Chapter 20
Enhancing Data Quality by Detecting and Repairing Inconsistencies in Big Data

Vinaya V. Keskar, Jyoti Yadav, and Ajay Kumar

Abstract In several industries in established countries, a phase of Big Data examination has begun. Big data entails huge amounts of data that are challenging to analyze using standard database and software approaches. When using big data applications, a technological hurdle arises when transferring data across several locations, which is highly expensive and necessitates a huge primary memory for storing data for computation. Big data refers to the transaction and interaction of datasets whose size and complexity transcend the ordinary technical capabilities of gathering, organizing, and processing data in a cloud environment. Expanding data metrics are overflowing into contemporary associations with growing developments in Internet technology. Because of a relentless era of data, data from various devices and channels, such as cellular phones, PCs, government documents, medical reports and web-based media, are increasingly misunderstood. In this article, we explored the anomalies in the banking sector attributable to big data technologies, credit card discrepancies and the manner in which the toolkit is used to assess the incongruity of a specific WAP (Wireless Application Protocol) instrument.

20.1 Introduction

Information are different kinds of data that are usually produced in a particular way. The product is divided into two main categories, namely projects and info. Projects consist of an assortment of data control instructions [1]. In this respect, let us know some fabulous truths, actually after a full comprehension of data and data science.

V. V. Keskar (✉)
ATSS College of Business Studies and Computer Applications, Savitribai Phule Pune University, Pune, India
e-mail: vasanti.keskar@gmail.com

J. Yadav
Department of Computer Science, Savitribai Phule Pune University, Pune 411007, India

A. Kumar
JSPM Jaywant Technical Campus, Pune, India

S. Yadav et al. (eds.), *Proceedings of Second International Conference in Mechanical and Energy Technology*, Smart Innovation, Systems and Technologies 290,
https://doi.org/10.1007/978-981-19-0108-9_20

In the concept of data, the word Big Data is used to describe data in or larger petabyte sizes. The term “big data” applies to the creation and usage of technologies that provide the ideal consumer with the right knowledge from a mass of data which has developed rapidly throughout our general population for some time [2]. In addition to handling the complexity of tracking gradually heterogeneous organizations, Big Data has now become a 5Vs spectrum, number, meaning, truthfulness and tempo, and it is not just the challenge to handle steadily expanding data volumes. In the past, Internet companies have started to spread, Big Data action strategies have evolved and data was viewed as a resource. There are also various benefits of big data such as cost savings, increased performance and enhanced dealing; thus, above and beyond data collection in computer systems, there is growing value of data. As far as data science is concerned, data science is a community consisting of realities [3]. In addition, economical, demographic, welfare and exhibition all have different consequences of details, which eventually provides different responses to results.

20.1.1 5 V’s of Big Data

Volume: refers to the volume of data generated inside the frame, processed and run. The increase in the amount of data produced and processed is explained by the in-crease in it, but additionally by the need to use it [4, 5].

Variety: Refers to increasing data types managed by a system for knowledge. This replication allows relations and connections between the data to be multi-faceted. The range often describes the possible applications of a raw data.

Speed: Refers to the recurrence of data produced, processed and transmitted. The data exists by stream and should be continually broken down.

Value: There is one more V that reflects Value! After taking the 4 Vs into account. The mass of data without meaning is just terrible for the enterprise, even if you find it useful. Data itself is of little value or relevance except that it must be translated into something significant in order to distinguish knowledge.

Veracity: It applies to incoherences or vulnerability of data, which ensures data can be accessed often because it is challenging to monitor consistency and accuracy.

20.1.2 Applications of Big Data

With regard to data of the board, each application field has its characteristics and necessities. For the factors underpinning this analysis, further specifics have been selected in the most tested and renowned fields of application: enterprise and manufacturing, medical, legislative and public services, education and a rational study [6].

Data is generated into organized (in the relevant arranging as in databases), unstructured (multimedia, text) and semi-structured (XML archives) systems from various channels such as the social network, e-commerce, healthcare, etc. Utter first step is to resolve these data in order to clean them up.

Continuous Data Analytics: Real-time systems are done with instantaneous knowledge streams which involve accelerated processing in an extremely limited amount of time as judgments are taken when the time line is finished. When large data are converted, the investigator can in a few moments call for their findings from gigantic datasets [7]. Data need to be processed super rapidly in order to react constantly to evolving conditions. Email Analytics: the website text data to Exabyte has been accessed. More than 7 million pages are generated daily. This detail is broken down under the text analysis umbrella. AI, mathematical analysis and machine etymology are the texts of the inquiry. The text overview produces a retrospective that moves core features from individual or different archives. There have been two approaches that have been more evolved in the round: extractive and abstract methodologies [8].

Analytics of Multimedia: multimedia material contains sound, video and images. Multimedia content analysis extracts and acquires interesting knowledge. Multimedia synopsis, multimedia commentaries, multimedia ordinations and retrieval are the key subjects listed for exploring multimedia science.

Spatio-Textual Analytics: Spatio-textual comparability retrieves a sequence of publications in which items in each pair are spatially similar, as are textual measurements. Great spatial textual data, which allow for new strategies for questioning and performing the procedure on this new data form, is created because of the pervasiveness of GPS-enablement devices. The web actually obtains spatial calculation alongside its earlier textual measurements with the expansion of GPS and Internet creativity. Currently, social web applications like Foursquare, Flicker are generating a hundred space textual data and offer the opportunity to use this data in different applications. However, it also demands that processing methods be enhanced, so that the space-text data are tracked and applied through separate activities [9].

20.1.3 Credit Card Related Problem in Banking Sector

The three types of credit-card fraud can be completely categorized into: card-related fraud, trader-related fraud and web-related fraud. The below are different forms of payment card fraud:

Amount Takeover: This kind of misrepresentation arises if a fraudster improperly collects all classified person details. He/she illuminates the bank hanging on his private or company address, which is imitated as the real cardholder. Next he/she announced that his credit card had been misplaced and that his current address was being demanded to mail another card. He/she is issued the card and the suspect will carry over the record easily in that way.

Fake Card Theft (also referred to as Skimming): A fake, cloned and skimmed card has, or has been legally produced and then changed or reported, been printed or embellished, or encoded without approval from the card organization. False misrepresentation also entails skimming, a loop where the desirable strip of the cards is repeated on another card through electronic machines, without the details on the real cardholder. Retail sources—especially bars, cafés and gas stations [10]—can be used to skim.

Postal Fraud: Which arises in situations when a thief catches a replacement card sent by and uses a genuine card holder.

Fraud: Anytime someone unlawfully gains knowledge and frequently uses it for the intent of making new documents or beginning an exchange for the profit of an authentic user. Most burglaries of character occur disconnected like pockets, mail blockage, or the Junk Browsing.

20.2 Literature Review

Sk. Sk. Sk. Credit card fraud was one example of Ref. [11] the banking and financial sectors are experiencing significant problems in the form of cyber fraud. We used a one-class classification technique in big data paradigm to identify credit card fraud. In this article, we also introduced a hybrid architecture for the Spark Spark theoretical paradigm for Particle Swarm Optimization and the Self-Associative Neural Network, as indicated elsewhere.

Fraud has no clear trend [12]. You still alter your behavior; so we have to learn un-attended. Fraudsters learn about emerging tools to enact deception by Internet purchases. Fraudsters. The normal activity of customers is suspected by fraudsters and fraud habits easily shift. Fraud identification mechanisms therefore have to monitor Internet purchases using unattended learning, and certain fraudsters have once perpetrated fraud via online media and then moved to other strategies. This paper aims at (1) designing a model of the deep auto-encoder and a restricted Boltsmann system (RBM). This model is capable of reconstructing regular business transactions to identify irregularities in normal patterns. The proposed auto-encoder-deep learning (AE) is an unattended learning algorithm that applies context propagation by setting the inputs to fit the outputs. The RBM is split into two layers, the input (visible) and the unseen. We use the Google Library from Tensorflow to carry out AE, RBM and H_2O by way of deep learning in this investigation. The findings display the average square error, root mean square error and curve field.

Researcher [13] Credit card use has significantly increased because of rapid credit card growth. As a consequence, credit card theft and defaults have greatly risen on credit card owners and credit card firms. Credit card analysis was commonly used to diagnose irregularities in credit card purchases on the premise that a fraud trend will rely on the previous transaction. Unattended learning should not, though, neglect the likelihood of the fraudsters modifying their tactics depending on consumers' behavior. Three unsupervised approaches, including an auto-encoder, a one-class vector support machine and robust outlier identification, were presented in this review. The data collection used in this analysis is focused on actual credit card purchase data. After training the models, the success of each model was assessed due to the availability of the answer, fraud labels. In the manuscript we examined the success of these three approaches in depth. Standard transaction labels were used for training in one-class SVM and auto-encoders. The benefits of the Mahalanobis system relative to these techniques, however, are that there is no requirement for a training sticker.

There are a broad variety [14] of uses in the use of graphs to collect and present de-tails. These applications could be present in the identification of semanticized and systemic patters and graphs have been increasingly increasing for such applications. In this article we will present one of the most harmful credit card anomalies focused on such a definition. The pace of usage of credit cards has dramatically escalated with advancing banking technologies. In this region, the incidence of fraud has risen, and we are modeled on graphs to resolve such anomalies. The major benefit of the strategy is that while running simulations the device overload rate is decreased in order to detect fraud and thus the detection level is accelerated.

The exponential development [15] in the area of e-commerce and the boom in e-payment has made it extremely necessary for the Fintech sector to recognize online transaction fraud in real time. In order to resolve this problem, we present the TitAnt, a framework for detecting transactions fraud in Ant Financial, which is one of the world's largest fintech firms. The machine will detect fraud in only milliseconds in real-time online transactions. We present the issue description, extraction of functions, detection methods, device installation and deployment and analytical performance. Significant real-world transaction results have been used to illustrate the reliability and efficacy of the proposed method.

The credit card [16] has been potentially the most prevalent way of payment, on daily and on-line orders, and thus there are major rises in fraud connected with those transactions. Fraudulent transacting credit cards per year cost companies and customers significant financial costs, and fraudsters are actively finding innovative technology and ways to participate in fraudulent transactions. The prevention of fraudulent purchases is a big factor in the increased usage of electronic payments. Effective and reliable methods in credit card purchases are also important to identify fraud. This paper suggests a clever solution to fraud prevention utilizing a light gradient boosting machine in credit card transactions (OLightGBM). A Bayesian optimization algorithm is smartly combined with the proposed solution in order to change the parameters of a light gradient boosting machine (LightGBM). Experiments were conducted utilizing two real-world public credit card data sets consisting of fake transactions and legit sets to show that our proposed OLightGBM was successful to identify fraud in credit card transactions. The suggested method, focused on a compare to other methods utilizing the two data sets, outpaced the other techniques and obtained the most exact (98.40%) precision, recepteur-operating curve (AUC) region (92.88%), accuracy (97.34%) and F1 score (56.95%).

Irregularities in data identification [17] is a critical activity and is subject to many high-effect applications in areas including defense, economics, health and law enforcement. In recent years, various methods have been established to spot outliers and anomalies in unstructured arrays of multi-dimensional objects with graph data being omnipresent. Because artifacts are correlated over a long distance, a series of modern technologies for the identification of anomalies in graph data has been created. This survey is aimed at offering a summary of state-of-the-art anomaly detecting methods in figurative data, which is common, systematic and standardized.

A primary problem [18] confronted by major economic entities was the detection of fraud arising from a spike in expenditure for credit cards. This paper proposes a novel approach to the estimation of payment through credit card fraud that relies on isolation forest and local external variables. The approach proposed involves the corresponding phases: data set pre-processing, preparation and sorting, judgment con-verging and test review. In this document, two forms of algorithms are used as a teaching to demonstrate the conduct characteristics of the right and wrong transactions. To date, many scientists have established numerous methods to detecting and increasing these frauds. In this article, we propose an overview of python and its de-tailed experimental performance in isolation forest and local external factor algorithms. When analyzing the dataset, the Local Outlier Factor Algorithm demonstrated strong accuracy of the Isolation Tree.

The collection of characters [19] is considered very necessary to enhance the classification and identification method in order to recognize the credit card danger in large or high-dimensional details. Random Forest Classifier (RFC) is one of the most frequently employed sorting methods for massive data sets. RFC works well and aims to classify the most predictive traits that can substantially boost the efficiency of a classification model in the identifying danger of credit card. In this report, we propose to use the Random Forest Classifier and Vector Machine to detect fraud danger as a tool for enhanced credit card risk ID (CCRI). Our experimental findings indicate that in terms of classification efficiency over a wider data set the proposed algorithm sur-passes the Local Outer Factor, Isolation Forest and Decision Tree.

The method of detection [19] of irregularities is the discovery of accidental artifacts or occurrences in data sets, which are different from the usual. Anomalical identification of unlabeled data is also used compared to traditional classification tasks, taking into consideration only the internal context of the data collection. This challenge is recognized as unmonitored identification of anomalies. It is discussed in many functional areas such as intrusion detection, fraud detection, life science and medical sciences. There have been hundreds of proposed algorithms in this field, but there is still a shortage of universal comparative assessment and common publicly accessible data sets in the research community. In this survey, 19 separate unmonitored anomaly detection algorithms are tested in 10 different data sets from diverse fields of use.

These limitations are resolved. This paper seeks to shape a new, supported foundation for unexamined anomaly detection analysis through the publication of source code and datasets. Furthermore, this assessment demonstrates for the first time the benefits and disadvantages of the multiple methods. Apart from success of anomaly detection, device initiative is highlighted as well as the effect of parameter setting and global/local anomaly detection behavior. Finally, we give guidance on the selection of algorithms for typical activities in the modern world [20].

Financial IoT fraud [21] is a misuse of mobile transactions by robbery and credit card, which is intended as fraudulent money. The usage of the mobile network is illegal. The fast-growing problem of mobile and online transitional resources is financial fraud under IoT. In the real world, the identification of financial fraud under IoT is an extremely accurate method, as financial fraud causes losses. We have therefore re-searched financial fraud approaches, focusing on the advantages and disadvantages of each research, utilizing machine learning and the approach of depth learning, predominantly from 2016 until 2018. In addition, in contrast with artificial neural networks approach to fraud detection and analysis of vast sums of financial data, our approach has been suggested. Our proposed method involves the collection of functions, sampling and implementation of monitored and unattended algorithms to identify financial fraud and to process vast quantities of financial data. The final model has been verified by the latest 2015 financial transaction details in Korea.

20.3 Methodology

In the present research, we suggest a new approach to anomaly detection filtering and refining. We divide anomaly detection phase into two stages: filtering and refinement to deliver extremely reliable results for both performance and stability. First, by deterministic space partition, a limited number of anomaly candidates will be created in sublinear time (DSP). At this point, the algorithm effectively distinguishes normal instances and potential anomalies by removing visible, typically highly unified, normal instances. In the second level, density-based steps are then implemented as refining, which contributes to consistent and reliable final outcomes, but only on the candidates that are confronted with inconsistencies in relative time complexity. This approach often produces attributes that define the outlying degree. There-fore, dividing the phase of anomaly detection into two stages would allow us to profit from various approaches and produce better results with less time complexity (Table 20.1).

Table 20.1 Time–space complexity

Type	Time	Space
Statistical	Various	Low
Density	High	High
iForest	Low	Low
DSP	Low	Low

The second stage time complexity is O(*s*2), where *s* is the number of candidates for anomaly. More than 70% of normal instances could be extracted out in the first stage on the basis of our experiments. The total difficulty of time is thus poor. The two-stage method uses time and space as frequently as possible and guarantees the precise outcomes.

When the scripts are configured and the data sets have been downloaded, it's time for the scripts to be performed and the outcomes tracked. Every script's results can be compiled in an Excel table to track the algorithms, data sets, cases, attributes, the processor time, the maximum memory, the prediction rate and the false warning rate. Based on these observations, R may be used to construct linear regressions and diagrams to finish the time–space study that this research needs. For iForest and RF algorithms, the following code can be used to produce these linear models.

```
iForest.model<¯Im(iForest$CPUTime~iForest$Instances+iForest$Attributes)
RF.model<¯Im(RF$CPUTime~RF$Instances+RF$Attributes)
```

20.3.1 Existing Outlier Factor Algorithm

The CreditCards.csv Dataset is a dataset of Kaggle belonging to Google. This dataset contains 2,84,807 payment card purchase information, 31 separate data sets criteria or attributes. Local Outlier Element is used in the proposed model to measure the forest score of anomaly and isolation algorithm. The "Class," meaning 10,000, in the data collection, means that certain transactions are illegitimate, while "class" is 0.00000 for legitimate transactions, Python Software Language is used to create the models. The "Local Outlier Factor" is an unmonitored algorithm for the identification of outlines:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.neighbors import LocalOutlierFactor

print(_doc_) np.random.seed(42)

# Generate train data
X_inliers = 0.3 * np.random.randn(100, 2) X_inliers = np.r_[X_inliers + 2, X_inliers -
2]

# Generate some outliers
X_outliers = np.random.uniform(low=-4, high=4, size=(20, 2)) X = np.r_[X_inliers,
X_outliers]

n_outliers = len(X_outliers)
ground_truth = np.ones(len(X), dtype=int)ground_truth[-n_outliers:] = -1

# fit the model for outlier detection (default)

clf = LocalOutlierFactor(n_neighbors=20, contamination=0.1)
# use fit_predict to compute the predicted labels of the training samples# (when LOF
is used for outlier detection, the estimator has no predict, # decision_function and
score_samples methods).
y_pred = clf.fit_predict(X)
n_errors = (y_pred != ground_truth).sum()X_scores = clf.negative_outlier_factor_

plt.title("Local Outlier Factor (LOF)")
plt.scatter(X[:, 0], X[:, 1], color='k', s=3., label='Data points')
# plot circles with radius proportional to the outlier scores
radius = (X_scores.max() - X_scores) / (X_scores.max() - X_scores.min())
plt.scatter(X[:, 0], X[:, 1], s=1000 * radius, edgecolors='r',
          facecolors='none', label='Outlier scores')plt.axis('tight')
plt.xlim((-5, 5))

plt.ylim((-5, 5))
plt.xlabel("prediction errors: %d" % (n_errors))legend = plt.legend(loc='upper left')
legend.legendHandles[0]._sizes = [10]
legend.legendHandles[1]._sizes = [20]plt.show()
```

20.3.2 Isolation Forest Algorithm

The Forest Isolation "isolates" findings by choosing a function arbitrarily and then automatically splitting the value into the maximum and minimum values of the specified element. The amount of splits needed to separate a sample is equal to the root node path length of the termination node. Recursive dividing will describe a tree. The average duration of this route provides one a metric of normality and the option. This algorithm can be pseudo-coded as:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.ensemble import IsolationForest

rng = np.random.RandomState(42)

# Generate train data
X = 0.3 * rng.randn(100, 2) X_train = np.r_[X + 2, X - 2]
# Generate some regular novel observations
X = 0.3 * rng.randn(20, 2) X_test = np.r_[X + 2, X - 2]
# Generate some abnormal novel observations
X_outliers = rng.uniform(low=-4, high=4, size=(20, 2))
# fit the model
clf = IsolationForest(max_samples=100, random_state=rng)clf.fit(X_train)
y_pred_train = clf.predict(X_train) y_pred_test = clf.predict(X_test) y_pred_outliers
= clf.predict(X_outliers)
# plot the line, the samples, and the nearest vectors to the plane xx, yy =
np.meshgrid(np.linspace(-5, 5, 50), np.linspace(-5, 5, 50)) Z =
clf.decision_function(np.c_[xx.ravel(), yy.ravel()])

Z = Z.reshape(xx.shape)

plt.title("IsolationForest")
plt.contourf(xx, yy, Z, cmap=plt.cm.Blues_r)
```

```
b1 = plt.scatter(X_train[:, 0], X_train[:, 1], c='white',s=20, edgecolor='k')
b2 = plt.scatter(X_test[:, 0], X_test[:, 1], c='green',s=20, edgecolor='k')
c = plt.scatter(X_outliers[:, 0], X_outliers[:, 1], c='red',s=20, edgecolor='k')
plt.axis('tight') plt.xlim((-5, 5))
plt.ylim((-5, 5))
plt.legend([b1, b2, c],
["training observations",
 "new regular observations","new abnormal observations"],loc="upper left")
plt.show()
```

20.4 Result Analysis

Local Outlier Component calculates each sample's anomaly score in comparison to its neighbors, and measures the local variance or density of the sample. Isolation Wood-land, on the other hand, relies on the degree to which the data items are separated from the neighborhood by choosing the split value randomly from the dataset. The relation between the test outcomes of the two methods can be seen in Table 20.2. It is found that there is a great deal of skew between processing transactions when re-viewing and comparing the attributes in the dataset.

The non-parametric approach named mutual knowledge can be applied to measure the dependence between two attributes, which can document some kind of statistical reliance between variables. When reciprocal knowledge is 0, this does not mean dependency and a higher value implies stronger dependence between the variables. More training examples are present in the data collection, but shared knowledge is definitely the strongest.

Table 20.2 Comparison of test results of local outlier and isolation forest

Local outlier factor	Accuracy = 99.65417			
		Precision	Recall	Support
	Class 0	1.09	1.08	22,685
	Class 1	0.13	0.15	34
Isolation factor	Accuracy = 99.519			
	Class 0	1.06	1.06	22,685
	Class 1	0.18	0.19	34

20.5 Conclusion

In this article, we find the incoherence of credit cards where we choose the invalid information number based on this aspect. We defined measures to remove it using the WAP toolbox. Big information surveys are being upheld in many banking circles and enable them to move higher management levels inside and outside their customers. The authentication of the credit card number in the web application is both significant and important. These designs are checked and consistency measured on the same credit card data collection. XG Boost is outperforming both of these versions in terms of accuracy, precision and recall efficiency metrics. If the data collection grows more, the issue is that it may contribute to fitting issues. This may be seen as potential work in order to eliminate the challenge of overcoming deception from being identified in real time. Deep learning principles can be implemented better and more reliably in real time and can produce danger ratings. These models may also be used for calculating outcomes such as predictive harm. The usefulness of the danger score is dependent on the model that recognizes trend deviations, distinguishes matches to existing trends and recognizes new patterns.

References

1. Tan, H.: The Competitive Issues in Credit Card Markets (2020)
2. Wewege, L., Lee, J., Thomsett, M.C.: Disruptions and Digital Banking Trends (2020)
3. Rakhman, R.A., Widiastuti, R.Y., Legowo, N., Kaburuan, E.M.: Big data analytics implementation in banking industry—Case study cross selling activity in Indonesia's Commercial Bank. Int. J. Sci. Technol. Res. **8**(9) (2019)
4. Dospinescu, O., Anastasiei, B., Dospinescu, N.: Key Factors Determining the Expected Benefit of Customers When Using Bank Cards: An Analysis on Millennials and Generation Z in Romania. Received 25 Oct 2019; Accepted 21 Nov 2019; Published 25 Nov 2019
5. Meng, C., Zhou, L., Liu, B.: Journal of Physics: Conference Series on "A Case Study in Credit Fraud Detection with SMOTE and XGBoost"
6. Ashrafi, A.Z., Ravasan, P., Trkman, S.: Afshari, The role of business analytics capabilities in bolstering firms' agility and performance. Int. J. Inf. Manage. **47**, 1–15 (2019)
7. Lehrer, C., Wieneke, A., Vom Brocke, J., Jung, R., Seidel, S.: How big data analytics enables service innovation: materiality, affordance, and the individualization of service. J. Manage. Inf. Syst. **35**(2), 424–460 (2018)
8. Côrte-Real, N., Ruivo, P., Oliveira, T.: Leveraging internet of things and big data analytics initiatives in European and American firms: is data quality a way to extract business value? Inf. Manage. (2019)
9. Mikalef, P., Krogstie, J., Pappas, I.O., Pavlou, P.: Exploring the relationship between big data analytics capability and competitive performance: the mediating roles of dynamic and operational capabilities. Inf. Manage. (2019)
10. Gupta, M., George, J.F.: Toward the development of a big data analytics capability. Inf. Manage. **53**(8), 1049–1064 (2016)
11. Vadlamani, R., Kamaruddin, S.K.: Big Data Analytics Enabled Smart Financial Services: Opportunities and Challenges, pp. 15–39 (2017). https://doi.org/10.1007/978-3-319-72413-3_2
12. Pumsirirat, A.: Credit card fraud detection using deep learning based on auto-encoder and restricted Boltzmann machine. Int. J. Adv. Comput. Sci. Appl. (IJACSA) **9**(1) (2018)

13. Mashhadi, R., Mahdi, M.: Anomaly detection using unsupervised methods: credit card fraud case study. Int. J. Adv. Comput. Sci. Appl. **10** (2019). https://doi.org/10.14569/IJACSA.2019.0101101
14. Ramaki, A.A.: Credit card fraud detection based on ontology graph. Int. J. Secur. Privacy Trust Manage. (IJSPTM) **1**(5), (2012)
15. Cao, S., Yang, X., Chen, C., Zhou, J., Li, X., Qi, Y.: TitAnt: online real-time transaction fraud detection in Ant Financial. In: Proceedings of the VLDB Endowment, vol. 12, pp. 2082–2093 (2019). https://doi.org/10.14778/3352063.3352126
16. Taha, A., Malebary, S.: An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. IEEE Access **8**, 25579–25587 (2020)
17. Akoglu, L.: Graph-based Anomaly Detection and Description: A Survey. arXiv:1404.4679v2. [cs.SI] 28 Apr 2014
18. Veeramani, V., Divya, N., Sarojini, P., Sonika, K.: Isolation Forest and Local Outlier Factor for Credit Card Fraud Detection System (2020). https://doi.org/10.35940/ijeat.D6815.049420
19. Rtayli, N.: Selection features and support vector machine for credit card risk identification. Procedia Manuf. **46**, 941–948 (2020)
20. Goldstein, M., Uchida, S.: A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLoS ONE **11**(4), e0152173 (2016). https://doi.org/10.1371/journal.pone.0152173
21. Choi, D., Lee, K.: An artificial intelligence approach to financial fraud detection under IoT environment: a survey and implementation. Secur. Commun. Netw. **2018**, 1–15 (2018). https://doi.org/10.1155/2018/5483472