UHASSELT

KNOWLEDGE IN ACTION

Faculteit Bedrijfseconomische Wetenschappen

master handelsingenieur in de beleidsinformatica

Masterthesis

Outlier detection and its applications in the fraud detection

Stef Breuls

Scriptie ingediend tot het behalen van de graad van master handelsingenieur in de beleidsinformatica

PROMOTOR :

dr. Gonzalo NAPOLES RUIZ

2019
2020

UHASSELT

KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT

KNOWLEDGE IN ACTION

Faculteit Bedrijfseconomische Wetenschappen

master handelsingenieur in de beleidsinformatica

Masterthesis

Outlier detection and its applications in the fraud detection

Stef Breuls

Scriptie ingediend tot het behalen van de graad van master handelsingenieur in de beleidsinformatica

PROMOTOR :

dr. Gonzalo NAPOLES RUIZ

This master thesis was written during the COVID-19 crisis in 2020. This global health crisis might have had an impact on the (writing) process, the research activities and the research results that are at the basis of this thesis.

Abstract

Fraud detection is an important task for many organizations is today's connected and rapidly changing world. The use of an outlier detection method is a common way of dealing with fraud detection. Numerous outlier detection techniques have been developed and researched within diverse research and application domains. In this paper, we try to present a comprehensive overview of different outlier detection methods and applications in the fraud detection. The choice of an appropriate method is important, therefore we identified some possible factors that can influence the method choice. Different methods for outlier detection are provided and structured by grouping them into categories. A basic explanation and some examples are given for each method, as well as advantages and disadvantages per category. Furthermore, we collected common fraud detection applications and analyzed how the chosen outlier detection methods handle specific outliers. We hope this paper provides a better understanding of the possible directions and challenges of outlier detection methods and their uses in the fraud detection domain.

1 Introduction

For many organizations, fraud detection is an essential task. It has become more important with the rapid development of digital technologies and e-services, as this created large networks that generate an enormous amount of data. Examples are telecommunication networks, banking and insurance networks and trading networks. The large data generation makes it easier to conceal fraudulent activities and creates possible opportunities for fraudsters (Pourhabibi, Ong, Kam, & Boo, 2020; X. Zhang, Han, Xu, & Wang, 2019). The detection of fraudulent activities is a necessity to reduce possible losses in which these activities can result. In the healthcare domain for example, financial losses due to fraud amount to 98 billion dollar per year in the United States alone (Branting, Reeder, Gold, & Champney, 2016). In auto-insurance claims, roughly 21 to 36% of claims involve suspected fraud elements (Tennyson & Salsas-Forn, 2002). Thus, there is an increasing need for detecting fraud to avoid economic losses for both insurance companies and policy holders (Nian, Zhang, Tayal, Coleman, & Li, 2016).

A widely used method in fraud detection is the use of outlier detection. Outlier detection is a generic term for various techniques and approaches to discover outlying observations in data. (Hodge & Austin, 2004). Besides fraud detection, it is also used in several other domains, such as cyber security, safety systems and smart homes. Applications within these domains that use outlier detection can generate actionable and potentially critical insights (Singh & Upadhyaya, 2012) as outliers often have a substantial relevance and may strongly influence the desired result. Outliers can be present in data due to numerous factors like human error, mechanical error, changes in system behaviour or fraudulent behaviour (Chandola, Banerjee, & Kumar, 2009; Peter J. Rousseeuw & Hubert, 2011).

However, there seems to be no commonly accepted definition for an outlier. In the literature, authors use many definitions to describe outliers. One of oldest and most used definitions is the one of Grubbs (1969):

An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs.

A second definition that is often used in the literature is the one of Hawkins (1980):

An outlier is an observation which deviates so much from the other observations as to arouse suspicion that it was generated by a different mechanism.

In the reference work of Barnett and Lewis (1994), outliers are described as:

An outlier is an observation (or subset of observations) which appear to be inconsistent with the remainder of the dataset.

A more recent definition is the one of Ramaswamy, Rastogi, and Shim (2000), which is based on the previously mentioned reference work:

An outlier in a set of data is an observation or a point that is considerably dissimilar or inconsistent with the remainder of the data.

As proven by the different definitions, outliers and outlier detection are widely reported in the literature. Extensive reviews on outlier detection techniques have been conducted within diverse fields such as machine learning and statistics. Many of these techniques apply concepts of various domains to a specific problem that might have a different notion of outliers than others. This makes it difficult to adopt certain techniques in other domains. In addition, approaches can be fundamentally the same, but named differently by the author which makes it more difficult to have an overview of information. Some of the most common names are outlier detection, novelty detection, anomaly detection or deviation detection (Hodge & Austin, 2004; Singh & Upadhyaya, 2012).

In this paper, we conduct a comprehensive review on outlier detection techniques by bringing together key information from multiple sources. The focus of our review study is on two main concerns. The first part describes outlier detection in general, what we do have to keep in mind when solving an outlier detection problem and what different methods exist to do so. In the second part, we elaborate on different applications of outlier detection techniques in the fraud detection, as described in the literature.

The remainder of this paper is structured as follows. In section 2 we describe the importance of outlier detection, where outlier detection is used and what challenges outlier detecting can imply. This is followed by an analysis of factors that can influence the outlier detection method choice for a certain outlier detection problem in section 3. A brief review of important existing methods, together with their strengths and weaknesses is given in section 4. Section 5 shows applications of outlier detection techniques within the fraud detection domain, how they handle outliers and what method they use. In section 6, we phrase our conclusions.

2 Outlier detection

2.1 Importance of outlier detection

Outlier detection is an important process because outliers may indicate errors, such as an incorrect entry or a missing value, but they can also indicate exceptional circumstances or fraudulent cases. A key characteristic of an outlier is that they are interesting to analyse and often have a considerable relevance (Chandola et al., 2009; Peter J. Rousseeuw & Hubert, 2011).

There are two possibilities when outliers are detected. The outlying observations is either an extreme manifestation of the random variability which is present in the data. If this is the case, the outlying observations must not be deleted and treated in the same way as the other observations. As an example, think of a few very tall people in the dataset of a certain population. The other possibility is when the outlying observation arose due to an error in the calculation or recording of the value or when there is a deviation from the prescribed experimental procedure. An investigation may be advisable to determine the reason of the outlying observation and if necessary, the value of the outlying observation may be rejected (Grubbs, 1969).

Outliers as a result of human errors or instrument reading error can be harmless and simply be corrected or deleted, while an outlier caused by an intrusion can be harmful, for example in a safety critical environment. They often need to be dealt with quickly to prevent possible damage (Hodge & Austin, 2004). From a machine learning perspective, outliers can be useful in a data cleaning task as the detection and removal can contribute to outlier-free datasets. This allows for more accurate modelling and prediction tasks (Domingues, Filippone, Michiardi, & Zouaoui, 2018).

2.2 Uses of outlier detection

Due to its importance, outlier detection is a frequently used method in numerous applications from different domains. For example, outliers in data regarding auto-insurance claims could indicate insurance fraud (Nian et al., 2016). Intrusion detection systems can analyse network traffic and detect unusual network behaviour and possible emerging cyberthreats (Kumar, 2005). Outlier detection systems in spacecraft can improve the necessary autonomy during space missions (Meß, Dannemann, & Greif, 2019).

A more exhaustive list of applications that use outlier detection (Hodge & Austin, 2004) is displayed in table 1:

Fraud detection	Detection of fraudulent applications for credit cards and fraudulent usage of credit cards (Panigrahi, Kundu, Sural, & Majumdar, 2009)

Loan application processing	Detection of fraudulent application or potentially problematical customers (Zhan & Yin, 2018)
Intrusion detection	Detection of unauthorised access in computer network (Lane & Brodley, 1997a)
Activity monitoring	Detection of mobile phone fraud or suspicious trades in equity markets (Fawcett & Provost, 1999)
Network performance	Detection of network bottlenecks by monitoring computer network performance (Weiss & Hirsh, 1998)
Fault diagnosis	Detection of faults in motors, generators, pipelines, space instruments,… by monitoring processes (Decoste & Levine, 2000)
Structural defect detection	Detection of faulty production runs by monitoring manufacturing lines (Susto, Terzi, & Beghi, 2017)
Satellite image analysis	Detection of novel or misclassified features (Meß, Dannemann, & Greif, 2019)
Image novelty detection	Detection of novelties for robot neotaxis or surveillance systems (Marsland, 2001)
Motion segmentation	Detection of image features moving independently of the background (Torr & Murray, 1997)
Structural health monitoring	Detection of changes or damages in safety critical applications such as drilling or high-speed milling (Gul & Necati Catbas, 2009)
Medical condition monitoring	Detection of outliers in patient-management decisions using electronic health records (Hauskrecht et al., 2013)
Text novelty detection	Detection of onset of news stories, for topic tracking or for traders (Allan, Carbonell, Doddington, Yamron, & Yang, 1998)
Mislabelled data detection	Detection of mislabelled data in training data sets (Brodley & Friedl, 1996)

Table 1: List of applications that use outlier detection

2.3 Challenges

In a simple form, outlier detection can be done by defining normal behaviour and identifying all observations which do not behave as defined. In reality, there are some challenges which increase the difficulty of this seemingly simple approach.

It is often very difficult to define normal behaviour because it can be impossible to be aware of or to include every possible normal behaviour. Besides that, the border between normal and abnormal behaviour can be imprecise. In today's rapidly changing world, normal behaviour can change and expand, which makes the current concept of normal behaviour possibly less representative in the future (Chandola et al., 2009).

When a malicious event is the cause of an outlying observation, they are often adapted to make them seem normal. This will complicate the process of defining normal behaviour. The domain in which the normal behaviour is defined is of importance, because applications in different domains can have different views on the concept of normal behaviour and often some domain-specific knowledge is needed (Chandola et al., 2009).

Outlier detection needs to be accurate as the capacity to detect them is usually limited. Human investigation is often necessary, which might take useful resource time. This asks for accurate outlier candidates that are interesting to the end user (Song, Wu, Jermaine, & Ranka, 2007)

For the training and validation of outlier detection models, there is a need for labelled data. The unavailability of this labelled data is often a main problem. Also the presence of noise in this data can be a factor of difficulty, because there is a great similarity between noise and actual outliers (Chandola et al., 2009; Singh & Upadhyaya, 2012).

The challenges mentioned above complicate the process of defining normal behaviour and therefore make it difficult to detect outliers. The approach varies along domains in which outlier detection is used, as it mostly is a specific formulation of the outlier detection problem. According to Chandola et al. (2009), the formulation is induced by various factors which can impact the approach used to handle the outlier detection problem.

3 Choosing an outlier detection method

To determine the best method for a certain outlier detection problem in a specific situation, there are some aspects that should be taken into consideration. According to Chandola et al. (2009), nature of the input data, which type of outliers to detect, availability of labelled data and other possible constraints or requirements are possible factors which impact the formulation of an outlier detection problem and therefore the choice of a suitable outlier detection method. The presence of these factors justifies the amount of different outlier detection techniques within various domains. It is clear that there is no single best solution to an outlier detection problem (Singh & Upadhyaya, 2012).

3.1 Factors impacting the method choice

3.1.1 Nature of input data

The input data often consists of objects, records, observations, entities, etc. In general, it is a selection of data instances which hold a set of attributes. These attributes can have a different character. The data instances can for example hold categorical, continuous, discrete, binary, or other attributes. When a data instance contains only one attribute, it is called univariate, while if it contains multiple attributes, it is called multivariate. A multivariate data instance does not necessarily contain attributes that have the same character. A mix of different types is also possible (Chandola et al., 2009).

It is important to know the nature of the input data, as this can impact the appropriateness of the outlier detection technique used. When using a model-based approach for example, a statistical or other model must be assumed and therefore these approaches are limited to certain data to which they can be applied. If a certain distribution is assumed by the model, the approach can only be applied to data with this particular distribution (Tan, Steinbach, Karpatne, & Kumar, 2005).

Proximity- or density-based approaches do not make this kind of assumption and can therefore be used for more data types. However, a certain proximity metric is used in proximity-based approaches. This metric must be chosen appropriately based on the input data. The variations in density all through a data set must also be taken into account when choosing the appropriate approach (Tan et al., 2005).

3.1.2 Outlier type

The outlier type that an outlier detection technique is trying to detect is of importance as well. Outliers are classified into three categories (Chandola et al., 2009):

Global outliers are the first category. A global or point outlier is the simplest type of outlier and it occurs when a data instance is considerably divergent from the rest of the data. Global outliers are the most common and the majority of the literature is focused on this type of outliers. They can occur in any type of data set.

A second category are contextual outliers. A contextual outlier is also a data instance that considerably diverges from the rest of the data, but only in a specific context. In a different context, the same value might not be considered an outlier. To determine this context for a certain data instance, the instance should be partitioned into two sets of attributes (Song et al., 2007).

1) Environmental or contextual attributes. These attributes describe the context for a certain data instance.
2) Indicator or behavioural attributes. These attributes are directly indicative for the non-contextual characteristics of the data instance.

The outliers are then detected by analysing the indicator data and searching for atypical values while keeping in mind the context or environmental attributes. The occurrence of contextual outliers is subject to availability of environmental or contextual attributes. Defining these environmental or contextual attributes is not always easy and straightforward and might not be meaningful in certain application domains (Chandola et al., 2009; Song et al., 2007). As time, dimensions or geographical data are mostly temporal, environmental attributes, it is therefore not surprising contextual outliers are very common and widely explored in time series data and spatial data (Kou, Lu, & Chen, 2006; Salvador, Chan, & Brodie, 2003; Shekhar, Lu, & Zhang, 2001; Weigend, Mangeas, & Srivastava, 1995)

A third category of outliers are collective outliers. These data instances are only considered outliers when a collection of related instances is considerably divergent from the entire data set. Their individual values are not necessarily outliers by themselves, but only considered outlying when occurring together (Chandola et al., 2009). An illustrative example is an employee of a company resigning. This is not considered an outlier as it is not rare that employees resign. When for example 90% of the employees resign at the same time, these data instances are considered collective outliers as it is very unusual that almost an entire company resigns at the same time, although their individual data instance are not considered outliers. Only when data instances in a data set are related to each other, collective outliers can appear. In the literature, this type of outliers have been analysed within graph data, spatial data and sequence data (Forrest, Warrender, & Pearlmutter, 1999; Shekhar et al., 2001; Sun, Chawla, & Arunasalam, 2006).

3.1.3 Data labels

All outlier detection methods can be categorised into three different types, which are supervised methods, unsupervised methods and semi-supervised methods (Hodge & Austin, 2004). The existence or availability of data labels plays a major role in selecting an appropriate method.

Supervised outlier detection needs availability of labelled instances for both normal and abnormal behaviour. Often a predictive model for both of these classes is built, so new data instances can be compared against this model (Chandola et al., 2009). A possible issue that can arise with supervised outlier detection is the unavailability of representative and accurate labels, particularly labels for abnormal behaviour. To overcome this, some approaches using

artificially generated outliers have been proposed. This is done by reducing the outlier detection problem to a classification problem and then making use of the variety of existing techniques for classification problems (Abe, Zadrozny, & Langford, 2006; Steinwart, Hush, & Scovel, 2005; Theiler & Cai, 2003). Another issue is imbalance in the class distribution as there are generally more normal than abnormal data instances. To deal with problems that can arise due to this imbalance, several approaches and techniques are presented in the literature (Chawla, Japkowicz, & Kotcz, 2004; Vilalta & Ma, 2002; Weiss & Hirsh, 1998). Some examples are boosting to improve classification accuracy (Joshi, Agarwal, & Kumar, 2002), using backpropagation together with other algorithms (Phua, Alahakoon, & Lee, 2004) and using two-phase rule induction (Joshi, Agarwal, & Kumar, 2001).

Semi-supervised outlier detection only needs data labels for normal behaviour. These methods often detect novel, previously unobserved events and subsequently determine whether the new observation lies within boundaries of the modelled normal behaviour. It can both be used for static or dynamic data. These techniques are more widely applicable as the availability of normal data labels is more common. Data labels regarding abnormal behaviour can be expensive or difficult to obtain (Chandola et al., 2009). However, there are some outlier detection techniques based solely on data labels of abnormal behaviour (D'haeseleer, Forrest, & Helman, 1996; Dasgupta & Nino, 2000). A semi-supervised outlier detection method is more likely to be able to detect and handle outliers from an unexpected, previously unseen region. On the other hand, this approach needs more normal behaviour to be modelled, as it needs the whole range of normality to be able to generalise and to be effective across different inputs and applications (Hodge & Austin, 2004; Markou & Singh, 2003a, 2003b; Singh & Upadhyaya, 2012).

Unsupervised outlier detection requires no data labels. Therefore, this is the most widely applicable technique (Chandola et al., 2009). Unsupervised outlier detection methods make some kind of assumption about the data and often handle the data as a static distribution. Remote points or points that do not seem to follow a certain pattern are identified and indicated as potential outliers. When a substantial database with good coverage is available, new items can be compared with existing data and outliers may be removed from future processing or incorporated in the distribution model to employ a robust classification method. These methods are called outlier diagnosis and outlier accommodation respectively (Hodge & Austin, 2004; Peter J. Rousseeuw & Hubert, 2011).

3.1.4 Outlier detection output

The way outliers are reported by the outlier detection technique can also be another important factor in the choice of the appropriate outlier detection method. According to Chandola et al. (2009), there are two typical outlier detection outputs:

A first one is using scores. Each instance is given a score corresponding to what extent the instance is considered an outlier. By using this method, the output of the outlier detection technique is a ranked list of outliers. For analysing purposes, one can choose to work with outliers that lie above a certain threshold. This threshold can be domain specific.

When working with the second outlier detection output, there is no possibility to specify a threshold. This second method is using labels to distinguish outliers from normal behaviour. A label stating whether a data instance is normal or abnormal will be assigned to every instance.

4 Outlier detection methods

There are copious amounts of outlier detection methods developed and reported in the literature. These methods are derived from different computing fields like statistics, machine learning and neural networks (Hodge & Austin, 2004). In the next sections, we will give a brief overview of some well-defined methods, together with their advantages and disadvantages.

4.1 Statistical methods

The first outlier detection approaches used were statistical approaches. Most of these approaches use the statistical properties of the data and fit a statistical model to this data. While using information regarding this model, a statistical inference test can be applied to determine whether or not any data instance comes from the same distribution and therefore might be an outlier (Hodge & Austin, 2004; Markou & Singh, 2003a). Statistical approaches vary from very simple, single dimensional techniques (Barnett & Lewis, 1994) to more complex approaches handling increasing dimensionality in data.

An example of a simple, single dimensional technique is Grubbs' method (Grubbs, 1969). In this method, a Z value for a query is calculated. The Z value is defined as the difference between the mean attribute value and the query value divided by the standard deviation. The mean and standard deviation are calculated based on all attribute values including the query value. Next, the Z value is compared to the critical value for a 5% or 1% significance level (Hodge & Austin, 2004). Another simple example for outlier detection is the informal box plot identification (Laurikkala, Juhola, & Kentala, 2000). Box plots typically show a five-number summary of lower extremes, lower quartile, median upper quartile and upper extremes (Markou & Singh, 2003a). In the next sections, we make a more detailed classification of several statistical approaches.

4.1.1 Parametric methods

Parametric approaches fit the data by applying a pre-selected distribution model. This allows a fast evaluation of the model for new data instances. This also makes them more scalable for large data sets in contrast to the previously mentioned methods, where adjustments or optimisations of the standard algorithms were necessary. In parametric approaches, prior knowledge of the data distribution in a data set increases the accuracy, but this also limits the applicability as data sets often fit more than one distribution model (Hodge & Austin, 2004).

Parametric methods can be Gaussian Model based, assuming a Gaussian distribution generated the data. Previously mentioned box plots and Grubbs' method can be categorised as Gaussian Model based (Grubbs, 1969; Laurikkala et al., 2000). Another common method is to use the 3σ technique. All data instances lying further than a distance of 3σ away from the distribution mean (μ) are considered outliers. σ represents the standard deviation for the underlying distribution. 99,7% of all data instances is covered by the region μ ±3σ (Chandola et al., 2009). This method is often applied in the domain of process quality control (Shewhart, 1931).

Another option is regression model-based outlier detection methods. Many variants of these methods have been proposed for both univariate and multivariate time-series data (Abraham & Chuang, 1989; Fox, 1972; P. Rousseeuw, Perrotta, Riani, & Hubert, 2019; Tsay, 2000). In basic regression model-based methods, a regression is fitted to the data. Next, the residual, which is the difference between the observed value and the predicted value of an instance, is used to calculate an outlier score (Hawkins, 1980). To manage possible outliers and reduce their influence while fitting the regression model, robust regression is used (P. J. Rousseeuw & Leroy, 1987).

It is also possible a mixture of parametric distributions is used to model the data. Normal instances can for example be generated from a certain distribution, while abnormal or outlying data instances can be generated from another distribution (Chandola et al., 2009). Normal and outlying instances can be modelled using the same type of distribution, but with other parameters such as variance (Abraham & Box, 1979). An alternative method is to model normal data with a mixture of distributions, like Gaussian mixture models.

4.1.2 Non-parametric methods

Unlike previous methods, non-parametric methods do not require a priori data knowledge, as they determine model structure from given data. Less assumptions are made on the statistical properties of the data (Chandola et al., 2009; Markou & Singh, 2003a). This makes them more autonomous and flexible compared to parametric methods (Hodge & Austin, 2004). A non-parametric method can be histogram based, where a histogram is built based on the data, or on certain attributes for multivariate data, and where outliers can be defined based on whether it lies in one of the histogram bins (Chandola et al., 2009; Eskin, 2000; Fawcett & Provost, 1999; Javitz & Valdes, 1991).

Kernel function based methods also exist (Parzen, 1962). They use the kernel functions to make for example an estimation of the normal instances probability distribution function (Desforges, Jacob, & Cooper, 1998). They are also often referred to as semi-parametric methods as they do not apply one global distribution model, but local kernel models. An advantage of these methods is the combination of the model flexibility of non-parametric methods with the speed and scalability of parametric methods (Hodge & Austin, 2004).

4.1.3 Advantages and disadvantages

Statistical techniques highly depend on their underlying data distribution assumption. This can be both an advantage as well as a disadvantage. When the distribution assumption is true, the methods offer a statistically justifiable solution which is very efficient. However, this is often not the case, especially for increasing dimensionality in data, and this limits their applicability and can increase their computational complexity. Besides this assumption of the distribution, constructing hypothesis verification tests with the right parameter values is also a complex, nontrivial task (Chandola et al., 2009; J. Zhang, 2013).

On the other hand, when a robust distribution estimation is made, statistical methods do not need labelled data and can be used in an unsupervised setting. Some of them are fairly easy

to implement, like histogram-based methods. They can handle continuously arriving data streams well and are effective when analysing a single feature. When there is a need for analysing interactions between different attributes, for example with contextual outliers, these methods do not perform well (Chandola et al., 2009; J. Zhang, 2013).

4.2 Proximity-based methods

When the proximity of a certain data point is very sparsely populated or significantly differs from the proximity of other points in the data set, this data point will be defined as an outlier by proximity-based techniques. The proximity of a certain data point is an indistinct element and therefore it is defined in different ways (Aggarwal, 2013). Proximity-based methods can be distance-based or density-based (J. Zhang, 2013).

4.2.1 Distance-based methods

Most distance-based algorithms use a distance-related metric using concepts of local neighbourhood and k nearest neighbour analysis (Ramaswamy et al., 2000; J. Zhang, 2013). A specific distance metric is used to calculate the distance. An example of such a metric is the Euclidean distance, which is given by following equation:

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

This Euclidean distance is a widely used metric for continuous attributes, but other measures can be used as well. This choice will often depend on attribute type and what kind of data it is applied to (Chandola et al., 2009; Tan et al., 2005). Distance-based outlier detection algorithms have been developed for different types of data dimensionalities and for dataset sizes ranging from small to very large (Knorr & Ng, 1998).

4.2.2 Density-based methods

Other proximity-based techniques use a density-based outlier detection method. Local density within a specified local region of a data instance is used to define outlier scores for the data instance. In general, these methods are more complex than distance-based methods because they use more complicated mechanisms for modelling the outlierliness (J. Zhang, 2013). This can for example be done by using the local outlier factor (LOF) (Breunig, Kriegel, Ng, & Sander, 2000). Even when potential outliers have a local neighbourhood density which is alike or significantly different to the neighbourhood of its neighbours, methods such as a Connectivity-based Outlier Factor scheme (COF)(Tang, Chen, Fu, & Cheung, 2002) or INFLO (Jin, Tung, Han, & Wang, 2006) still perform well.

4.2.3 Advantages and disadvantages

The main advantage of proximity-based methods is that most can work unsupervised and they do not rely on assumptions of the underlying data distribution. By changing the distance measure appropriately, distance-based methods are generally easy to adapt to and implement on different types of data. However, their effectiveness relies on the distance metric, which is

challenging to define with complex data. This makes distance-based methods not effective for high-dimensional data (Chandola et al., 2009; J. Zhang, 2013).

More effective are density-based methods. This is due to the more complicated mechanisms than distance-based methods. A drawback of this complexity is increasing computational power which often is very expensive. Another disadvantage is the non-updatability of used outlierliness measurements in most density-based methods (Chandola et al., 2009; J. Zhang, 2013).

4.3 Clustering-based methods

Clustering methods generally are unsupervised methods using clusters to group similar data instances (Tan et al., 2005). Clustering-based outlier detection methods make an assumption about when data instances are considered outliers. A first category is assuming data instances are outliers when they do not belong to any cluster (Ester, Kriegel, Jörg, & Xiaowei, 1996; Guha, Rastogi, & Shim, 2000; Yu, Sheikholeslami, & Zhang, 2002). A second category assumes outliers are data instances located too far from the closest cluster centroid (Brockett, Xia, & Derrig, 1998; Kohonen, 2001). A third category considers data instances outliers when the size of the closest cluster is too small (Eskin, Arnold, Prerau, Portnoy, & Stolfo, 2002; He, Xu, & Deng, 2003).

Clustering-based techniques are similar to distance-based techniques as they both require distance computation between instances. The difference is how the methods evaluate instances. In clustering-based methods, instances are evaluated relating to the cluster it belongs to while in distance-based methods this evaluation is done based on the local neighbourhood (Chandola et al., 2009).

4.3.1 Advantages and disadvantages

Clustering-based methods can mostly operate in unsupervised mode, but their effectiveness depends on the accuracy of the clustering algorithm in defining the clusters for normal data points. As clustering-based techniques also require distance computations, choosing the appropriate distance measure is key for a good performance of the clustering algorithm. The detection of outliers through clustering is in line with the human perception of outliers and this makes it quite intuitive (Chandola et al., 2009; J. Zhang, 2013).

Clustering algorithms exist for many and complex data types, so adapting the outlier detection method to a specific data type can often be easily done by using the appropriate clustering algorithm. However, the main goal of many clustering algorithms is not detecting outliers, and this makes them not optimised for outlier detection. When using a complex clustering algorithm, the computational complexity can increase, requiring a lot of computational power (Chandola et al., 2009; Eskin et al., 2002).

4.4 Classification based methods

Learning a model from a set of labelled data instances and then using this learnt model to classify other test instances into different classes is called classification (Tan et al., 2005). The

learning is done through training. The classifying itself is called testing. Outlier detection methods using classification also work in two phases. During training, the model learns to classify based on normal or abnormal data instances. During testing, it classifies new instances and can detect outliers based on the classification made (Chandola et al., 2009).

For training classifiers, most of the time labels are needed. There are some possibilities that do not require labels, like unsupervised neural networks. If the labels used during training belong to multiple normal classes, this is called multi-class classification. Multi-class outlier detection is able to distinguish between different normal classes and instances are considered outliers if it cannot be classified as normal. One-class classification is when the labels used during training consist of a single class. In this case, instances are considered outliers when they do not fall withing the learnt class boundary (Upadhyaya & Singh, 2012).

4.4.1 Neural networks

Neural networks traverse data sets many times in order to create a network that tries to model the data accurately. They need training and testing to be able to form a classifier for new data. They can work in both multi-class as one-class classification scenarios (Upadhyaya & Singh, 2012). In general, neural network approaches are non-parametric and model based. During training, they try to inflect the network and determine specific thresholds accordingly. This makes neural networks able to learn complicated class boundaries. Increasing dimensionality in data can be a problem for neural networks as a lot of them are vulnerable to it. However, compared to statistical techniques, neural networks overall suffer less as they often are able to focus on key attributes by reducing input features (Hodge & Austin, 2004). We can divide neural network approaches in supervised and unsupervised methods.

4.4.1.1 Supervised neural networks

The supervised methods make use of labelled data during the creation of the network. An example is the multi-layer perceptron (Augusteijn & Folkert, 2002; Bishop, 1994), which is a feedforward neural network. This means input passes through the network in one direction and no back-propagation is present. MPL are the most widely used class of neural networks (Markou & Singh, 2003b). It is used with different types of data, including time-series data (Nairac et al., 1999). When a neural network is trained using data coming from a specific set of distributions, it will be confused when data comes from another distribution (Bishop, 1994).

Other examples of supervised neural networks are radial basis function based (Bishop, 1994; Brotherton, Johnson, & Chadderdon, 1998; Nairac et al., 1999) and auto-associative neural networks (Japkowicz, Myers, & Gluck, 1995), such as Hopfield networks. Auto-associative neural networks are very accurate for outlier detection, but their training is slow, like MLP. They also do have some parameters which are specific to the data and need some empirical testing and purification to be set (Hodge & Austin, 2004).

4.4.1.2 Unsupervised neural networks

When there is no pre-labelled data available, the above supervised neural networks cannot be used for learning. In this case, unsupervised neural networks are needed. In an unsupervised

neural network, nodes compete to serve as portions of the data set. Normal and outlying classes are differentiated by modelling the underlying data distribution based on autonomous clustering of input vectors through node placement (Hodge & Austin, 2004).

An example of unsupervised neural networks are Self organising maps (Kohonen, 2001) which use vector quantisation and non-linear mapping techniques. Several extended versions of SOMs have been introduced by various authors (Marsland, 2001; Saunders & Gero, 2002; Ypma & Duin, 1998). Other examples are neural trees (Martinez, 1998), evolutionary neural network growth, like the grow when required evolutionary neural network (Marsland, 2001) or adaptive resonance theory based neural networks (Dasgupta & Nino, 2000; Moya, Koch, & Hostetler, 1993).

4.4.2 Machine learning

In machine learning methods, the focus is often on categorical data without an implicit ordering (Hodge & Austin, 2004). An example of a method used to detect outliers in categorical data is using C4.5 decision trees (John, 1995; Skalak & Rissland, 1990). This is a robust method that does not require any prior knowledge of the data. Decision trees can be used with large data sets and will operate well with increasing dimensionality in data. Similar to other classifiers, they are withal depending on the coverage of training data. Decision trees are often not very generalisable to new data instances as they can suffer from over-fitting (Hodge & Austin, 2004). To overcome this, pruning of the decision tree (John, 1995) or pre selection of normal cases (Skalak & Rissland, 1990) can be used.

Pruning is used in other methods as well, such as set-based machine learning. In this method outliers are detected based on a comparison of an instance and an already examined set of instances (Arning, Agrawal, & Raghavan, 1996). It can be used with large data sets and even without prior knowledge of the data, it will be feasible.

Other examples are rule-based systems and similarity-based systems. These methods can also operate in both multi-class and one-class classification scenarios. Rule-based systems are similar to decision trees, but they are generally more flexible as they may add or exclude rules without disturbing the process while decision trees might generate a completely new tree. In a similarity-based system example, outliers are detected by comparing a sequence to profiled sequences using a similarity measure. This system learns to classify behaviour based on past positive examples (Lane & Brodley, 1997a, 1997b).

4.4.3 Advantages and disadvantages

A big advantage of classification-based outlier detection methods is the possibility to use very powerful algorithms which can classify different data instances that belong to multiple classes. This is especially the case in multi-class classification methods. The training phase can become complex and slow when using more powerful algorithms that involve for example quadratic optimization. Because this model is learned during the training, it can be used during testing by comparing test instances against this pre-computed model. This makes the testing phase very fast (Chandola et al., 2009; Upadhyaya & Singh, 2012).

Most of the classification-based methods require accurate data labels. In the case of multi-class classification, labels for multiple classes are necessary. This is a big disadvantage as these labels are often unavailable or very expensive. The output of a classification based is most of the time a label for a data instance. When a score corresponding to what extent the instance is considered an outlier is the desired output of the outlier detection method, classification based methods are unsuitable (Chandola et al., 2009; Upadhyaya & Singh, 2012).

4.5 Hybrid systems

Hybrid systems are methods that adopt algorithms or techniques from two or more fields described above. In this way it is possible to handle occurring limitations or shortcomings of specific methods by combining and using advantages of other methods (Hodge & Austin, 2004).

The multi-layer perceptron discussed in the supervised neural networks section is not able to deduce information from new instances coming from other distributions than the ones it has been trained for. To deal with this problem it has been combined with multiple other techniques, like a Parzen window that estimates the probability density to induce a confidence estimate (Bishop, 1994). It has also been combined with a *k*-means module for partitioning and modelling graph shape normality (Nairac et al., 1999) and with hidden Markov models for output stabilisation (Smyth, 2006) or determination of certain parameters (Hollmén & Tresp, 1999).

Ensembles of several machine learning techniques and classifiers are also beginning to gain popularity. An example is the Java Agents for Meta-Learning system where multiple machine learning techniques are combined (Stolfo et al., 1997). When combining multiple classifiers, it should be kept in mind that there is a minimum of redundancy to not waste resources as complexity and processing time increases (Hodge & Austin, 2004).

5 Fraud detection

A domain in which outlier detection is important and widely used is fraud detection. When thinking of fraud, one often thinks of financial fraud, like fraudulent credit card transactions. This is because credit, banking and insurance fraud is extensively reported and explored in the literature. Other types of fraud detection, such as in the games of chance sector (Christou et al., 2011), in healthcare (Thornton, van Capelleveen, Poel, van Hillegersberg, & Mueller, 2014) or in telecommunications (Fawcett & Provost, 1997) exist as well. In this section, we will examine multiple types of fraud detection to see what kind of methods are adopted to detect fraudulent behaviour.

As fraud detection is an umbrella term for multiple types of detection in various fields, we start with defining fraud. According to the Oxford English Dictionary (1999), fraud is wrongful or criminal deception intended to result in financial or personal gain. Fraud can thus be specified as criminal activities that occur in diverse organizations or companies in both financial and non-financial sectors. The crimes are committed by malicious users which are either actual customers of the concerned organisation or ostensible customers possibly using a sort of identity theft (Singh & Upadhyaya, 2012).

A general way of detecting fraudulent behaviour is by monitoring activity (Fawcett & Provost, 1999) to maintain a usage profile that defines normal behaviour. Deviations from this normal usage pattern may be fraudulent. This is a common method in fraud detection approaches. How these deviations are detected can vary along the different domains. It often depends on the properties of the data within this domain.

5.1 Credit card fraud

Detection of credit card fraud is one of the most examined fields within fraud detection (Dal Pozzolo, Caelen, Le Borgne, Waterschoot, & Bontempi, 2014). This is due to the fact that credit card payments are gaining popularity when purchasing goods and services. Online shopping and e-commerce are booming and this results in increasing transactions, with corresponding transactional and customer data (Georgieva, Markova, & Pavlov, 2019).

Credit card fraud can occur in two ways. The first one is a fraudulent application for a credit card. This can be done using false information or by using the identity of someone else (Bhattacharyya, Jha, Tharakunnel, & Westland, 2011). To detect this type of fraud, mostly user data is used and then compared to a normal behaviour profile. When values that are not common for a particular type of user occur, this can raise alarm and escalate the specific case to an expert in the application domain. The detection of this kind of fraud is comparable to detecting insurance fraud (Ghosh & Reilly, 1994). The second way is fraudulent usage of credit cards. This is often the case when a credit card is stolen or when counterfeit credit cards are used (Georgieva et al., 2019). This type of fraud can be detected when unusual purchases occur

Credit card data is typically multi-dimensional as it keeps track of users, amounts, time between usages, locations, etc. The most occurring type of outlier in credit card fraud is a

point outlier as they often resemble unusual purchases, high amounts, high purchase rates, etc compared to the normal usage profile. Data labels are available as the companies providing the service in general have access to complete data. A problem with these data labels is the imbalance due to unavailability of many fraudulent labels. Most of the data available will be of normal transactions, which makes the non-fraudulent class significantly larger than the fraudulent one (Georgieva et al., 2019). The detection of credit card fraud can become expensive because data of all transactions needs to be stored (Singh & Upadhyaya, 2012). Output of a credit card fraud detection method can be a label assigned to an instance stating whether it is considered an outlier or not, but it may also be a score corresponding to the extent of what it is considered an outlier.

Data available for credit card fraud detection is, as described above, typically collected throughout the history of transactions. Neural networks are an effective way for handling these large volumes of customer and transactional data to recognise irregularities in the behaviour. As the datasets are mostly imbalanced, machine learning techniques are likely to generate imprecise classifiers. They often show bias towards the majority class (Georgieva et al., 2019). This needs to be kept in mind when implementing a machine learning method, together with the fact that standard classification metrics, like accuracy, are not suitable for imbalanced problems (Dal Pozzolo et al., 2014).

The transactional behaviour of customers is additionally subject to other factors. Examples are holiday seasons or special occasions. This needs to be kept in mind when detecting fraudulent behaviour. Because of this, a lot of statistical approaches are not favourable (Georgieva et al., 2019). However, clustering-based techniques are typically used for profiling of credit card users as their data falls into distinct profile clusters (Dheepa & Dhanapal, 2009; Singh & Upadhyaya, 2012). A combination of more than one method, the so-called hybrid systems, can help improving credit card fraud detection approaches because of compensation of the individual deficiencies by other methods (Krivko, 2010).

5.1.1 Examples of credit card fraud detection methods

Examples widely used by banks are rule-based checks (Brause, Langsdorf, & Hepp, 1999). The banks develop rules against which all credit card behaviour is reviewed. Examples of rules can be number of transactions in a day or amount of purchases (Ghosh & Reilly, 1994). This can be extended by using an artificial neural network and combining it with a clustering approach to compare user data, as done by Hanagandi, Dhar, and Buescher (1996). Clustering is also analysed and applied by Bolton and Hand (2001).

Because neural networks are performing well in credit card fraud detection, a lot of examples using them can be found. The neural network created by Georgieva et al. (2019) uses real historical data to make a classification of credit card transactions. This method is used in the neural network of Modi (2017) as well, together with an oversampling technique to handle the imbalance in data. Another neural classifier presented by Dorronsoro, Ginel, Sgnchez, and Cruz (1997) acts solely on the immediate previous history and the data of an operation itself.

An enormous amount of other example credit card fraud detection approaches is available, many of them combining multiple methods (Carcillo et al., 2019; Panigrahi et al., 2009). The majority of the approaches seems to be using a classification based method and neural networks are a popular choice as they handle data with the same properties as the data available in the credit card domain very well (Georgieva et al., 2019).

5.2 Insurance fraud

Fraud in insurance claims is a challenging problem for insurance companies. The traditional way of insurance fraud detection relies on costly expert inspections to detect unusual values in claims, which is often inefficient. However, roughly 21 to 36% of auto-insurance claims involve suspected fraud elements (Tennyson & Salsas-Forn, 2002). Thus, there is a need for detecting fraud to avoid economic losses for both insurance companies and policy holders (Nian et al., 2016). Another form of insurance fraud is healthcare insurance fraud. Fraudulent practitioners, large criminal networks or regular people making unintended mistakes are factors contributing to fraudulent payments by healthcare payers (Thornton et al., 2014).

As in credit card fraud detection, data regarding both auto-insurance fraud and healthcare insurance fraud is often multi-dimensional. Insurance companies and healthcare payers keep track of many attributes within their data. Outliers are not necessarily point outliers but can be very contextual. A high level of subject matter is necessary to be able to understand and adapt existing techniques to specific environments like healthcare insurance. Therefore insurance claim fraud detection methods usually generate a score corresponding to the level of outlierliness instead of producing labels stating whether an instance is normal or outlying (Nian et al., 2016; Thornton et al., 2014). Claims containing odd values or scenarios with an outlierliness score above a defined threshold could be considered outliers.

Clear fraudulent labels for training are not available and obtaining them is very costly. This makes supervised techniques unfeasible (Nian et al., 2016). However, some supervised and semi-supervised techniques exist, utilizing labels from manually investigated insurance claims (Singh & Upadhyaya, 2012). Examples are activity monitoring to find deviations from a normal usage pattern (Fawcett & Provost, 1999) and neural networks to identify fraudulent insurance claims (Brockett et al., 1998; He et al., 2003).

Unsupervised insurance fraud detection examples include multivariate clustering, the use of boxplots and detection based on a linear model. These techniques applied by van Capelleveen, Poel, Mueller, Thornton, and van Hillegersberg (2016) in the Medicaid dental domain are adapted to the complex industry of medical insurance in association with domain experts. Another example is an unsupervised model for detecting prescription fraud, which causes substantial monetary loss in health care systems. This example uses, among other techniques, a distance-based technique to generate an automated fraud detection method (Aral, Guvenir, Sabuncuoglu, & Akar, 2012).

Multiple methods have been applied for insurance fraud detection and there is no clear majority of a certain adopted method. Due to the properties of the data available in the insurance domain, there is a favour for using unsupervised methods (Nian et al., 2016).

5.3 Mobile phone fraud

Mobile phone fraud is a substantial problem for network providers and users. It happens when the telecommunications network is used in an unauthorised way through deception. Examples include unauthorised access through phone cloning, subscription fraud and phone theft. Phone cloning is done by reproducing the identification code of legitimate phones, often collected with scanners from public places. Subscription fraud can be compared to a fraudulent credit card application where a frequent occurring method is identity theft. Phone theft exists when a stolen or lost phone is used without permission by an unauthorised user (Barson, Field, Davey, McAskie, & Frank, 1996).

Mobile phone fraud is a dynamic fraud. Unauthorised users are very inventive in bypassing security measures and frauds occur on different levels in different geographical locations. Nevertheless, they cause significant costs amounting to billions of dollars of a worldwide uncollectible debt per day (Barson et al., 1996; Cox, Eick, Wills, & Brachman, 1997; Olasoji, 2014).

Telecommunications data consists of enormous databases storing detailed multi-dimensional information on users and calls. Information such as caller names, numbers, call durations, local times and destination countries is stored. Originally, telecommunication companies kept track of this data for billing purposes, but fraud detection can use this enormous data stream as well (Cox et al., 1997). The generated data volume increases significantly, so it is necessary to extract useful knowledge from it. Outliers occurring in telecommunications data are often point outliers as they correspond to extreme values of call duration, long distance to destination country, high debt/payment ratio, etc. Complete data is available to telecommunications companies, however, for clear labels determining whether a case is fraudulent or not, expertise in the telecommunications domain is necessary. This often results in using a score to formulate a degree of outlierliness in mobile phone fraud detection methods, like with insurance fraud (Barson et al., 1996; Taniguchi, Haft, Hollmen, & Tresp, 1998).

Multiple methods have been applied to telecommunications networks (Singh & Upadhyaya, 2012) ranging from neural networks (Barson et al., 1996) to density-based approaches. Taniguchi et al. (1998) present a supervised neural network, an unsupervised technique including a Bayesian network and another unsupervised density-based approach using Gaussian mixture model. The three of them are validated using real mobile communications data. The authors suggest a combination of the three methods to improve mobile phone fraud detection. Other examples include statistical profiling using histograms (Fawcett & Provost, 1999), visualization to display suspicious patterns (Cox et al., 1997), parametric statistical approaches (Aggarwal, 2005; Scott, 2000) and rule-based systems (Phua et al., 2004; Taniguchi et al., 1998).

As the methods used for mobile phone fraud detection vary among all different methods, there seems to be no favoured or better performing method for detecting this type of fraud.

5.4 Other types of fraud

The most common and most reported types of fraud are described in the previous sections. Detection of other types of fraud exists as well. We will summarize some examples in this section.

5.4.1 Games of chance fraud

A first example that is only sparsely examined is the detection of fraudulent operations in the games of chance sector. A prevalent problem in this sector is money laundering. This does not necessarily affect the immediate financial terms of organizations, but it can affect their reputation in the long run. Another problem is the threat of insider attacks, where participating agents or users try to increase their own profit by scamming systems (Christou et al., 2011).

A lot of transactional data is processed in the games of chance sector, up to fourteen million transactions in a single day. This makes it difficult to detect possible fraud as criminals try to hide their criminal intents within these large data volumes. According to Christou et al. (2011), outliers indicating possible fraud can be for example unusual values of player gross amount. This can be indicating game-fixing or money laundering. A high cancellation frequency might be an indication for agents stealing from the organization by cancelling valid tickets. Challenges like the enormous amount of transactions and user anonymity can complicate the outlier detection process.

To be able to detect possible frauds, records of certain statistics are maintained and are analysed. The authors use combinations of statistical test and cluster analysis to present a novel clustering-based outlier detection technique that works unsupervised (Christou et al., 2011).

5.4.2 Insider trading fraud

Another type of fraud that relates with the previous example is committing fraud trough insider trading in stock markets. The key here is to detect illegal profits generated by actions taken on inside information before this information is made public. This inside information is generally any form of non-public information that affects the stock prices (Singh & Upadhyaya, 2012). Early detection of insider trading fraud is important because when detected after the news becomes public for a while, investors have been disadvantaged and the fraud has been committed already (Donoho, 2004; Islam, Khaled Ghafoor, & Eberle, 2018).

Possible outliers in this domain are abnormally high trading volumes, unusual price movements and unusually distributed trading among various contract types (Donoho, 2004). Different methods have been applied to this insider trade outlier detection problem. Logistic regression, a decision tree and a neural network are used to tackle this problem. There was no method clearly outperforming the others (Donoho, 2004).

5.4.3 Tax fraud

A last example is fraud in the tax domain. The biggest problem in this domain is the unavailability of known fraud or legal cases and therefore labelled data is missing or not representative. This makes it difficult to define outliers within this domain. The fraud class distributions change over time as this domain is dynamic in nature. The notion of normal behaviour varies with sector as all sectors have different market conditions and requirements (Vanhoeyveld, Martens, & Peeters, 2019).

For this reason, outlier detection is not a common method in the tax fraud detection (Ngai, Hu, Wong, Chen, & Sun, 2011). There are some useful supervised classification methods but they are not representative for the population due to the dynamic tax domain This means fraud detection needs to be developed and assessed per sector (Vanhoeyveld et al., 2019).

An example from the value added tax (VAT) domain is given by Vanhoeyveld et al. (2019). To be able to detect fraud, they created tax ratios that can be obtained from VAT declarations. Unusual values in these ratios can signify fraud. The authors handle this problem as a contextual outlier detection problem by keeping in mind characteristics of the sector. Because of the lack of any labelled data, they use unsupervised methods. By using combinations of clustering, nearest neighbour analysis and LOF, they propose a novel approach for VAT fraud detection (Vanhoeyveld et al., 2019).

6 Conclusion

It has become clear that there is no single best outlier detection method. Authors have applied numerous methods from diverse computing fields like statistics, machine learning and neural networks. With the large amount of different methods and applications available in the literature, it is difficult to have an overview. Especially because of the use of different names for fundamentally similar approaches. We have tried to provide an overview by bringing together key information about the selection of methods, the wide variety of methods available and the possible applications in the fraud detection, but we are unable to describe all existing approaches in one paper.

When solving an outlier detection problem, the developer will encounter challenges with defining normal behaviour, distributing available capacity, and obtaining data labels if necessary. The developer should keep in mind several factors when deciding which methods are suitable for a specific problem. These factors are nature of the input data, availability of data labels, which outliers need to be detected in what context and the desired output of the method. A large amount of outlier detection methods is available including statistical, classification based, proximity-based and clustering-based methods. Based on the described factors and the strengths and weaknesses of the reviewed methods, a suitable method, or combination of suitable methods must be chosen.

In fraud detection, a commonly used method is outlier detection. We reviewed applications of outlier detection methods in the fraud detection domain to provide an overview. A common method applied in this domain is the recognition of deviations in a normal usage pattern. This is done by using different techniques from the computing fields described earlier. It appears that domain knowledge is an important necessity to be able to choose the appropriate method. A lot of different methods are used in different fraud detection domains. Neural networks are frequently occurring methods, especially in credit card fraud detection. A trend that is also visible in the fraud detection is the combination of different techniques to overcome shortcomings of others, applied to fields including telecommunication fraud, insurance fraud and fraud in the games of chance sector.

References

Abe, N., Zadrozny, B., & Langford, J. (2006). *Outlier detection by active learning*. Paper presented at the Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06, Phildelphia, PA, USA.

Abraham, B., & Box, G. E. P. (1979). Bayesian analysis of some outlier problems in time series. *Biometrika, 66*(2), 229-236. doi:10.1093/biomet/66.2.229

Abraham, B., & Chuang, A. (1989). Outlier Detection and Time Series Modeling. *Technometrics, 31*(2), 241-248. doi:10.2307/1268821

Aggarwal, C. C. (2005). On Abnormality Detection in Spuriously Populated Data Streams. In *Proceedings of the 2005 SIAM International Conference on Data Mining* (pp. 80-91): Society for Industrial and Applied Mathematics.

Aggarwal, C. C. (2013). Proximity-Based Outlier Detection. In *Outlier Analysis*: Springer, New York, NY.

Allan, J., Carbonell, J., Doddington, G., Yamron, J., & Yang, Y. (1998). Topic Detection and Tracking Pilot Study: Final Report. *Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*.

Aral, K. D., Guvenir, H. A., Sabuncuoglu, I., & Akar, A. R. (2012). A prescription fraud detection model. *Comput Methods Programs Biomed, 106*(1), 37-46. doi:10.1016/j.cmpb.2011.09.003

Arning, A., Agrawal, R., & Raghavan, P. (1996). *A linear method for deviation detection in large databases*. Paper presented at the Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon.

Augusteijn, M. F., & Folkert, B. A. (2002). Neural network classification and novelty detection. *International Journal of Remote Sensing, 23*(14), 2891-2902. doi:10.1080/01431160110055804

Barnett, V., & Lewis, T. (1994). *Outliers in Statistical Data*: John Wiley and Sons.

Barson, P., Field, S., Davey, N., McAskie, G., & Frank, R. (1996). The detection of fraud in mobile phone networks. *Neural Network World, 6*(4), 7.

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems, 50*(3), 602-613. doi:10.1016/j.dss.2010.08.008

Bishop, C. M. (1994). Novelty detection and neural network validation. *IEE Proceedings - Vision, Image and Signal Processing, 141*(4), 217-222. doi:10.1049/ip-vis:19941330

Bolton, R., & Hand, D. (2001). Unsupervised Profiling Methods for Fraud Detection. *Conference on Credit Scoring and Credit Control, 7*.

Branting, L. K., Reeder, F., Gold, J., & Champney, T. (2016, 18-21 Aug. 2016). *Graph analytics for healthcare fraud risk estimation.* Paper presented at the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

Brause, R., Langsdorf, T., & Hepp, M. (1999). *Neural Data Mining for Credit Card Fraud Detection*. Paper presented at the Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence.

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). *Lof: Identifying Density-Based Local Outliers*. Paper presented at the Proceedings of the 2000 ACM SIGMOD international conference on Management of data - SIGMOD '00.

Brockett, P. L., Xia, X., & Derrig, R. A. (1998). Using Kohonen's Self-Organizing Feature Map to uncover Automobile Bodily Injury Claims Fraud. *The Journal of Risk and Insurance, 65*(2), 30. doi:10.2307/253535

Brodley, C. E., & Friedl, M. A. (1996). *Identifying and eliminating mislabeled training instances*. Paper presented at the Proceedings of the thirteenth national conference on Artificial intelligence - Volume 1, Portland, Oregon.

Brotherton, T., Johnson, T., & Chadderdon, G. (1998, 4-9 May 1998). *Classification and novelty detection using linear models and a class dependent-elliptical basis function neural network.* Paper presented at the 1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36227).

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*. doi:https://doi.org/10.1016/j.ins.2019.05.042

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection. *ACM computing surveys, 41*(3), 1-58. doi:10.1145/1541880.1541882

Chawla, N. V., Japkowicz, N., & Kotcz, A. (2004). Editorial: Special Issue on Learning from Imbalanced Data Sets. *SIGKDD Explorations Newsletter, 6*, 6. doi:10.1145/1007730.1007733

Christou, I. T., Bakopoulos, M., Dimitriou, T., Amolochitis, E., Tsekeridou, S., & Dimitriadis, C. (2011). Detecting fraud in online games of chance and lotteries. *Expert Systems with Applications, 38*(10), 13158-13169. doi:10.1016/j.eswa.2011.04.124

Cox, K., Eick, S., Wills, G., & Brachman, R. (1997). Brief Application Description; Visual Data Mining: Recognizing Telephone Calling Fraud. *Data Mining and Knowledge Discovery, 1*. doi:10.1023/A:1009740009307

D'haeseleer, P., Forrest, S., & Helman, P. (1996). *An Immunological Approach to Change Detection: Algorithms, Analysis and Implications*. Paper presented at the IEEE Conference on Security and Privacy, Oakland, California.

Dal Pozzolo, A., Caelen, O., Le Borgne, Y.-A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Systems with Applications, 41*, 4915–4928. doi:10.1016/j.eswa.2014.02.026

Dasgupta, D., & Nino, F. (2000). *A comparison of negative and positive selection algorithms in novel pattern detection*. Paper presented at the SMC 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics. 'Cybernetics Evolving to Systems, Humans, Organizations, and their Complex Interactions' (Cat. No.00CH37166), Nashville, TN, USA.

Decoste, D., & Levine, M. (2000). Automated Event Detection in Space Instruments: A Case Study Using IPEX-2 Data and Support Vector Machines.

Desforges, M. J., Jacob, P. J., & Cooper, J. E. (1998). Applications of probability density estimation to the detection of abnormal conditions in engineering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 212*(8), 687-703. doi:10.1243/0954406981521448

Dheepa, V., & Dhanapal, R. (2009). Analysis of Credit Card Fraud Detection Methods. *SHORT PAPER International Journal of Recent Trends in Engineering, 2*.

Domingues, R., Filippone, M., Michiardi, P., & Zouaoui, J. (2018). A comparative evaluation of outlier detection algorithms: Experiments and analyses. *Pattern Recognition, 74*, 406-421. doi:10.1016/j.patcog.2017.09.037

Donoho, S. (2004). *Early detection of insider trading in option markets*. Paper presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA. https://doi.org/10.1145/1014052.1014100

Dorronsoro, J. R., Ginel, F., Sgnchez, C., & Cruz, C. S. (1997). Neural fraud detection in credit card operations. *IEEE Transactions on Neural Networks, 8*(4), 827-834. doi:10.1109/72.595879

Eskin, E. (2000). *Anomaly Detection over Noisy Data using Learned Probability Distributions*. Paper presented at the International Conference on Machine Learning.

Eskin, E., Arnold, A., Prerau, M., Portnoy, L., & Stolfo, S. (2002). A Geometric Framework for Unsupervised Anomaly Detection: Detecting Intrusions in Unlabeled Data. In *Applications of Data Mining in Computer Security* (pp. 77-101): Springer, Boston, MA.

Ester, M., Kriegel, H.-P., Jörg, S., & Xiaowei, X. (1996). *A density-based algorithm for discovering clusters in large spatial databases.* Paper presented at the KDD-96.

Fawcett, T., & Provost, F. (1997). Adaptive Fraud Detection. *Data Mining and Knowledge Discovery, 1*, 291-316. doi:10.1023/A:1009700419189

Fawcett, T., & Provost, F. (1999). *Activity monitoring: noticing interesting changes in behavior*. Paper presented at the Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, San Diego, California, USA. https://doi.org/10.1145/312129.312195

Forrest, S., Warrender, C., & Pearlmutter, B. (1999). *Detecting intrusions using system calls: Alternate data models.* Paper presented at the IEEE Symposium on Security and Privacy, Oakland, United States.

Fox, A. J. (1972). Outliers in Time Series. *Journal of the Royal Statistical Society: Series B (methodological), 34*(3), 350-363. doi:10.1111/j.2517-6161.1972.tb00912.x

Georgieva, S., Markova, M., & Pavlov, V. (2019). *Using neural network for credit card fraud detection* (Vol. 2159).

Ghosh, & Reilly. (1994, 4-7 Jan. 1994). *Credit card fraud detection with a neural-network.* Paper presented at the 1994 Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences.

Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics, 11*(1), 21. Retrieved from http://www.jstor.org/stable/1266761

Guha, S., Rastogi, R., & Shim, K. (2000). Rock: A robust clustering algorithm for categorical attributes. *Information Systems, 25*(5), 345-366. doi:10.1016/S0306-4379(00)00022-3

Gul, M., & Necati Catbas, F. (2009). Statistical pattern recognition for Structural Health Monitoring using time series modeling: Theory and experimental verifications. *Mechanical Systems and Signal Processing, 23*(7), 2192-2204. doi:https://doi.org/10.1016/j.ymssp.2009.02.013

Hanagandi, V., Dhar, A., & Buescher, K. (1996, 24-26 March 1996). *Density-based clustering and radial basis function modeling to generate credit card fraud scores.* Paper presented at the IEEE/IAFE 1996 Conference on Computational Intelligence for Financial Engineering (CIFEr).

Hauskrecht, M., Batal, I., Valko, M., Visweswaran, S., Cooper, G. F., & Clermont, G. (2013). Outlier detection for patient monitoring and alerting. *Journal of Biomedical Informatics, 46*(1), 47-55. doi:https://doi.org/10.1016/j.jbi.2012.08.004

Hawkins, D. (1980). *Identification of Outliers*. Netherlands: Springer.

He, Z., Xu, X., & Deng, S. (2003). Discovering Cluster Based Local Outliers. *Pattern Recognition Letters, 24*, 9. doi:10.1016/S0167-8655(03)00003-5

Hodge, V. J., & Austin, J. (2004). A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review, 22*.

Hollmén, J., & Tresp, V. (1999). *Call-based fraud detection in mobile communication networks using a hierarchical regime-switching model*. Paper presented at the Proceedings of the 1998 conference on Advances in neural information processing systems II.

Islam, S. R., Khaled Ghafoor, S., & Eberle, W. (2018). *Mining Illegal Insider Trading of Stocks: A Proactive Approach*. Paper presented at the 2018 IEEE International Conference on Big Data (Big Data).

Japkowicz, N., Myers, C., & Gluck, M. (1995). *A novelty detection approach to classification*. Paper presented at the Proceedings of the 14th international joint conference on Artificial intelligence - Volume 1, Montreal, Quebec, Canada.

Javitz, H. S., & Valdes, A. (1991, 20-22 May 1991). *The SRI IDES statistical anomaly detector.* Paper presented at the Proceedings. 1991 IEEE Computer Society Symposium on Research in Security and Privacy.

Jin, W., Tung, A. K. H., Han, J., & Wang, W. (2006). *Ranking outliers using symmetric neighborhood relationship*. Paper presented at the Proceedings of the 10th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining, Singapore. https://doi.org/10.1007/11731139_68

John, G. H. (1995). *Robust decision trees: removing outliers from databases*. Paper presented at the Proceedings of the First International Conference on Knowledge Discovery and Data Mining, Montréal, Québec, Canada.

Joshi, M. V., Agarwal, R. C., & Kumar, V. (2001). *Mining needle in a haystack classifying rare classes via two-phase rule induction*. Paper presented at the ACM SIGMOD international conference on Management of data, Santa Barbara, California, USA. https://doi.org/10.1145/375663.375673

Joshi, M. V., Agarwal, R. C., & Kumar, V. (2002). *Predicting Rare Classes: Can Boosting Make Any Weak Learner Strong?* Paper presented at the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada.

Knorr, E. M., & Ng, R. T. (1998). *Algorithms for Mining Distance-Based Outliers in Large Datasets*. Paper presented at the International Conference on Very Large Data Bases, San Francisco, CA, USA.

Kohonen, T. (2001). *Self-Organizing Maps* (3 ed. Vol. 30): Springer-Verlag Berlin Heidelberg.

Kou, Y., Lu, C.-T., & Chen, D. (2006). *Spatial Weighted Outlier Detection*. Paper presented at the Proceedings of the Sixth SIAM International Conference on Data Mining, Bethesda, MD, USA.

Krivko, M. (2010). A hybrid model for plastic card fraud detection systems. *Expert Systems with Applications, 37*(8), 6070-6076. doi:10.1016/j.eswa.2010.02.119

Lane, T., & Brodley, C. E. (1997a). *An application of Machine Learning to Anomaly Detection*. Paper presented at the The 20th National Information Systems Security Conference.

Lane, T., & Brodley, C. E. (1997b). Sequence Matching and Learning in Anomaly Detection for Computer Security. 7.

Laurikkala, J., Juhola, M., & Kentala, E. (2000). Informal Identification of Outliers in Medical Data. *Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology*.

Markou, M., & Singh, S. (2003a). Novelty detection: a review—part 1: statistical approaches. *Signal Processing, 83*(12), 2481-2497. doi:10.1016/j.sigpro.2003.07.018

Markou, M., & Singh, S. (2003b). Novelty detection: a review—part 2. *Signal Processing, 83*(12), 2499-2521. doi:10.1016/j.sigpro.2003.07.019
Marsland, S. (2001). *ON-LINE NOVELTY DETECTION THROUGH SELF-ORGANISATION, WITH APPLICATION TO INSPECTION ROBOTICS.* Victoria University of Wellington,
Martinez, D. (1998). Neural tree density estimation for novelty detection. *Trans. Neur. Netw., 9*(2), 330–338. doi:10.1109/72.661127
Meß, J.-G., Dannemann, F., & Greif, F. (2019). *Techniques of Artificial Intelligence for Space Applications - A Survey*. Paper presented at the European Workshop on On-Board Data Processing (OBDP2019).
Modi, K. (2017). Fraud Detection Technique in Credit Card Transactions using Convolutional Neural Network. *International Journal of Advance Research in Engineering, Science & Technology, 4*, 2394-2444.
Moya, M. M., Koch, M. W., & Hostetler, L. D. (1993). *One-class classifier networks for target recognition applications*. Retrieved from https://ui.adsabs.harvard.edu/abs/1993STIN...9324043M
Nairac, A., Townsend, N., Carr, R., King, S., Cowley, P., & Tarassenko, L. (1999). A System for the Analysis of Jet Engine Vibration Data. *Integrated Computer-Aided Engineering, 6*, 53-66. doi:10.3233/ICA-1999-6106
Ngai, E. W. T., Hu, Y., Wong, Y. H., 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. doi: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. doi:10.1016/j.jfds.2016.03.001
Olasoji, B. (2014). FRAUD DETECTION IN MOBILE TELECOMMUNICATION.
Oxford English Dictionary. (1999). *Fraud, n.*: Oxford University Press.
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), 354-363. doi:10.1016/j.inffus.2008.04.001
Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *Ann. Math. Statist., 33*(3), 1065-1076. doi:10.1214/aoms/1177704472
Phua, C., Alahakoon, D., & Lee, V. C. S. (2004). Minority Report in Fraud Detection: Classification of Skewed Data. *ACM SIGKDD Explorations Newsletter, 6*(1), 9. doi:10.1145/1007730.1007738
Pourhabibi, T., Ong, K.-L., Kam, B. H., & Boo, Y. L. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems*. doi:10.1016/j.dss.2020.113303
Ramaswamy, S., Rastogi, R., & Shim, K. (2000). Efficient Algorithms for Mining Outliers from Large Data Sets. *ACM SIGMOD Record, 29*(2), 12. doi:10.1145/335191.335437
Rousseeuw, P., Perrotta, D., Riani, M., & Hubert, M. (2019). Robust Monitoring of Time Series with Application to Fraud Detection. *Econometrics and Statistics, 9*, 108-121. doi:10.1016/j.ecosta.2018.05.001
Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1*(1), 73-79. doi:10.1002/widm.2
Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*: John Wiley & Sons.
Salvador, S., Chan, P. K., & Brodie, J. (2003). *Learning States and Rules for Time Series Anomaly Detection.* Florida Institute of Technology Melbourne,
Saunders, R., & Gero, J. (2002). Designing for Interest and Novelty - Motivating Design Agents.
Scott, S. (2000). Detecting Network Intrusion Using a Markov Modulated Nonhomogeneous Poisson Process.
Shekhar, S., Lu, C.-T., & Zhang, P. (2001). *Detecting graph-based spatial outliers: Algorithms and Applications (a summary of results).* University of Minnesota,
Shewhart, W. A. (1931). *Economic Quality Control of Manufactured Product*: D. Van Nostrand Company, New York NY.
Singh, K., & Upadhyaya, S. (2012). Outlier Detection: Applications And Techniques. *IJCSI International Journal of Computer Science Issues, 9*(1).
Skalak, D. B., & Rissland, E. L. (1990). *Inductive learning in a mixed paradigm setting*. Paper presented at the Proceedings of the eighth National conference on Artificial intelligence - Volume 2, Boston, Massachusetts.
Smyth, P. (2006). Markov monitoring with unknown states. *IEEE J.Sel. A. Commun., 12*(9), 1600–1612. doi:10.1109/49.339929

Song, X., Wu, M., Jermaine, C., & Ranka, S. (2007). Conditional Anomaly Detection. *IEEE Transactions on Knowledfe and Data Engineering, 19*, 631–645.
Steinwart, I., Hush, D., & Scovel, C. (2005). A Classification Framework for Anomaly Detection. *Journal of Machine Learning Research, 6*, 22. doi:10.5555/1046920.1058109
Stolfo, S., Prodromidis, A. L., Tselepis, S., Lee, W., Fan, D. W., & Chan, P. K. (1997). *JAM: java agents for meta-learning over distributed databases*. Paper presented at the Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, Newport Beach, CA.
Sun, P., Chawla, S., & Arunasalam, B. (2006). *Mining for outliers in sequential databases*. Paper presented at the SIAM International conference on Data Mining.
Susto, G. A., Terzi, M., & Beghi, A. (2017). Anomaly Detection Approaches for Semiconductor Manufacturing. *Procedia Manufacturing, 11*, 2018-2024. doi:https://doi.org/10.1016/j.promfg.2017.07.353
Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2005). *Introduction to Data Mining*: Pearson.
Tang, J., Chen, Z., Fu, A. W.-C., & Cheung, D. W.-L. (2002). *Enhancing Effectiveness of Outlier Detections for Low Density Patterns*. Paper presented at the Proceedings of the 6th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining.
Taniguchi, M., Haft, M., Hollmen, J., & Tresp, V. (1998, 15-15 May 1998). *Fraud detection in communication networks using neural and probabilistic methods.* Paper presented at the Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181).
Tennyson, S., & Salsas-Forn, P. (2002). Claims Auditing in Automobile Insurance: Fraud Detection and Deterrence Objectives. *Journal of Risk and Insurance, 69*, 289-308. doi:10.1111/1539-6975.00024
Theiler, J., & Cai, M. D. (2003). *Resampling Approach for Anomaly Detection in Multispectral Images*. Paper presented at the SPIE - The International Society for Optical Engineering 5093, Orlando, FL.
Thornton, D., van Capelleveen, G., Poel, M., van Hillegersberg, J., & Mueller, R. M. (2014). *Outlier-based Health Insurance Fraud Detection for U.S. Medicaid Data*. Paper presented at the Proceedings of the 16th International Conference on Enterprise Information Systems.
Torr, P. H. S., & Murray, D. W. (1997). *Outlier detection and motion segmentation*. Paper presented at the SPIE 2059.
Tsay, R. S. (2000). Outliers in multivariate time series. *Biometrika, 87*(4), 789-804. doi:10.1093/biomet/87.4.789
Upadhyaya, S., & Singh, K. (2012). Classification Based Outlier Detection Techniques. *International Journal of Computer Trends and Technology, 3*(2012), 290-294.
van Capelleveen, G., Poel, M., Mueller, R. M., Thornton, D., & van Hillegersberg, J. (2016). Outlier detection in healthcare fraud: A case study in the Medicaid dental domain. *International Journal of Accounting Information Systems, 21*, 18-31. doi:10.1016/j.accinf.2016.04.001
Vanhoeyveld, J., Martens, D., & Peeters, B. (2019). Value-added tax fraud detection with scalable anomaly detection techniques. *Applied Soft Computing*. doi:10.1016/j.asoc.2019.105895
Vilalta, R., & Ma, S. (2002). *Predicting Rare Events In temporal Domains*. Paper presented at the IEEE International Conference on Data Mining, Maebashi City, Japan.
Weigend, A. S., Mangeas, M., & Srivastava, A. N. (1995). Nonlinear gated experts for time series: discovering regimes and avoiding overfitting. *Int J Neural Syst, 6*(4), 373-399. doi:10.1142/s0129065795000251
Weiss, G. M., & Hirsh, H. (1998). *Learning to Predict Rare Events in Event Sequences*. Paper presented at the International Conference on Knowledge Discovery and Data Mining, New York, NY.
Ypma, E., & Duin, R. (1998). Novelty detection using Self-Organizing Maps. *2*.
Yu, D., Sheikholeslami, G., & Zhang, A. (2002). Findout: Finding Outliers in Very Large Datasets. *Knowledge Information Systems, 4*(4), 26. doi:10.1007/s101150200013
Zhan, Q., & Yin, H. (2018). *A loan application fraud detection method based on knowledge graph and neural network*.
Zhang, J. (2013). Advancements of Outlier Detection: A Survey. *ICST Transactions on Scalable Information Systems, 13*(1). doi:10.4108/trans.sis.2013.01-03.e2
Zhang, X., Han, Y., Xu, W., & Wang, Q. (2019). HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. *Information Sciences*. doi:10.1016/j.ins.2019.05.023