic random forest algorithm and the minimum risk model. Patil, Nemade and Soni (2018) used the big data analysis framework and machine learning algorithms for real credit card fraud detection. Deployment of a fraud detection system based on machine learning methods in a large e-tail merchant was explored and described by Carneiro, Figueira and Costa (2017). Ensemble learning is a common method used in various practical problems. Zareapoor and Shamsolmoali (2015) evaluated and compared various data mining techniques for credit card fraud detection. They presented the decision tree based bagging classifier as the best classifier to construct the fraud detection model. Deep learning neural networks, Generative Adversarial Networks, were used to improve the effectiveness of classifiers for credit card fraud detection by Fiore et al. (2019). Tu, He, Shang, Zgou and Li (2019) proposed convolutional neural networks for the enhancement of anti-fraud systems in the area of e-commerce payments.

Several pieces of research have been conducted in the area of insurance. Yan, Li, Liu, and Qi (2020) used an adaptive genetic algorithm with a backpropagation neural network for simulation and prediction of frauds in the automobile insurance claim data. An Artificial Bee Colony algorithm-based Kernel Ridge Regression was proposed for automobile insurance fraud detection by Yan et al. (2019). An Artificial Bee Colony was used for global optimization and to optimize the parameter combination of the Kernel Ridge Regression. Wang and Xu (2018) proposed a deep learning model for automobile insurance fraud detection based on text mining. They used the Latent Dirichlet Allocation-based text analytics to extract text features of the descriptions of the accidents in the claims. Deep neural networks are used for detecting fraudulent claims. Neural networks were used to detect fraud in the automobile insurance industry, with the aim of fraud detection when it comes to personal injury claims (Viaene, Dedene, & Derig, 2005). Machado and Santos (2015) used five strategies for auditing vehicle claims and concluded that neural networks perform the best. Šubelj, Furlan, and Baje (2011) proposed an expert system for the detection of groups of

automobile insurance fraudsters by using an Iterative Assessment Algorithm (IAA). Patel and Singh (2013) used genetic algorithms to detect fraudulent activities in credit card transactions. Fuzzy C-Means clustering and supervised classifiers comprise the novel hybrid approach that was proposed for detecting fraud in an automobile insurance dataset (Subudhi & Panigrahi, 2017). Nian, Zhang, Tayal, Coleman and Li (2016) proposed a spectral ranking method for automobile insurance fraud detection, while Caldeira, Gassenferth, Machado and Santos (2015) used neural networks for the same purpose.

Additionally, neural networks were used to detect fraud in the context of bank direct marketing (Zakaryazad & Duman, 2016) and card payments and operations (Dorronsoro, Ginel, Sánchez, & Cruz, 1997). Recurrent neural networks were used for the detection of stock price manipulation activities by Wang, Xu, Huang, and Yang (2019). The authors concluded that the method could be used to identify unusual trading activities among huge amounts of data.

2.3 Kohonen self-organising maps in fraud research

Self-organising maps (SOMs), Kohonen Map or Kohonen Neural Networks are feed-forward neural networks based on unsupervised learning and a clustering algorithm that produces two dimensional and nonlinear mappings of multidimensional data (Urueña López et al., 2019).

SOMs are widely used for research in different contexts of the financial industry, including banking, insurance and so on (Van Hulle, 2012).

Pejić Bach, Juković, Dumičić and Šarlija (2014) identified three clusters by using self-organising maps for business clients' segmentation in the context of the Croatian banking industry, and authors suggested marketing activities for the identified clusters. Holmbom, Eklund and Back (2011) described how self-organising maps could be used for customer portfolio analysis. Merkevicius, Garšva, and Simutis (2004) explored the usage of self-organising maps for forecasting of credit classes.

Only several researchers investigated the usage of SOMs in fraud. Urueña López et al. (2019) used self-organising maps for finding hidden relationships in data about fraud on the Internet, computer users' behaviour, as well as security incidents. Balasupramanian, Ephrem and Al-Barwani (2017) proposed an architectural framework that uses big data analytics and the self-organising maps to handle card fraud effectively. Olszewski (2014) presented how self-organising maps can be used for visualisation of user profiles and comparison of frauds in credit card transactions, telecommunications, and networks. Almendra and Enachescu (2013) present an algorithm that combines the self-organising map with the supervised learning paradigm with labelled data in the context of online auction

sites. Quah and Sriganesh (2008) described a real-time fraud detection approach aimed at a better understanding of fraudulent spending patterns based on self-organising maps. Zaslavsky and Strizhak (2006) derived the model of a typical cardholder's behaviour and analysed suspicious transactions by using self-organising maps. Brockett, Xia, and Derrig (1998) classified suspicious automobile bodily injury claims by using self-organising maps.

Data mining has been extensively used in fraud detection and prevention, with various areas of applications, such as credit card fraud and insurance fraud. Several researchers indicated that neural networks outperform other methods for fraud prevention and detection. To our best knowledge, no research presents the application of data mining in fraud prevention and detection in the leasing industry.

3 Methods

3.1 Self-organising maps (SOMs)

The goal of using the SOMs is to discover similarities among elements in the set of instances and to organise the neurons in the computational layer into clusters associated with patterns in the set of instances. Therefore, SOMs are visual representations of learned structures that appear as clusters of similar objects.

The basic SOMs algorithm can be described as follows (Bação, Lobo, & Painho, 2005). The neighbourhood function is a function that decreases with the distance to the winning node and is responsible for the interactions among nodes. During training, the radius of this function decreases, so each node becomes more isolated from the effects of its neighbours. The winning node changes its weight vector to become more similar to the input vector. All neighbours of the winning node also change their weights to the direction of the input vector. Thus, the weight vectors of neighbouring nodes become similar because of their convergence with the winning node towards the input data vector.

The corresponding error function $E(w)$ with an expectation value converging to a minimum during the training process (distortion measure) is:

$$E = \int \sum_i h_{ci} |w - x| g(x) d_n x, \quad (1)$$

where h_{ci} is the neighbouring function of node i to the corresponding winner $c(x)$, and $g(x)$ the density function of the vectors x in the n -dimensional data space. The Kohonen net is obtained in a discrete data space by computing the optimal weight vectors for minimizing $E(w)$ using a gradient descent (Viscovery, 2019).

In addition, SOMs can be seen as a form of k -means clustering in which every unit corresponds to a "cluster", and the number of clusters is defined by the size of the

grid (Wehrens & Buydens, 2007). In comparison to the k-means clustering, Kohonens's self-organizing maps showed more accuracy in classifying most of the objects when the number of clusters is lower than eight (Abbas, 2008). Bação, Lobo, and Painho (2005) also proposed the use of SOMs as a possible substitution for the k-means clustering. They concluded that during the search, space is better explored by SOM, and by the end of the search process, the SOM is the same as k-means, which allows for a minimization of the distances between the nodes and the winning node. The main reason for the usage of SOMs in this research is that the k-means clustering algorithm is mainly used for minimizing the sum of squared distances between the input and the prototype vectors, but it does not perform topological mapping like Kohonen self-organizing maps do (Van Laerhoven, 2001).

SOMs are used in state-of-the-art software. Viscovery SOMine software is specialised software, which enables clustering by using two algorithms that are based on the classical hierarchical agglomerative cluster method of Ward (Viscovery, 2019). The first algorithm is based on the Ward method, which uses the variance criterion as a distance measure. The second algorithm is the SOM Ward algorithm based on the modified Ward method. It is developed on the ground of the soft computing paradigm. In this method, the topological neighbourhood influences the cluster merge steps (Viscovery, 2019). The nodes with many corresponding data records have a higher impact in comparison with the nodes with fewer matching records (Viscovery, 2019).

As a distance measure, a modified Ward distance is used. This distance observes the topological locations of the clusters. It means that two clusters that are not neighbouring in the SOM are never considered to be merged (Viscovery, 2019):

$$d_{rs} = \begin{cases} 0 & \text{if } n_r = n_s = 0, \\ \frac{n_r n_s}{n_r + n_s} \|\bar{x}_r + \bar{x}_s\|^2 & \text{otherwise.} \end{cases}$$

$$\rightarrow d'_{rs} = \begin{cases} d_{rs} & \text{if clusters } r \text{ and } s \text{ are adjacent in the SOM.} \\ \infty & \text{otherwise.} \end{cases}$$
(2)

Then, the SOM – Ward distance is normalized with an exponential function (Viscovery, 2009):

$$\mu(c) = d(c) * e^{\beta}, \quad (3)$$

where $d(c)$ indicates the SOM-Ward distance used to merge c clusters into $c-1$ clusters and β is a linear regression coefficient ($3 \leq c < C$).

For this research, the SOM-Ward algorithm supported by Viscovery SOMine software was used.

3.2 Sample description and statistical analysis

In this research, the analysis was performed using the client base of one Croatian leasing company containing data on 13,057 small and medium enterprises (SMEs) and sole proprietorships as clients with expired leasing contracts. The dataset contains numerous attributes. The following attributes were used.

- Client sector - a nominal attribute related to demographic characteristics of clients, and it has eight modalities: *agriculture, chemical, construction, financial, trade, other services, public, and tourism*
- Client New/Old - a nominal attribute related to behavioural characteristics of the client, and it is represented by two modalities: *new* and *old*
- Leasing object - a nominal attribute related to operational characteristics of the lease agreements. 12 modalities describe it: *car, light commercial vehicle, truck_tugboat, machine, equipment, trailer_semitrailer, motor, agri_forest, forklift, vessel, public* and *other*
- Leasing object New/Used - a nominal attribute describing operational characteristics of the lease agreements. It is represented by two modalities: *new* and *used*
- Leasing type – an operational attribute with two modalities: *financial* and *operative*
- Client type – a demographic attribute with two modalities: *company* and *sole proprietorship*
- Client County – a demographic attribute with 20 modalities: all the Croatian counties
- Client rating – a behavioural attribute with four modalities: R2, R3, R4, and R5 (*R2-the lowest risk, R5-the highest risk*)
- Risk mark – a behavioural attribute with two modalities: *No_estimation, No_risk, and Risk*

For the purpose of this research, fraud is defined as every act of a client that decreases the possibility to regain a leasing object or payment during collection (Pejić Bach, Vlahović, & Pivar, 2018). The goal of the research is to develop an algorithm that could be used for the purpose of fraud prevention and detection. The goal attribute in our research is:

- Fraud/Default attribute is represented by two modalities: (1) - *fraud_default*, which describes the situation when a lease is terminated because fraud or default occurred; and (0) - *non-fraud* or *non-default* cases which refer to the contracts that were terminated in cases of pre-term repurchase, normal, pre-term termination and harms.

A cluster analysis was performed by using the SOM-Ward algorithm implemented in Viscovery SOMine software. The first step in performing a cluster analysis is to define the map size, training parameters, and a clustering method. The map size is the granularity of the map that is determined by a number of nodes. More nodes require more time for training. For the cluster analysis in this research, a map with 14000 nodes was trained with a normal training schedule. In Viscovery SOMine, the number of clusters should be set before running the SOM-Ward algorithm. Therefore, the algorithm was run with varying numbers of clusters and the most appropriate clustering result was selected using the domain knowledge, by consulting the expert in the field (Uribe & Isaza, 2012). The number of five clusters was determined, and the SOM-Ward clustering method was chosen. In the following steps, the map was explored in order to identify clusters, which are presented in the results section.

Chi-square tests results were used to describe (i) characteristics of clusters according to the characteristics of the leasing contracts, and (ii) characteristics of clusters according to the occurrence of frauds/defaults within clusters.

4 Results

4.1 Cluster identification

The SOM algorithm revealed five clusters in the leasing company dataset. Clusters can be described and clearly distinguished based on all the attributes used for the analysis. Figure 1 shows the self-organising map in which clusters are labelled according to the modal values of the attribute leasing object and the attribute client sector. Since the difference between the clusters is the largest in relation to the leasing objects and industry, clusters were named after them, like the following: Cluster 1 – new cars / trade; Cluster 2 – used trucks and truckboat / other services; Cluster 3 – new machine / construction; Cluster 4 – new motors / trade; and Cluster 5 – new machines and tractors / agriculture.

Table 1 presents the structure of the total number of leasing contracts per cluster. Cluster 1 contains the majority of the leasing contracts (72.18%), and Cluster 5 contains only 1.86% of the total number of leasing contracts.

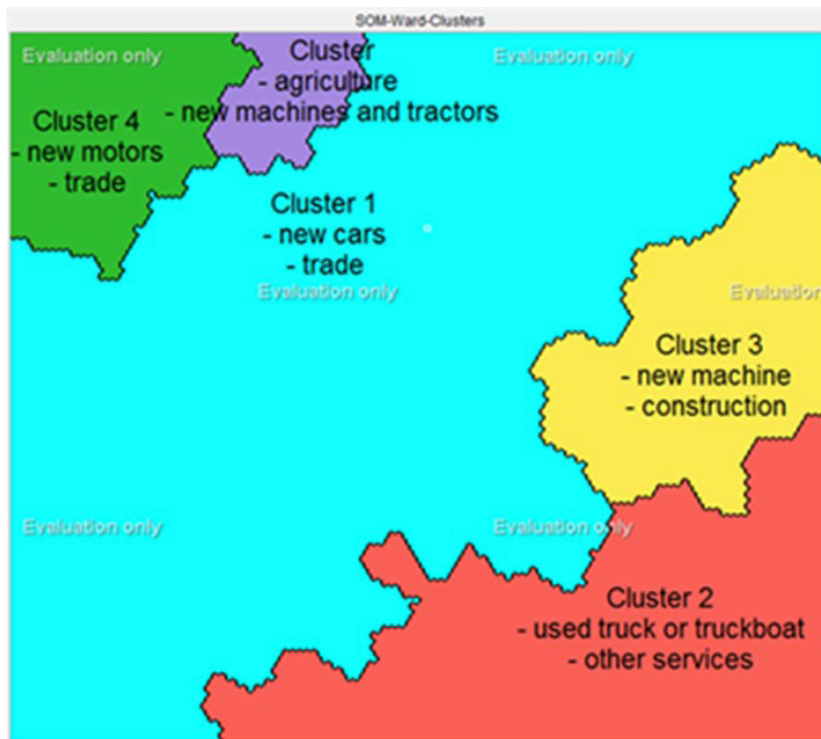


Figure 1: SOM-Ward clusters of clients in leasing. Source: Authors' work based on Viscovery SOMine output

Table 1: Clusters according to the number of leasing contracts

Cluster	Number of leasing contracts in the cluster	% of the total number of leasing contracts
C 1	9425	72.18%
C 2	1828	14.00%
C 3	951	7.28%
C 4	611	4.68%
C 5	243	1.86%
Total	13057	100%

Source: Authors' work based on Viscovery SOMine output

The quantization error is a measure of how well the data vectors from the source data set are matched by a specific node. It is calculated by the average of the squared distance of all data records associated with a node. Averaging over the quantization errors of all nodes yields the quantization error of the map (Viscovery, 2019). The value of the quantization errors (Table 2) suggests that the map is well trained. The errors are distributed evenly over the map.

Table 3 presents the clusters according to the demographic attributes of clients. Clusters differ significantly according to the client sector. For example, in Cluster 1 majority of clients perform trade activities (41.4%). Furthermore, in Cluster 1 companies perform other services (27.9%) and construction activities (17.3%). In Cluster 2 56.2% of clients and Cluster 3 31.4% of clients perform other services. In Cluster 4 46.8% of clients perform trade activities, followed by other services (29%). The agriculture sector is dominant in Cluster 5 (75.35%).

Furthermore, Chi-squares show a significant association between the clusters and the attribute client type as well as associations with the client county. It can be noticed that in all the five clusters there is a high percentage of SMEs, although in Cluster 4 this percentage is the highest. Cluster 5 is distinguished from the others by the highest percentage of sole proprietorships (73.7%).

All the clusters have a high percentage of clients doing business in Zagreb County. However, Cluster 1 clients from Zagreb County are followed by those from Primorje-Gorski Kotar County, and in Cluster 2, 3 and 4 by clients from Split-Dalmatia County. Compared with the other clusters, Cluster 4 contains the highest percentage of clients from Zagreb County. Similarly, Cluster 3 contains the highest percentage of clients from Split-Dalmatia County. It can be noticed that Cluster 5 contains a high percentage of clients from counties that are traditionally related to agricultural activities. The clusters differ according to the demographic characteristics of clients.

Table 2: Training report

Data records: 13057	Attributes: 26	Principal plane: 100:90
Nodes: 14096	Rows: 121	Columns: 117
Schedule: Nrmal	Training cycles: 115	Tension: 0,5
Final errors were:	Normalized distortion: 0,00045	Quantization error: 0

Source: Authors' work based on Viscovery SOMine output

Table 3: Clusters according to client sector, type and county

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Chi-square (p-value)
Client sector						
Agriculture	5.7%	0.5%		3.1%	75.3%	3411.889 (0.000***)
Chemical	0.3%					
Construction	17.3%	16.1%	43.3%	15.5%	2.1%	
Financial	0.5%					
Other_services	27.9%	56.2%	31.4%	29.0%	5.3%	
Public	2.0%			0.2%		
Tourism	4.9%	0.7%		5.4%	0.8%	
Trade	41.4%	26.5%	25.4%	46.8%	16.5%	
Client type						
Company	75.1%	55.5%	64.7%	80.0%	26.3%	5653.550 (0.000***)
Sole proprietorship	24.9%	44.5%	35.3%	20.0%	73.7%	
Client County						
Bjelovar-Bilogora	1.0%	2.1%	1.1%	0.5%	10.3%	1482.537 (0.000***)
Brod-Posavina	5.1%	3.3%	2.3%	1.0%	2.5%	
Dubrovnik-Neretva	1.3%	1.5%	2.1%	1.8%		
Istria	6.4%	2.0%	4.9%	8.7%	2.5%	
Karlovac	2.2%	3.4%	2.0%	4.7%	1.6%	
Koprivnica-Križevci	1.4%	1.4%	0.5%	1.0%	8.6%	
Krapina-Zagorje	2.0%	3.6%	2.3%	1.0%	1.6%	
Lika-Senj	0.9%	1.0%	2.3%	0.8%	0.4%	
Međimurje	0.9%	1.3%	0.4%	0.7%		
Osijek-Baranja	3.9%	2.7%	3.9%	1.5%	21.8%	
Požega-Slavonia	0.4%	0.6%	0.3%	0.2%	0.8%	
Primorje-Gorski Kotar	13.0%	6.5%	7.6%	7.7%	1.2%	
Sisak-Moslavina	1.6%	3.1%	1.9%	0.7%	4.9%	
Split-Dalmatia	11.1%	15.9%	21.8%	13.7%	2.1%	
Šibenik-Knin	1.0%	2.9%	4.3%	2.8%	0.4%	
Varaždin	3.1%	3.7%	2.0%	1.1%	0.4%	
Virovitica-Podravina	0.4%	0.6%	0.5%	1.0%	9.1%	
Vukovar-Srijem	1.6%	2.4%	1.3%	0.5%	9.5%	
Zadar	1.7%	2.5%	3.9%	2.9%		
Zagreb	41.1%	39.4%	34.5%	47.8%	22.2%	

Source: Authors' work based on Viscovery SOMine output;

Note: ***statistically significant at 1%

Table 4 compares clusters according to the type of leasing object, whether the leasing object was new or used, and the leasing type (financial or operative). As for the leasing object, Cluster 1 is the only one that contains leasing contracts related to cars and light commercial vehicles as

leasing objects. In Cluster 2 those are trucks and tugboats, as well as trailers and semitrailers. Leasing contracts related to machines are assigned to Cluster 3 and Cluster 5 contains only leasing contracts related to agricultural machines and tractors. Cluster 4 is diverse when it comes to

leasing objects, and it included leasing contracts related to machines, forklifts, and vessels.

When it comes to the status of the leasing object, whether it is new or used, it can be noticed that in Cluster 2 the majority of leasing objects are used. In the other clusters, those are mostly new leasing objects with the highest percentage in Cluster 4.

Financial leasing is the most common leasing type in all the clusters. However, the highest percentage of financial leasing contracts is related to Cluster 5 and the highest percentage of operative leasing contracts is related to Cluster 1.

Table 5 compares clusters according to a different client and contract characteristics. When it comes to the status of the client, it can be noticed that in Cluster 5 new clients are in the majority, while in others those are old clients. Within Cluster 3 the highest percentage of the clients has the lowest rating R5, followed by R2. In the other clusters, rating R3 is most common. Finally, according to the attribute client risk, it can be noticed that the leasing company did not have data on the estimated client risk. The highest percentage of fraud or default cases occurred within Cluster 5 (18.55%) and Cluster 3 (17.1%).

Table 4: Clusters according to leasing object and leasing type

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Chi-square (p-value)
Leasing object						
Agri_forest					100.0%	50454.327 (0.000***)
Cars	66.8%					
Equipment	5.6%					
Forklifts				37.5%		
Light_commercial_vehicles	25.8%					
Machines	0.8%		100.0%			
Motors				40.9%		
Public		3.8%				
Trailers_semitrailers		24.9%				
Trucks_tugboats	1.0%	71.2%				
Vessels				21.3%		
Other				0.3%		
Leasing object New/Used						
New	67%	19.4%	59.4%	84.5%	84%	853.240 (0.000***)
Used	33.1%	80.6%	40.6%	15.5%	16%	
Leasing type						
Financial	64.4%	87.9%	85.8%	86.6%	98.8%	716.436 (0.000***)
Operative	35.6%	12.1%	14.2%	13.4%	1.2%	

Source: Authors' work based on Viscovery SOMine output;

Note: ***statistically significant at 1%

Table 5: Clusters according to client characteristics and fraud/default attributes

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Chi-square (p-value)
Client new/old						
New	48.5%	44.3%	42.5%	38.6%	67.9%	80.270 (0.000***)
Old	51.5%	55.7%	57.5%	61.4%	32.1%	
Client rating						
R2	2.2%	0.7%	0.7%	1.3%	1.2%	198.575 (0.000***)
R3	43.6%	43.2%	33.2%	38.3%	43.6%	
R4	29.5%	25.6%	22.8%	33.2%	25.1%	
R5	24.7%	30.6%	43.3%	27.2%	30.0%	
Client risk						
No_estimation	63.6%	86.2%	83.3%	85.3%	97.5%	762.799 (0.000***)
No_risk	24.9%	5.5%	6.2%	3.8%		
Risk	11.5%	8.3%	10.5%	11.0%	2.5%	
Fraud/default						
No	89.3%	87.7%	82.9%	90.2%	81.5%	48599.000 (0.000***)
Yes	10.7%	12.3%	17.1%	9.8%	18.5%	

Source: Authors' work based on Viscovery SOMine output;

Note: ***statistically significant at 1%

4.2 Fraud according to client and leasing characteristics

In this section, we present the ratio of fraud/default leasing contracts within clusters.

Table 6 presents details on fraud/default cases in each cluster according to the client sector, the client type, and the client county. Chi-squares results show significant associations between clusters 1, 4 and 5 and the attribute client sector. For example, in Cluster 1 31.8% of fraud/default cases were committed by clients doing business in the trade sector, followed by other services (27.40%) and construction (26.40%), which was significant at 1%. It can be noticed that the clients from the same sectors were fraudsters in Cluster 4 as well, which is also significant at a 1% level.

The client type was shown to be significant when it comes to frauds/defaults for clusters 2, 4 and 5. The highest percentage of SMEs committed fraud in Cluster 4. Sole proprietorships committed the highest percentage of frauds/defaults in Cluster 5.

Additionally, Chi-squares reveal significant associations between frauds in clusters 1, 2, 3 and 4 and the attribute client county.

Table 7 presents details on fraud/default cases within the clusters according to the operational attributes. Chi-squares results show significant associations between frauds and defaults in clusters 1, 2 and 4, and the attribute of the leasing object. In Cluster 1 fraud/default cases are mostly related to cars (62.50%), in Cluster 2 to trucks and tugboats, and in Cluster 4 to forklifts. The status of the leasing object was shown to be significant in Cluster 2 and 5.

Frauds/defaults in Cluster 2 are related to used leasing objects, and in Cluster 5 to new leasing objects. The leasing type is significant for fraud in Cluster 1 and 2.

In Cluster 1 frauds/defaults are related to operative leasing contracts in 57.60% of cases. For Cluster 2 those are financial leasing contracts in 83% cases.

Table 8 presents details on fraud/default cases in the clusters according to the behavioural attributes of clients. The attribute client new/old, and the attribute client risk are significantly associated with fraud/default cases for Cluster 1. New clients mostly committed fraud or default in Cluster 1. The client rating R5 is significant for frauds/defaults in all the clusters.

Table 6: Fraud and default cases according to the behavioural attributes of clients

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Client sector					
Agriculture	6.7%	1.3%		3.3%	97.8%
Chemical	0.1%				
Construction	26.4%	11.6%	44.4%	30.0%	
Financial	0.6%				
Other services	27.4%	58.6%	30.2%	20.0%	
Public	0.8%				
Tourism	6.2%	0.4%			
Trade	31.8%	28.1%	25.3%	46.7%	2.2%
Chi-square (p-value)	97.087 (0.000***)	6.771 (0.148)	0.142 (0.932)	14.498 (0.013**)	15.050 (0.005***)
Client type					
Company	74.5%	67.0%	68.5%	91.70%	8.9%
Sole proprietorship	25.5%	33.0%	31.5%	8.3%	91.1%
Chi-square (p-value)	0.198 (0.657)	13.527 (0.000***)	1.224 (0.269)	5.636 (0.018**)	8.667 (0.003***)
Client County					
Bjelovar-Bilogora			34.0%	51.7%	26.7%
Brod-Posavina	4.4%		17.9%	3.3%	33.3%
Dubrovnik-Neretva			6.8%	5.0%	
Istria	7.90%				2.2%
Karlovac					15.6%
Koprivnica-Križevci					13.3%
Krapina-Zagorje					4.4%
Primorje-Gorski Kotar	9.8%	4.9%			
Split-Dalmatia	7.5%	6.7%			
Zagreb	44.2%	46.9%			
Chi-square (p-value)	94.452 (0.000***)	52.968 (0.000***)	52.544 (0.000***)	72.064 (0.000***)	18.728 (0.283)

Source: Authors' work based on Viscovery SOMine output;

Note: ***statistically significant at 1%; ** 5%

Table 7: Fraud and default cases according to operational attributes of leasing contracts

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Leasing object					
Agri_forest					18.5%
Cars	62.5%				
Equipment	4.2%				
Forklifts				65.0%	
Light_commercial_vehicles	31.3%				
Machines	0.8%		17.1%		
Motors				15.0%	
Public		0.9%			
Trailers_semitrailers		36.6%			
Trucks_tugboats	1.3%	62.5%			
Vessels				20.0%	
Chi-square (p-value)	21.916 (0.001***)	22.395 (0.000***)	/	24.635 (0.000***)	/
Leasing object New/Used					
New	70.0%	25.0%	60.5%	86.7%	62.0%
Used	30.5%	75.0%	39.5%	13.3%	38.0%
Chi-square (p-value)	3.327 (0.068)	5.079 (0.024**)	0.103 (0.749)	0.249 (0.618)	19.352 (0.000***)
Leasing type					
Financial	42.4%	83.0%	81.5%	86.7%	100.0%
Operative	57.6%	17.0%	18.5%	13.3%	
Chi-square (p-value)	239.348 (0.000***)	5.707 (0.017**)	2.973 (0.085)	0.000 (0.983)	0.690 (0.406)

Source: Authors' work based on Viscosity SOMine output;

Note: ***statistically significant at 1%; ** 5%

Table 8: Fraud and default cases according to behavioural attributes of clients

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Client new/old					
New	60.1%	44.2%	42.6%	43.3%	82.2%
Old	39.9%	55.8%	57.40%	56.7%	17.8%
Chi-square (p-value)	61.095 (0.000***)	0.001 (0.971)	0.000 (0.985)	0.622 (0.430)	5.197 (0.023**)
Client rating					
R2	0.2%				
R3	1.7%	4.50%	0.6%	1.7%	2.2%
R4	4.9%	3.10%		1.7%	
R5	93.2%	92.40%	99.4%	96.70%	97.8%
Chi-square (p-value)	2861.991 (0.000***)	459.771 (0.000***)	250.596 (0.000***)	162.408 (0.000***)	120.594 (0.000***)
Client risk					
No_estimation	48.6%	83.9%	78.4%	83.3%	100.0%
No_risk	38.5%	5.4%	9.3%	3.3%	
Risk	12.8%	10.7%	12.3%	13.3%	
Chi-square (p-value)	127.931 (0.000***)	1.928 (0.381)	4.093 (0.129)	0.402 (0.818)	1.398 (0.237)

Source: Authors' work based on Viscovery SOMine output;

Note: ***statistically significant at 1%; ** 5%

4.3 Summary of cluster characteristics in relation to fraud

Based on our research, this section provides the characteristics of clusters according to characteristics of leasing contracts and fraudulent behaviour occurrence. It also presents the occurrence of fraud in relation to the type of vehicle.

Cluster 1 - New cars / Trade. Cluster 1 contains the largest percentage of clients doing business in the trade sector (41.4%). It is also the only cluster in which clients are interested in cars as leasing objects (66.8% of clients), but it also consists of clients that are interested in light commercial vehicles (25.8% of clients). The leasing object is new in 67% cases, reflecting the fact that the new cars are an object of the majority of the leasing agreements. Leasing contracts of this cluster were financial in 64.4% of cases. In this cluster, clients are also mostly small or medium companies (75.1%) from Zagreb County (41.1%), and they are old company's clients (48.5%). Their rating is on average R3, and their risk is not estimated (63.6%).

More than 10% of leasing contracts in this cluster were fraudulent. In 93.20% of cases, fraudsters had low client rating R5. Therefore, when it comes to fraudster profiles, leasing companies should take care of new clients that do

business in trade, other services and construction in Zagreb County. Additionally, a client's risk in 48.6% of cases was not estimated, or they were no risky clients (38.5%). Operative leasing contracts were prone to risk, and they were related to cars and light commercial vehicles. An additional analysis showed that VW, Peugeot, and Citroen vehicles are riskier than the other vehicle brands.

Cluster 2 - Used trucks or tugboats / Other services. Clients in Cluster 2 perform business in other services or trade (82.7%). This is the only cluster in which clients are interested in trucks and tugboats (in 71.2% of cases) as well as trailers and semitrailers (in 24.9% of cases). Most of the leasing objects are used (80.6%), and 87.9% leasing contracts are financial. Furthermore, more than half of the clients in this cluster are from Zagreb or Split-Dalmatia County. When it comes to behavioural attributes of a client, 55.7% of them in this cluster are old clients. Their risk is not estimated in 86.2% of cases. Client rating in this cluster is R3 in 43.2% of cases, but there is also a large proportion of clients with low rating R5 (30.6% of cases).

In Cluster 2 12.3% of leasing contracts were fraudulent. An analysis of fraudulent cases for this cluster showed small and medium companies were fraudsters in 67% of fraud cases, and they were mostly from Zagreb County (46.9%) with rating R5. Additionally, used vehicles as well as trailers and semitrailers, especially MAN

trucks, were the object of most of the fraud cases. Financial leasing contracts are especially risky in this cluster.

Cluster 3 - New machinery / Construction. Cluster 3 is the one with the largest percentage of clients doing business in the construction sector (43.3%), followed by other services and trade. Most of the clients are small or medium companies (64.7) in Zagreb or Split-Dalmatia County. Machines are the only type of leasing objects in this cluster, and they are new in 59.4% of the cases. The leasing contracts are financial. In this cluster, clients are known to the company from previous agreements (57.5%). This is the cluster with the highest percentage of R5 rated clients (43.3%), but their risk is not estimated (83.3%).

This cluster has a high percentage of fraud/default cases (17.1%). Clients from the construction sector committed 44.4% of fraud cases. Fraudsters are from Bjelovar-Bilogora County in 34% of cases, and their rating is R5.

Cluster 4 - New motors / Trade. Cluster 4 has the highest percentage of clients doing business in the trade sector (46.8%) followed by other services (29%). In 80% of cases, those are small or medium companies, and in more than 60% they are from Zagreb or Split – Dalmatia County. These companies are interested in motors (40.9%) and forklifts (37.5%) as leasing objects, especially new ones. Financial leasing contracts are the main type of leasing in this cluster. When it comes to behavioural attributes of clients, in this cluster, clients are mostly old clients, with the average rating R3 or R4.

This cluster has the lowest fraud rate of all the clusters (9.8%). Fraudulent cases are related to small and medium companies from Bjelovar-Bilogora County with an R5 rating. They also do business in the trade or the construction sector. In 65% of cases, the leasing object in fraudulent leasing contracts is a forklift.

Cluster 5 - New machinery and tractors / Agriculture. Clients in Cluster 5 do business in the agricultural or the trade sector. Sole proprietorships are the main type of clients. This is the only cluster in which agricultural machinery and tractors are objects of leasing contracts. Additionally, leasing objects are new in most cases. The primary type of leasing is financial. This can be explained by the fact that agricultural sole proprietorships, in reali-

ty, want to keep machinery and tractors for a longer term, after the termination of a lease. This cluster contains the largest percentage of new clients without a risk estimation.

Furthermore, this cluster contains the largest percentage of fraud cases (18.5%). Frauds are likely to be committed by new, low rated agricultural proprietorships interested in new agricultural machinery and tractors.

5 Practical recommendations

In order to check the validity of our approach, we asked experts from leasing companies to evaluate whether the observed results are useful to them and whether they are in line with their observations in practice. In our research, we followed the approach of Osei-Bryson (2010), who used expert evaluation of clustering results in one data mining application. We asked four experts, from four different Croatian companies, to provide their opinion in relation to the cluster characteristics. They confirmed that the given results are applicable in their day-to-day business operations, as well as tactical and strategic planning.

Finally, with the support of experts, we have developed Table 8, which presents the summary of the characteristics of fraudulent contracts within each cluster, which can be useful to leasing companies in their development of fraud prevention and detection programmes. The table presents the characteristics of leasing contracts within clusters, which have been proved as statistically significant and thus useful for the identification of fraudulent clients. For example, fraud occurs most often in Cluster 1 with the construction industry, other sectors, and trade, which is significant at 1%.

There are several practical recommendations that could be derived from Table 9. For example, companies should take special care of the clients coming from the construction industry, other services and trade, which operate in Zagreb County, with new cars and light commercial vehicles as leasing objects, especially if the operative leasing is used. Similar recommendations could be derived from other clusters.

Table 9: Fraud or default profiles within clusters

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Client sector	○ (1%, construction, other, trade)	/	○ (5%, construction, other, trade)	○ (5%, trade, construction, other)	○ (1%, agriculture)
Client type	/	○ (1%, SME)	/	○ (5%, SME)	○ (1%, sole proprietorship)
Client County	○ (1%, Zagreb)	○ (1%, Zagreb)	○ (1%, Bjelovar-Bilogora, Brod-Posavina)	○ (1%, Bjelovar-Bilogora)	/
Leasing object	○ (1%, cars and light commercial vehicles)	○ (1%, trailer semitrailer; truck_tugboat)	/	○ (1%, forklift)	/
Leasing object (New/Used)	/	○ (5%, used)	/	/	○ (1%, new)
Leasing type	○ (1%, operative)	○ (5%, financial)	/	/	/
(Client new/old)	○ (1%, new)	/	/	/	○ (5%, new)
Client rating	○ (1%, R5)	○ (1%, R5)	○ (1%, R5)	○ (1%, R5)	○ (1%, R5)
Client risk	(1%, no_estimation; no_risk)	/	/	/	/

Source: Authors 'calculations and Viscovery SOMine output;

Note: ○ Statistically significant at 1% and 5%; / - no significance

6 Conclusions

The objective of this work is to shed some light on the area of fraud in the leasing industry, with support of the data mining approach utilizing cluster analysis with self-organising maps. Research goals were: (i) to investigate whether SOMs is an appropriate method for identifying and describing clusters of clients in the context of the leasing industry; (ii) to detect specific attributes that could explain the fraud in the leasing industry.

We applied the SOM algorithm with the usage of Viscovery SOM software on a database of one Croatian leasing company, which resulted in the identification of five clusters of leasing contracts according to their characteristics, such as the client sector, the leasing object, and the leasing type. The application of the SOM algorithm resulted in the extraction of five clusters, with the significant differences in relation to the leasing contract characteristics. We have asked several experts from other leasing companies to evaluate the usefulness of our results. They have confirmed that the results are in line with their observations, as well as the practices of their companies. Therefore, the usage of the SOM-Ward algorithm with the support of Viscovery SOMine software proved to be useful for the cluster analysis of the clients of the leasing company, which indicates a positive answer to our first research question.

In order to detect specific attributes that could explain the fraud in the leasing industry, the clusters were inter-

preted according to all clustering and other attributes used for the analysis. We used Chi-Square tests in order to detect a significant association between attributes' modalities and the occurrence of fraud and default cases within each of the clusters. Based on our results, we identified fraudster profiles based on the attributes that explain committed frauds or defaults.

Our work indicates the potential practical implications. Although our work is based on a client database from one Croatian leasing company, the expert evaluation of the clustering results indicates that other leasing companies could also benefit from the developed fraudster profiles. In addition, other leasing companies could develop their own analyses, based on the same methodology and implement fraud alert mechanisms in their client databases. This means that they also can increase their efficiency and effectiveness by creating customised business strategies for different clusters of clients. The findings of this paper can be used for further adaptation of the methodology in fraud profiling in contexts of different industries.

However, several limitations should be taken into account when it comes to this research. First, we focused only on small and medium companies and sole proprietorships in one industry. Second, data are provided only by one leasing company. Therefore, future research should include data provided by more companies in order to enhance the robustness of the results. Additionally, the methodology should be tested in the context of other industries, such as insurance. Testing the proposed method on other

case studies and industries could enhance the robustness of the results.

Acknowledgement

This research has been fully supported by the Croatian Science Foundation under the PROSPER (Process and Business Intelligence for Business Performance) project (IP-2014-09-3729).

Literature

- Abbas, O.A. (2008). Comparison between Data Clustering Algorithms. *The International Arab Journal of Information Technology*, 5(3), 320-325.
- Alavi, M., & Leidner, D.E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, 25(1), 107-136. <http://doi.org/10.2307/3250961>
- Almendra, V. D., & Enachescu, D. (2013). Using Self-Organizing Maps for Fraud Prediction at Online Auction Sites. *2013 15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, 281-288. <http://doi.org/10.1109/synasc.2013.44>
- Baço, F., Lobo, V., & Painho, M. (2005). Self-organizing Maps as Substitutes for K-Means Clustering. *Lecture Notes in Computer Science Computational Science – ICCS 2005*, 3516, 476-483. http://doi.org/10.1007/11428862_65
- Balasupramanian, N., Ephrem, B. G., & Al-Barwani, I. S. (2017). User pattern based online fraud detection and prevention using big data analytics and self organizing maps. *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, 691-694. <http://doi.org/10.1109/icicict1.2017.8342647>
- Bănărescu, A. (2015). Detecting and Preventing Fraud with Data Analytics. *Procedia Economics and Finance*, 32, 1827-1836. [http://doi.org/10.1016/S2212-5671\(15\)01485-9](http://doi.org/10.1016/S2212-5671(15)01485-9)
- Basel Committee on Banking Supervision. (2002). *Operational Risk Data Collection Exercise*. Retrieved March 28, 2019, from <http://www.bis.org/bcbs/qis/oprdata.pdf>
- Boobyer, C. (2003). *Leasing and Asset Finance: The Comprehensive Guide for Practitioners*. London: Euromoney Books.
- 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), 245-274.
- Caldeira, A. M., Gassenferth, W., Machado, M. A., & Santos, D. J. (2015). Auditing Vehicles Claims Using Neural Networks. *Procedia Computer Science*, 55, 62-71. <http://doi.org/10.1016/j.procs.2015.07.008>
- Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2019). Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences, In Press*. <http://doi.org/10.1016/j.ins.2019.05.042>
- Carneiro, N., Figueira, G., & Costa, M. (2017). A data mining based system for credit-card fraud detection in e-tail. *Decision Support Systems*, 95, 91-101. <http://doi.org/10.1016/j.dss.2017.01.002>
- Chen, Y.-KJ., Liou, W.-C., Chen, Y.-M., & Wu, J.-H. (2019). Fraud detection for financial statements of business groups. *International Journal of Accounting Information Systems*, 32, 1-23., ISSN 1467-0895, <https://doi.org/10.1016/j.accinf.2018.11.004>.
- Chouiekh, A., & Haj, E. H. (2018). ConvNets for Fraud Detection analysis. *Procedia Computer Science*, 127, 133-138. <http://doi.org/10.1016/j.procs.2018.01.107>
- Dorransoro, J., Ginel, F., Sánchez, C., & Cruz, C. (1997). Neural fraud detection in credit card operations. *IEEE Transactions on Neural Networks*, 8(4), 827-834. <http://doi.org/10.1109/72.595879>
- Eshghi, A., & Kargari, M. (2019). Introducing a new method for the fusion of fraud evidence in banking transactions with regards to uncertainty. *Expert Systems with Applications*, 121, 382-392. <http://doi.org/10.1016/j.eswa.2018.11.039>
- European Commission. (2011). EU Accounting Rule 8 Leases. Retrieved April 5, 2019, from https://ec.europa.eu/info/sites/info/files/about_the_european_commission/eu_budget/eu-accounting-rule-8-leases_2011_en.pdf
- Fiore, U., Santis, A. D., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, 479, 448-455. <http://doi.org/10.1016/j.ins.2017.12.030>
- Flath, D. (1980). The economics of short-term leasing. *Economic inquiry*, 18(2), 247-259.
- Folorunso, O., & Ogunde, A. (2005). Data mining as a technique for knowledge management in business process redesign. *Information Management & Computer Security*, 13(4), 274-280. <http://doi.org/10.1108/09685220510614407>
- Hainaut, D. (2019). A self-organizing predictive map for non-life insurance. *European Actuarial Journal*, 9(1), 173-207.
- Holmbom, A. H., Eklund, T., & Back, B. (2011). Customer portfolio analysis using the SOM. *International Journal of Business Information Systems*, 8(4), 396-412. <http://doi.org/10.1504/ijbis.2011.042397>
- Horvat, I., Pejić Bach, M., & Merkač Skok, M. (2014). Decision tree approach to discovering fraud in leasing agreements. *Business Systems Research Journal*, 5(2), 61-71. <http://doi.org/10.2478/bsrj-2014-0010>
- Jian, L., Ruicheng, Y., & Rongrong, G. (2016). Self-organizing map method for fraudulent financial data detection. In *2016 3rd International Conference on Information Science and Control Engineering (ICISCE)* (pp.

- 607-610). <http://doi.org/10.1109/ICISCE.2016.135>
- Leite, R. A., Gschwandtner, T., Miksch, S., Gstrein, E., & Kuntner, J. (2018). Visual analytics for event detection: Focusing on fraud. *Visual Informatics*, 2(4), 198-212. <http://doi.org/10.1016/j.visinf.2018.11.001>
- Lucas, Y., Portier, P.-E., Laporte, L., He-Guelton, L., Calen, O., Granitzer, M., & Calabretto, S. (2020). Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs. *Future Generation Computer Systems*, 102, 393-402. <http://doi.org/10.1016/j.future.2019.08.029>
- Merkevicius, E., Garšva, G., & Simutis, R. (2004). Forecasting of credit classes with the self-organizing maps. *Information Technology and Control*, 33(4), 61-66. Retrieved March 25, 2019, from <http://itc.ktu.lt/index.php/ITC/article/view/11956>
- Moradi, M., Salehi, M., Ghorgani, M. E., & Yazdi, H. S. (2013). Financial distress prediction of Iranian companies using data mining techniques. *Organizacija*, 46(1), 20-27. <http://dx.doi.org/10.2478/orga-2013-0003>
- Nami, S., & Shajari, M. (2018). Cost-sensitive payment card fraud detection based on dynamic random forest and k -nearest neighbors. *Expert Systems with Applications*, 110, 381-392. <http://doi.org/10.1016/j.eswa.2018.06.011>
- Ngai, E., Hu, Y., Wong, Y., Chen, Y., & Sun, X. (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), 559-569. <http://dx.doi.org/10.1016/j.dss.2010.08.006>
- Nian, K., Zhang, H., Tayal, A., Coleman, T., & Li, Y. (2016). Auto insurance fraud detection using unsupervised spectral ranking for anomaly. *The Journal of Finance and Data Science*, 2(1), 58-75. <http://doi.org/10.1016/j.jfds.2016.03.001>
- Olszewski, D. (2014). Fraud detection using self-organizing map visualizing the user profiles. *Knowledge-Based Systems*, 70, 324-334. <http://doi.org/10.1016/j.knsys.2014.07.008>
- Osei-Bryson, K. M. (2010). Towards supporting expert evaluation of clustering results using a data mining process model. *Information Sciences*, 180(3), 414-431.
- Patel, R., & Singh, D. (2013). Credit Card Fraud Detection & Prevention of Fraud Using Genetic Algorithm. *International Journal of Soft Computing*, 6. Retrieved March 25, 2019, from <http://www.ijscce.org/attachments/File/v2i6/F1189112612.pdf>
- Patil, S., Nemade, V., & Soni, P. K. (2018). Predictive Modelling For Credit Card Fraud Detection Using Data Analytics. *Procedia Computer Science*, 132, 385-395. <http://doi.org/10.1016/j.procs.2018.05.199>
- Pejić Bach, M., Juković, S., Dumičić, K., & Šarlija, N. (2014). Business Client Segmentation in Banking Using Self-Organizing Maps. *South East European Journal of Economics and Business*, 8(2), 32-41. <http://doi.org/10.2478/jeb-2013-0007>
- Pejić Bach, M., Vlahović, N., & Pivar, J. (2018). Self-organizing maps for fraud profiling in leasing. *41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, 2018*, 1203-1208. <http://doi.org/10.23919/MIPRO.2018.8400218>
- Quah, J. T., & Sriganesh, M. (2008). Real-time credit card fraud detection using computational intelligence. *Expert Systems with Applications*, 35(4), 1721-1732. <http://doi.org/10.1016/j.eswa.2007.08.093>
- Rousseuw, P., Perrotta, D., Riani, M., & Hubert, M. (2019). Robust Monitoring of Time Series with Application to Fraud Detection. *Econometrics and Statistics*, 9, 108-121. <http://doi.org/10.1016/j.ecosta.2018.05.001>
- Ryman-Tubb, N. F., Krause, P., & Garn, W. (2018). How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. *Engineering Applications of Artificial Intelligence*, 76, 130-157. <http://doi.org/10.1016/j.engappai.2018.07.008>
- Sadgali, I., Sael, N., & Benabbou, F. (2019). Performance of machine learning techniques in the detection of financial frauds. *Procedia Computer Science*, 148, 45-54. <http://doi.org/10.1016/j.procs.2019.01.007>
- Singleton, T. W., & Singleton, A. J. (2007). Why don't we detect more fraud? *Journal of Corporate Accounting & Finance*, 18(4), 7-10.
- 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), 1039-1052. <http://doi.org/10.1016/j.eswa.2010.07.143>
- 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*, In press. <http://doi.org/10.1016/j.jksuci.2017.09.010>
- Tu, B., He, D., Shang, Y., Zhou, C., & Li, W. (2019). Deep feature representation for anti-fraud system. *Journal of Visual Communication and Image Representation*, 59, 253-256. <http://doi.org/10.1016/j.jvcir.2019.01.031>
- Uribe, C., & Isaza, C. (2012). Expert knowledge-guided feature selection for data-based industrial process monitoring. *Revista Facultad De Ingenieria Universidad De Antioquia*, 65, 112-125. Retrieved March 25, 2019, from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-62302012000400009
- Urueña López, A., Mateo, F., Navío-Marco, J., Martínez-Martínez, J. M., Gómez-Sanchís, J., Vila-Francés, J., & Serrano-López, A. J. (2019). Analysis of computer user behavior, security incidents and fraud using Self-Organizing Maps. *Computers & Security*, 83, 38-51. <http://doi.org/10.1016/j.cose.2019.01.009>
- Van Hulle, M.M. (2012). *Self-organizing Maps*. In Rozenberg G., Bäck T., & Kok J.N. (Eds.) *Handbook of Natural Computing*. Berlin: Springer, Heidelberg.

- Van Laerhoven K. (2001) Combining the Self-Organizing Map and K-Means Clustering for On-Line Classification of Sensor Data. In: *Dorffner G., Bischof H., Hornik K. (eds) Artificial Neural Networks — ICANN 2001. ICANN 2001. Lecture Notes in Computer Science, 2130*, 464–469, Springer, Berlin, Heidelberg.
- Viaene, S., Dedene, G., & Derrig, R. (2005). Auto claim fraud detection using Bayesian learning neural networks. *Expert Systems with Applications*, 29(3), 653–666. <http://doi.org/10.1016/j.eswa.2005.04.030>
- Viscovery. (2019). The Ward cluster algorithm of Viscovery SOMine. Retrieved April 5, 2019, from <https://www.viscovery.net/download/public/The-SOM-Ward-cluster-algorithm.pdf>
- Wang, D., Cheng, B. & Chen, J. (2019). Credit card fraud detection strategies with consumer incentives. *Omega*, 88, 179–195. <http://doi.org/10.1016/j.omega.2018.07.001>
- Wang, H., & Wang, S. (2008). A knowledge management approach to data mining process for business intelligence. *Industrial Management & Data Systems*, 108(5), 622–634.
- Wang, Q., Xu, W., Huang, X., & Yang, K. (2019). Enhancing intraday stock price manipulation detection by leveraging recurrent neural networks with ensemble learning. *Neurocomputing*, 347, 46–58. <http://doi.org/10.1016/j.neucom.2019.03.006>
- Wang, Y., & Xu, W. (2018). Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems*, 105, 87–95. <http://doi.org/10.1016/j.dss.2017.11.001>
- Wehrens, R., & Buydens, L. (2007). Self- and Super-organizing Maps in R: The kohonen Package. *Journal of Statistical Software*, 21(5). Retrieved March 25, 2019, from <https://www.jstatsoft.org/article/view/v021i05>
- West, J., & Bhattacharya, M. (2016). Some Experimental Issues in Financial Fraud Mining. *Procedia Computer Science*, 80, 1734–1744. <http://doi.org/10.1016/j.procs.2016.05.515>
- Yan, C., Li, M., Liu, W., & Qi, M. (2020). Improved adaptive genetic algorithm for the vehicle Insurance Fraud Identification Model based on a BP Neural Network. *Theoretical Computer Science*, 817, 12–23. <http://doi.org/10.1016/j.tcs.2019.06.025>
- Yan, C., Li, Y., Liu, W., Li, M., Chen, J., & Wang, L. (2019). An artificial bee colony-based kernel ridge regression for automobile insurance fraud identification. In press: *Neurocomputing*. <http://doi.org/10.1016/j.neucom.2017.12.072>
- Zakaryazad, A., & Duman, E. (2016). A profit-driven Artificial Neural Network (ANN) with applications to fraud detection and direct marketing. *Neurocomputing*, 175, 121–131. <http://doi.org/10.1016/j.neucom.2015.10.042>
- Zareapoor, M., & Shamsolmoali, P. (2015). Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier. *Procedia Computer Science*, 48, 679–685. <http://doi.org/10.1016/j.procs.2015.04.201>
- Zaslavsky, V., & Strizhak, A. (2006). Credit Card Fraud Detection Using Self-Organizing Maps. *Information & Security: An International Journal*, 18, 48–63. <http://doi.org/10.11610/isij.1803>
-
- Mirjana Pejić Bach**, is a Full Professor at the Department of Informatics at the Faculty of Economics & Business. She graduated at the Faculty of Economics & Business – Zagreb, where she also received her Ph.D. degree in Business in the area of system dynamics applications. She is the recipient of the Emerald Literati Network Awards for Excellence for the paper Influence of strategic approach to BPM on financial and non-financial performance published in *Baltic Journal of Management*. Mirjana was also educated at MIT Sloan School of Management in the field of System Dynamics Modelling, and at OliviaGroup in the field of data mining. She participates in number of EU FP7 projects, and is an Expert for Horizon 2020. Her current research interests are big data, project management, data mining, simulation games and system dynamics.
-
- Nikola Vlahović**, is an Associate Professor at the Department of Informatics at the Faculty of Economics & Business. He received his Ph.D. in Information and Communication Sciences at Faculty of Organisation and Informatics in Varaždin, University of Zagreb, Croatia. He participated in international and national scientific research projects and commercial projects. His current research interest are decision support methods, expert systems, artificial intelligence, and business application development.
-
- Jasmina Pivar**, is currently employed as a teaching and research assistant at the Department of Informatics, Faculty of Economics and Business, University of Zagreb, where she is also enrolled in a postgraduate doctoral study program. She graduated with a degree in economics from the Faculty of Organisation and Informatics in Varaždin and earned the Dean's Award for excellence in 2009, 2010, 2011, 2012 and 2013. Her main research interests are big data, smart city, data mining, Internet of Things and technology adoption.

Preprečevanje goljufij pri lizingu z uporabo Kohonen-ovih samoorganizirajočih zemljevidov

Ozadje in namen: Tehnike rudarjenja podatkov se intenzivno uporabljajo v različnih panogah za preprečevanje in odkrivanje goljufij. Raziskav, ki se osredotočajo na lizing industrijo, je malo, čeprav se goljufije na tem področju pojavljajo precej pogosto. V študiji najprej identificiramo grozde poslovnih strank v izbrani lizingi družbi po metodi samoorganizirajočih zemljevidov, ki temeljijo na atributih lizing pogodb. Nato primerjamo grozde na podlagi prisotnosti goljufivih strank, da bi razvili profile prevarantov.

Zasnova / metodologija / pristop: Za odkrivanje značilnosti goljufivih strank smo uporabili bazo strank, ki vsebuje attribute lizing pogodb ene od hrvaških lizingi družb. Za razvoj profilov goljufivih odjemalcev smo uporabili postopek združevanja s Kohonen samoorganizirajočimi zemljevidi, ki jih podpira programska oprema Viscovery SOMine.

Rezultati: Identificirali smo pet skupin in jih označili v skladu z modalnimi vrednostmi atributov, ki opisujejo predmet lizinga in panogo, v kateri stranka posluje: (i) novi avtomobili / trgovina; (ii) rabljeni tovornjaki ali vlačilci / druge storitve; (iii) novi stroji / gradbeništvo; (iv) novi motorji / trgovina; in (v) novi stroje in traktorji / Kmetijstvo.

Zaključek: Samoorganizirajoči zemljevidi so se izkazali kot uporabna metodologija za razvoj profilov goljufivih strank v lizingi družbah. Podjetja lahko naše rezultate za ciljno spremljanje strank iz opredeljenih panog, ki kupujejo specifične lizing predmete. Poleg tega lahko podjetja uporabijo našo metodologijo v lastnih bazah podatkov, da razvijejo profile prevarantov za njihove posebne namene in v svoje baze podatkov strank vgradijo mehanizme za opozarjanje na goljufije.

Ključne besede: *prevara, zakup, Kohonenovi samoorganizirajoči zemljevidi, Viscovery SOMine, Ward algoritem, Hrvaška, rudarjenje podatkov*