A Systematic Review of Data Mining Approaches to Credit Card Fraud Detection

IGOR MEKTEROVIĆ, LJILJANA BRKIĆ, MIRTA BARANOVIĆ
Department of Applied Computing
University of Zagreb Faculty of Electrical Engineering and Computing
Unska 3, Zagreb Croatia
CROATIA
igor.mekterovic@fer.hr, ljiljana.brkic@fer.hr, mirta.baranovic@fer.hr

Abstract: - Credit card fraud is a serious and ever-growing problem with billions of dollars lost every year due to fraudulent transactions. Fraud has always been present and will always be. It is also ever changing, as the technology and usage patterns change over time, which makes CCFD (credit card fraud detection) a particularly hard problem. Traditionally, fraud detection relied solely on domain experts' detection rules, but in the past decade or two, such solutions are being augmented with data mining models for fraud detection. The progress in this area is impeded both by the sensitive nature of the data and great commercial potential – the industrial solutions are understandably kept secret and authentic datasets are rare and few. In this paper we study the CCFD problem with its typical problems and state of the art solution. We survey the recent literature and bring a structured overview of relevant fraud detection features and data mining approaches to this problem.

Key-Words: - credit card, fraud detection, machine learning, systematic review

1 Introduction

Global electronic commerce business is in a steady rise for years: cumulative data from Statista anticipates a 246.15% increase in worldwide ecommerce sales, from $1.3 trillion in 2014 to $4.5 trillion in 2021. That's a nearly threefold lift in online revenue[1]. Of course, electronic commerce relies heavily on user adoption of credit cards which is, in turn, based on increased confidence in electronic payments and general ease of use. Put differently, credit card processing must be safe, streamlined and fast. Not surprisingly, safety and speed are two conflicting requirements – it is hard to assess transactions validity on a such large scale in a matter of milliseconds. Fraud remains an open problem requiring constant care and adjustment of detection processes. It is estimated that in USA alone the credit card fraud losses could exceed 12 billion dollars by the year 2020 [2]. Association of Certified Fraud Examiners defines fraud as "the use of one's occupation for personal enrichment through the deliberate misuse or application of the employing organization's resources or assets" [3]. Another simpler definition is "third party unauthorized use of a card". There are fundamentally two types of frauds: application (internal) and behavioural (external) frauds [4]. Application fraud represents frauds where cards are acquired from the issuing companies using false and counterfeit data. This kind of fraud is small scale and not of interest in this context. Behavioural fraud entails card lost and/or stolen scenarios, be it a physical card (e.g. through mail theft or wallet theft), or just card details. In both cases card details are obtained without the card holder knowledge and can be used in the "card holder not present" (CNP) transaction to commit fraud, typically over the Internet. In the former case, when the physical card is obtained, one can also commit "card present" frauds such as withdrawing money from the ATM or buying goods at some POS. With the rise of electronic commerce, CNP frauds are also rising and have become a dominant type of fraud. For instance, Figure 1 shows the total value of card fraud using cards issued in SEPA in 2013. by different fraud channels, showing that CNP has become an ever more important fraud channel [5].

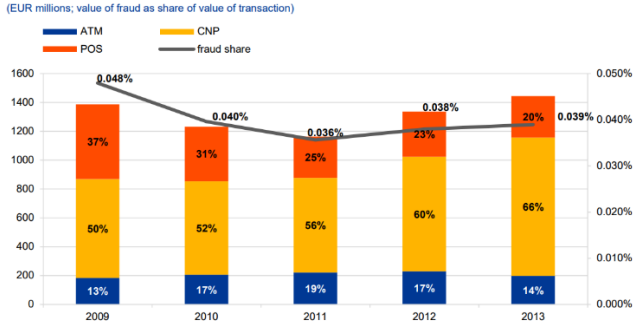

Figure 1. The total value of card fraud using cards issued in SEPA amounted to €1.44 [5]

There are several stakeholders involved in a transaction:

- Issuer: the bank issuing the credit cards and holding the customers' accounts.
- Card-holder – the customer

- Acquirer: the bank holding the merchants' accounts
- Payment processor: the entity which is responsible for linking the issuer and the acquirer every time a transaction is triggered.
- Payment schema: entity responsible for the card brands and interconnections between processors and/or banks

A more comprehensive general background and non-technical description of credit card systems can be found in [6], [7]. Various fraud detections and validations are performed at all stakeholders (except the card-holder) and can broadly be divided to fundamental card validation (e.g. checking the PIN number, checking the account balance, etc.) and more advanced fraud detection which is performed both automatically by an algorithm or "offline" - manually by a human expert. Fundamental validation is performed in real-time and it must be successful for the money to be transferred from one account to another. Near real-time is performed post-festum, but as soon as possible in order to limit the loss and break the potential chain of fraudulent transactions. Figure 2 shows the authorization and fraud detection flow. Credit card processing has to be fast, typically within a few tens of milliseconds. That is the reason that only fundamental checks are performed in real-time, as it is a major technological challenge to execute data-driven models against the stream of data in real-time.

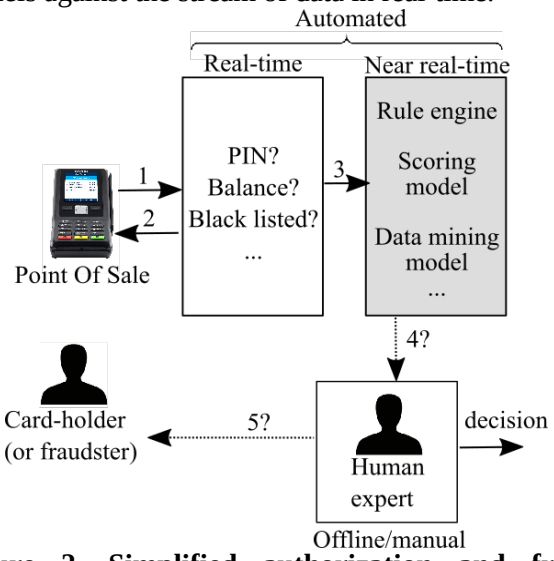

Figure 2. Simplified authorization and fraud detection flow

In this paper, we focus on automated, near real-time approaches (shaded in Figure 2). Initially, fraud detection was performed by a set of rules written by human experts, e.g. "if abroad and more than some amount then raise alert". Such rules are domain/bank specific and are hard to maintain as they grow in size and complexity and get outdated. Authors in [8] addressed this issue by proposing the system that determines the "best" adaptation to existing rules (an NP-hard problem) to capture all fraudulent transactions and, respectively, omit all legitimate transactions. The proposed modifications can then be further refined by users. On the other hand, such rules have a beautiful property of being explainable, which is very important when dealing with clients whose cards got rejected. That is also a property that most data mining models do not have, which is why rule-based fraud detection is probably still in play today to some extent.

To address the shortcomings of rule-based systems, fraud detection employed machine learning techniques. Machine learning methods can be divided into two broad categories: supervised and unsupervised. Unsupervised methods deal with outlier detection and "unusual" transactions while supervised methods classify new transactions based on samples of known data. Since banks and processors have a vast number of high quality fraud-labelled data records, supervised methods dominate the field. Firstly, neural networks were used to detect fraud, due to their immense popularity in the 90's (e.g. [9], [10]) and now again, with the advent of deep learning (e.g. [11]), but ever since, practically "all" machine learning algorithms were tested against this problem, with logistic regression ([12], [13], [14], [15]), SVM ([16], [17], [12], [13], [14]) and random forests ([12], [13], [14], [18], [17], [19]) being the most popular ones. Models are usually tweaked to accommodate fraud detection problem properties (e.g. uneven cost function or concept-drift) and various bagging/boosting methods and ensemble methods are often employed. In Chapter III we bring an overview of used methods in the past ten years. Building on "classic" rule-based systems, one might be tempted to mine association rules from the large data set ([20], [21]). This method is not suitable for fraud mining because of the highly imbalanced classes typical for this domain (fraudulent transactions are usually below 0.4%). "Moreover, since these approaches typically rely on some frequent item-set mining algorithm, they are usually designed to learn patterns for normal behaviour from which fraudulent behaviour must be extracted via outlier detection" [22]. In the rest of the paper we present the problems characteristic for CCFD and comment on the solutions, followed by a systematic review of data mining approaches to this issue.

2 CCFD Challenges

CCFD faces a number of characteristic challenges. We comment on each in the following chapters:

2.1 Lack of data

Most papers cite the lack of literature as a problem in this area. We respectfully disagree, as there is vast number of papers available on this topic as can be seen from the proposed review, and even books ([23], [24]). It is true, on the other hand, that for the confidentiality reasons a number of relevant information are obscured or omitted (e.g. aggregated variables definitions). Another often-cited problem is the lack of publicly available datasets. This is a real problem, to the extent that some researchers even used synthetic datasets, and a number of papers used one, relatively small dataset [25] containing 284,807 transactions made by credit cards in two days in September 2013 by European cardholders, released as a part of PhD thesis [26]. Furthermore, for confidentiality reasons, the meaning of most variables is not revealed in this dataset. Recently one more larger dataset was made public: in [22] authors used "synthetic dataset which is representative of real credit card transaction streams" and can be downloaded from [27]. Lack of public high-quality datasets impairs model comparisons and facilitates "personal truths". On the other hand, even with more data, this domain suffers from high volatility (things change over time and datasets become obsolete) and locality (usage patterns are different in different markets).

2.2 Feature engineering

There is a consensus that transactional data alone is not enough to drive the data mining model for CCFD. Transactional data is enriched with aggregated data from past transactions, and if available, additional data from other sources (e.g. demographic data). For instance, one such derived attribute (feature) could be the average amount spent in previous transactions in some predefined time frame. These aggregated attributes together relate to a customer (behaviour) profile and are never fully shared in the literature. It is not unusual to build several hundred such variables. The most comprehensive list we could find denotes 16 aggregated variables in [13].

2.3 Scalability

Credit card data is big data. It can also be thought of as stream data. Many transactions need to be processed in a second, and fraud estimates must be measured in milliseconds. It is challenging to design such scalable system. Most papers simply ignore this issue and deal with the fraud detection problem offline. This is not surprising as streaming/big data technologies are still emerging. A notable approach is presented in [18] where authors present a scalable framework for streaming CCFD with Kafka as the messaging queue system, Spark as the streaming and data mining platform and Cassandra as distributed database.

2.4 Unbalanced class sizes

This problem is present in every CCFD application: fraud detection is a needle in a haystack problem. The percentage of fraud is typically well under 0.4%! In other words, the baseline classifier with 99.6% accuracy is unacceptable and needs to be improved. Learning from unbalanced datasets is a difficult task since most learning systems are not prepared to cope with a large difference between the number of cases belonging to each class [28]. This problem is dealt with on either data or algorithmic level. On data level, various custom (over- and under-) sampling techniques are used, falling into following categories:

- Undersampling of non-fraud transactions, where authors usually remove at random the instances of majority class (or use e.g. stratified sampling) and experiment with various fraud to non-fraud ratios (e.g. [13]). This is the most common approach, showing good results.
- Oversampling of fraud by generating new fraud instances. For instance, SMOTE [29] generates synthetic minority data in the vicinity of the observed ones. Oversampling increases the risk of overfitting.
- Hybrid approaches: under sampling of majority and oversampling of minority, e.g. [30]
- Filtering: majority data is filtered to balance the data. For instance, in [11] authors use cascade of two filters (based on neural networks) to reduce the ratio of 5000:1 to 420:1 and then to 100:1.
- Ensemble methods – combining balancing techniques with a classifier, e.g. EasyEnsemble samples several subsets from the majority class, trains a learner using each of them, and combines the outputs of those learners [31]

At algorithmic level, modifications are made to the underlying classification algorithms. These can be divided into imbalanced learning and cost-sensitive learning. Former increase the accuracy of the minority class and latter decrease the associated cost. It should be noted that, in fraud detection, the cost is different for false positives and false negatives. For instance, in [18] the authors used Balanced Random Forest algorithm which is a modification of the classic Random Forest.

2.5 Concept Drift

Credit card fraud patterns change over time. Fraudsters agilely adopt new approaches as the old ones become obsolete. Also, the market changes, as new products and new ways of performing transactions appear. This changes the underlying data distribution, and is often referred to as "concept drift" [32]. Most papers ignore this problem and produce a one-off model based on

some data snapshot. However, based on such models one can devise various updating strategies that fall into two categories. The first approach is to use the "sliding window" to forget older transactions. The "window" can actually consist of a number of data chunks that produce a number of models that can work as a weighted ensemble. The second approach is to forget only non-fraudulent transactions and keep all fraudulent transactions (to a certain threshold). This approach has the benefit of "remembering" fraud patterns for a much longer period and, to some extent, mitigating the class unbalance problem. Papers [33] and [34] provide a nice overview of these techniques.

2.6 Performance Measures

The traditional measures of classifier performance, like accuracy and AUC (Area Under the Curve) are not well suited for this problem. Due to the data unbalance, the overall accuracy appears excellent and AUC gives equal weight to false positives and false negatives (though [33] suggests the use of AUC based on the MannWhitney statistic). False positives incur the cost of human analyst that must evaluate the transaction and maybe even contact the customer. Blocking the genuine transaction (card) carries a hard to quantify cost of customer dissatisfaction. Also, more expensive transactions are more valuable than the less expensive ones, which further complicates the issue. Usually, authors consider multiple measures, looking for high accuracy on recall (the minority class). In the following table we provide an overview of performance measures based on the confusion matrix:

Table I Credit card fraud confusion matrix

	Predicted fraud	**Predicted genuine**
Actual fraud	**TP** – true positive	**FN** – false negative
Actual genuine	**FP** – false positive	**TN –** true negative

Table II Classification performance measures

Measure	**Definition**
Sensitivity (Recall)	TP/(TP + FN)
Specificity	TN/(TN + FP)
Precision	TP/(TP + FP)
F-measure	2 * Precision * Recall / (Precision + Recall)
G-mean	$\sqrt{Sensitivity * Specificity}$
Mathew's Correlation Coefficient	$\frac{(TP * TN) - (FP * FN)}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$
Balanced classification rate	$\frac{1}{2} * (\frac{TP}{\mathrm{TP}+\mathrm{FP}} + \frac{TN}{\mathrm{TN}+\mathrm{FN}})$

Apart from *classification* perspective, it is also beneficial to consider this task from the *detection* perspective: how well are the transactions ranked and whether the model can rank fraud ahead of others? As described, potentially fraudulent transactions must be examined by human analyst and it is possible that there will not be enough analysts to evaluate all suspicious transactions in near real-time, and that is why it is of great interest that model performs well for top ranked items. E.g. one such detection performance measure, the average precision is proposed in [33]:

$AP = \sum_{r=1}^{N} (P(t_r)\, \Delta R(t_r))$

where $P(t_r)$ and $R(t_r)$ are the precision and recall of the rth ranked observation, and $\Delta R(t_r) = R(t_r) - R(t_{r-1})$. In other words, "An algorithm A is superior to an algorithm B only if it detects the frauds before algorithm B" [33]. Performance measures can also be weighted where transaction amounts are used as weights. At the end of the day, measure selection is primarily a business decision.

3 Literature review

For this purpose, in July 2018 we explored Web of Science (WoS) database and focused on peer-reviewed articles in journals and conference publications published in last 10 years (between 2009 and 2018) in the 'Computer science artificial intelligence' category. We searched for keywords 'credit card fraud detection' in the title, abstract and in the keywords of the articles and received 86 hits. Our search strategy is presented in Table III.

Table III WOS search strategy

Database	**Search Strategy**
Web of Science (WoS)	TOPIC: (credit card fraud detection) Refined by: WEB OF SCIENCE CATEGORIES: (COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE) Timespan: 2009-2018. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

The 'Computer science artificial intelligence' category was chosen for two reasons: (i) fraud detection fits naturally to this category; (ii) to reduce the number of papers since the number of papers regardless of the category rises to 205. The features of the fraud detection systems we wanted to explore are stated and described in Table IV. Out of 86 papers 60 were omitted from further consideration, because they were inaccessible due to lack of access rights (5 papers) or not relevant for this research (55 papers). The main reasons for classifying articles in the 'not relevant'

category are: (i) the searched terms are present in the abstract only and not studied in the remainder of the paper (15 papers); (ii) the paper deals with specific problems like outliers, sampling and sequence classification, important and relevant topics, but not a subject of our research (18 papers); (iii) the proposed solution does not employ any machine learning algorithm or experimental dataset characteristics are not presented (18 papers); (iv) the paper presents comparative study of the scientific work in the fraud detection field (4 papers). Some papers were relevant but due to confidentiality reasons, authors did not present any detail about dataset used. Such papers are also omitted.

Table IV Investigated features of fraud detection systems

Feature abreviation	Description
Year	Year of publication
DSType	The type of dataset used. Two types of datasets are identified: real-world (RL) and synthetic (Syn).
DSSize	The size of original dataset (when stated, abbreviated Or) and/or the size of dataset used in experiments (abbreviated Ex).
Fraud%	The percentage of fraudulent transactions in the original or the experimental data.
FType	The type of transactions features/attributes used as model input. Two types are identified: numerical only and mixed. Mixed type includes numerical and categorical data.
FNo	Number of features/attributes.
Sampling	Sampling method used: US =Under-sampling, NFT= non-fraudulent transactions; OS FT=Over-sampling fraudulent transactions.
Machine learning algorithm	Machine learning algorithm used.

The remaining 26 papers were carefully reviewed and based on this, Table V and Table VI are populated (NS in both tables is used for Not specified). Table V presents an overview of features we consider relevant in for this topic while Table VI gives a survey of applied machine learning algorithms in the surveyed WOS papers. In almost all researches real-life data sets were used (DSType column), typically in cooperation with the industrial partner, and never publicly available. This is a typical problem described in Chapter II.A. The size of the dataset (DSSize) used to obtain the experimental dataset or in experiments themselves is very varied and ranges from several thousand to several hundred million. An even bigger difference is present in the expressed percentage of fraudulent transactions (Fraud%). The typical fraudulent transaction percentage is below 0.4% in the original dataset and is typically oversampled to train the model. Not all papers state both percentages, so the table conveys only what is stated. Regarding the characteristics and number of features (FType, FNo) used in DM models, in addition to numerical and categorical transaction data, aggregated attributes, reflecting customer profile, are generally used. The number of features used varies from 3 ([42]) to 293 ([16]). It is unlikely that only 3 features will reflect the customer profile and be sufficient for model construction resulting in reliable transaction fraud detection.

To overcome the problem of an unbalanced dataset and avoid subsequent problems with the ML algorithm application, almost all researchers used some kind of sampling technique (Sampling). The prevailing sampling techniques are under-sampling non-fraudulent transactions and over-sampling fraudulent transactions.

Table V Overview of fraud detection features

Paper	Year	DS Type	DS Size	Fraud%	FType	F No	Sampling
[36]	2009	RL	Ex:30876	< 0.1	Mixed	41	US NFT
[12]	2009	RL	OrA:175*10^6 OrB:1.1*10^6	NS	Mixed Mixed	45 47	Yes to fit 0.1% fraud
[21]	2009	NS	12,107	NS	Mixed	4	NS
[37]	2011	RL	Or:60*10^6 Ex:1,1*10^6	0.00007 0.07	NS	NS	US NFT
[15]	2012	RL	Or:50*10^6 Ex:15099	 0.16	Mixed	28	OS FT
[30]	2013	RL	Or:22*10^6	0.004	Mixed	NS	US NFT
[39]	2013	RL	Ex:1000	30	NS	20	
[40]	2013	RL	Or:80*10 Ex:750,000	0.025 0.467	Mixed	274	OS FT
[41]	2013		Or:50*10^6	0.0044	NS	15	US NFT
[42]	2014	RL	Or:10000 accounts *2 months	NS	Mixed	3	NS
[33]	2014	RL	Or:422 days	0.4	Mixed	NS	
[43]	2014	RL	Or:41647	3.74	NS	18	NS
[44]	2015	RL	Or:3.3*10^6	<1	Mixed	28	US NFT
[45]	2015	RL	Or:120*10^6 Ex:236,735	0.025 1.5	Mixed	277	US NFT, OS FT
[46]	2015	RL	9387	10	NS	102	No
[47]	2015	RL	21.8*10^6 7.6*10^6	0.19 0.22	NS	51 51	NS
[16]	2016	RL	Or:120*10^6	0.025	Mixed	293	US NFT
[48]	2016	RL	41647	3.74	Mixed	18	Cost sens. appr.
[49]	2016	RL	Or:1,1*10^6	0.07	Mixed	18	OS FT
[50]	2016	RL	9388 5960 3463	0.11 0.2	Mixed	27 20 14	NS
[51]	2017	RL	Or:450000	NS	Mixed	71	US NFT
[52]	2017	RL	Or:152,706	0.0065	Mixed	6	NS
[11]	2018	RL	Or:900*10^6	0.02	Mixed	62	NN filter
[18]	2018	RL	Or:8*10^6	0.4	Mixed	18	US NFT
[19]	2018	RL	Or:2.9*10^6 Or:4.3*10^6	1.48 0.81	Mixed	9	US NFT
[53]	2018	RL	Or:3*10^6	NS	Mixed	9	NS

			Ex:277,721				

The list of applied ML algorithms is quite long (Table VI), and the most frequently used are: neural networks, random forest, logistic regression, support vector machines, decision tree and naïve Bayes. These methods are systematically used in papers published at the beginning of the observed ten-year period as well as in the recently published papers while some novel methods (e.g. deep learning) appear only in recent publications. Some authors tested several ML algorithms while trying to solve the problem so those papers' references appear more than once (e.g. [37], [19]).

Table VI Overview of machine learning methods used

Machine learning algorithm	**Papers**	**Count**
Neural networks	[36] [37] [42] [33] [44] [46] [48] [50] [19]	9
Random forest	[37] [33] [44] [45] [47] [51] [52] [18] [19]	9
Logistic regression	[37] [15] [44] [45] [16] [49] [51]	7
Support vector machines	[36] [37] [39] [33] [16] [51] [52]	7
Decision tree	[12] [45] [46] [16] [50] [52]	6
Naïve Bayes	[37] [46] [50] [52]	4
Artificial immune system	[43]	1
Association rules	[21]	1
Bayes minimum risk	[40]	1
Bayesian network	[52]	1
Cost-based DT	[30]	1
Hidden Markov model	[53]	1
k Nearest neighbours	[12]	1
Deep learning	[18]	1
Linear discriminant analysis	[46]	1
Migrating birds optimization	[41]	1
Quadratic discriminant	[12]	1

4 Conclusion

CCFD is a difficult problem mainly due to the class imbalance, concept drift and complicated cost-structure. It is also a very closed area where industrial solutions are understandably kept secret. In this paper we have performed a systematic review on this subject and detected the characteristic problems and proposed state-of-the-art solutions and surveyed a number of data mining approaches to this problem. It is evident that there are many approaches to this problem which are sadly hard to compare, but it is also notable that some appear more than others (e.g. random forest). CCFD is also complicated from the technological standpoint, as credit card transactions are a stream of data that needs to be processed in near real-time. This area is in the constant flux and it is unlikely to be "solved" in the near future. It will probably remain an everlasting race with the fraudsters as the market and technology changes

References:

[1] A. Orendorff, "No Title," *Global Ecommerce: Statistics and International Growth Trends*. [Online]. Available: https://www.shopify.com/enterprise/global-ecommerce-statistics.

[2] R. Aitken, "No Title," *U.S. Card Fraud Losses Could Exceed 12B USD By 2020*, 2016. [Online]. Available: http://www.forbes.com/sites/rogeraitken/2016/10/26/us-card-fraud-losses-could-exceed-12bn-by-2020/.

[3] A. of C. F. Examiners, "No Title." [Online]. Available: http://www.acfe.com/rttn-introduction.aspx.

[4] R. J. Bolton, D. J. Hand, and D. J. H, "Unsupervised Profiling Methods for Fraud Detection," *Proc. Credit Scoring Credit Control VII*, pp. 5–7, 2001.

[5] E. C. Bank, "Fourth report on card fraud," 2015.

[6] D. J. Hand and G. Blunt, "Prospecting for gems in credit card data," *IMA J. Manag. Math.*, vol. 12, no. 2, pp. 173–200, 2001.

[7] V. Hanagandi, A. Dhar, and K. Buescher, "Density-based clustering and radial basis function modeling to generate credit card fraud scores," in *Computational Intelligence for Financial Engineering, 1996., Proceedings of the IEEE/IAFE 1996 Conference on*, 1996, pp. 247–251.

[8] T. Milo and W. Tan, "Interactive Rule Refinement for Fraud Detection."

[9] J. R. Dorronsoro, F. Ginel, C. Sánchez, and C. Santa Cruz, "Neural fraud detection in credit card operations," *IEEE Trans. Neural Networks*, vol. 8, no. 4, pp. 827–834, 1997.

[10] Ghosh and Reilly, "Credit card fraud detection with a neural-network," *1994 Proc. Twenty-Seventh Hawaii Int. Conf. Syst. Sci.*, vol. 3, pp. 621–630, 1994.

[11] J. A. Gómez, J. Arévalo, R. Paredes, and J. Nin, "End-to-end neural network architecture for fraud scoring in card payments," *Pattern Recognit. Lett.*, vol. 105, pp. 175–181, 2018.

[12] C. Whitrow, D. J. Hand, P. Juszczak, D. Weston, and N. M. Adams, "Transaction aggregation as a strategy for credit card fraud detection," *Data Min. Knowl. Discov.*, vol. 18, no. 1, pp. 30–55, 2009.

[13] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decis. Support Syst.*, vol. 50, no. 3, pp. 602–613, 2011.

[14] P. Ravisankar, V. Ravi, G. Raghava Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining techniques," *Decis. Support Syst.*, vol. 50, no. 2, pp. 491–500, 2011.

[15] S. Jha, M. Guillen, and J. Christopher Westland, "Employing transaction aggregation strategy to detect credit card fraud," *Expert Syst. Appl.*, vol. 39, no. 16, pp. 12650–12657, 2012.

[16] A. Correa Bahnsen, D. Aouada, A. Stojanovic, and B.

Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Syst. Appl.*, vol. 51, pp. 134–142, 2016.

[17] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," *IEEE Access*, vol. 6, pp. 14277–14284, 2018.

[18] F. Carcillo, A. Dal Pozzolo, Y. A. Le Borgne, O. Caelen, Y. Mazzer, and G. Bontempi, "SCARFF: A scalable framework for streaming credit card fraud detection with spark," *Inf. Fusion*, vol. 41, pp. 182–194, 2018.

[19] J. Jurgovsky *et al.*, "Sequence classification for credit-card fraud detection," *Expert Syst. Appl.*, vol. 100, pp. 234–245, 2018.

[20] J. Xu, A. H. Sung, and Q. Liu, "Behaviour mining for fraud detection," *J. Res. Pract. Inf. Technol.*, vol. 39, no. 1, pp. 3–18, 2007.

[21] D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, "Association rules applied to credit card fraud detection," *Expert Syst. Appl.*, vol. 36, no. 2 PART 2, pp. 3630–3640, 2009.

[22] A. Artikis *et al.*, "Industry paper: A prototype for credit card fraud management," *DEBS 2017 - Proc. 11th ACM Int. Conf. Distrib. Event-Based Syst.*, 2017.

[23] L. W. Vona, *Fraud Data Analytics Methodology: The Fraud Scenario Approach to Uncovering Fraud in Core Business Systems*. John Wiley & Sons, 2017.

[24] W. V. Bart Baesens, Veronique Van Vlasselaer, *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection*. John Wiley & Sons, 2015.

[25] M. L. G.- ULB, "Credit Card Fraud Detection Anonymized credit card transactions labeled as fraudulent or genuine." [Online]. Available: https://www.kaggle.com/mlg-ulb/creditcardfraud.

[26] A. Dal Pozzolo, "Adaptive Machine Learning for Credit Card Fraud Detection Declaration of Authorship," no. December, p. 199, 2015.

[27] S. Project, "No Title." [Online]. Available: http://speedd-project.eu/data.

[28] G. E. A. P. A. Batista, A. C. P. L. F. Carvalho, and M. C. Monard, "Applying one-sided selection to unbalanced datasets," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 1793 LNAI, pp. 315–325, 2000.

[29] N. Chawla and K. Bowyer, "SMOTE: Synthetic Minority Over-sampling Technique Nitesh," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[30] Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," *Expert Syst. Appl.*, vol. 40, no. 15, pp. 5916–5923, 2013.

[31] X. Y. Liu, J. Wu, and Z. H. Zhou, "Exploratory under-sampling for class-imbalance learning," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 965–969, 2006.

[32] T. R. Hoens, R. Polikar, and N. V. Chawla, "Learning from streaming data with concept drift and imbalance: an overview," *Prog. Artif. Intell.*, vol. 1, no. 1, pp. 89–101, 2012.

[33] A. Dal Pozzolo, O. Caelen, Y. A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4915–4928, 2014.

[34] T. R. Hoens, R. Polikar, and N. V. Chawla, "Learning from streaming data with concept drift and imbalance: an overview," *Prog. Artif. Intell.*, vol. 1, no. 1, pp. 89–101, 2012.

[35] S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning," *Inf. Fusion*, vol. 10, no. 4, pp. 354–363, 2009.

[36] B. Wiese and C. Omlin, "Credit card transactions, fraud detection, and machine learning: Modelling time with LSTM recurrent neural networks," *Stud. Comput. Intell.*, vol. 247, pp. 231–268, 2009.

[37] N. F. Ryman-Tubb and P. Krause, "Neural network rule extraction to detect credit card fraud," *IFIP Adv. Inf. Commun. Technol.*, vol. 363 AICT, no. PART 1, pp. 101–110, 2011.

[38] N. Wong, P. Ray, G. Stephens, and L. Lewis, "Artificial immune systems for the detection of credit card fraud: An architecture, prototype and preliminary results," *Inf. Syst. J.*, vol. 22, no. 1, pp. 53–76, 2012.

[39] M. Hejazi and Y. P. Singh, "One-class support vector machines approach to anomaly detection," *Appl. Artif. Intell.*, vol. 27, no. 5, pp. 351–366, 2013.

[40] A. C. Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten, "Cost sensitive credit card fraud detection using bayes minimum risk," *Proc. - 2013 12th Int. Conf. Mach. Learn. Appl. ICMLA 2013*, vol. 1, pp. 333–338, 2013.

[41] E. Duman and I. Elikucuk, "Solving Credit Card Fraud Detection Problem by the New Metaheuristics Migrating Birds Optimization," pp. 62–71, 2013.

[42] D. Olszewski, "Fraud detection using self-organizing map visualizing the user profiles," *Knowledge-Based Syst.*, vol. 70, pp. 324–334, 2014.

[43] N. Soltani Halvaiee and M. K. Akbari, "A novel model for credit card fraud detection using Artificial Immune Systems," *Appl. Soft Comput. J.*, vol. 24, pp. 40–49, 2014.

[44] V. Van Vlasselaer *et al.*, "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions," *Decis. Support Syst.*, vol. 75, pp. 38–48, 2015.

[45] A. C. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive decision trees," *Expert Syst. Appl.*, vol. 42, no. 19, pp. 6609–6619, 2015.

[46] N. Mahmoudi and E. Duman, "Detecting credit card fraud by Modified Fisher Discriminant Analysis," *Expert Syst. Appl.*, vol. 42, no. 5, pp. 2510–2516, 2015.

[47] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection and concept-drift adaptation with delayed supervised information," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2015–Septe, 2015.

[48] F. Ghobadi and M. Rohani, "Cost sensitive modeling of credit card fraud using neural network strategy," *Proc. - 2016 2nd Int. Conf. Signal Process. Intell. Syst. ICSPIS 2016*, pp. 8–10, 2017.

[49] T. Liu and S. Liu, "Fraud detection model & application for credit card acquiring business based on data mining technology," vol. 50, no. Iceeecs, pp. 963–967, 2016.

[50] A. Zakaryazad and E. Duman, "A profit-driven Artificial Neural Network (ANN) with applications to fraud detection and direct marketing," *Neurocomputing*, vol. 175, no. PartA,

pp. 121–131, 2016.

[51] N. Carneiro, G. Figueira, and M. Costa, “A data mining based system for credit-card fraud detection in e-tail,” *Decis. Support Syst.*, vol. 95, pp. 91–101, 2017.

[52] Y. Kültür and M. U. Çağlayan, “Hybrid approaches for detecting credit card fraud,” *Expert Syst.*, vol. 34, no. 2, pp. 1–13, 2017.

[53] W. N. Robinson and A. Aria, “Sequential fraud detection for prepaid cards using hidden Markov model divergence,” *Expert Syst. Appl.*, vol. 91, pp. 235–251, 2018.