Volume 7, Issue 12, December 2022 ISSN (online): 2456-0006
International Journal of Science Technology Management and Research
Available online at: www.ijstmr.com

Recent Survey on Various Outlier Detection Techniques and Advantage Disadvantages

Narendra Suman
M.Tech 4th Sem.
Computer Science and Engineering Department
Lakshmi Narain College of Technology and Science
Indore M. P. India

Miss Sanjana Sharma
Assistant Professor
Computer Science and Engineering Department
Lakshmi Narain College of Technology and Science
Indore M. P. India

Abstract-: ***Outlier detection is a key consideration within the development and deployment of any model and also data analysis. Identifying and dealing with outliers is an integral part of working with data, and data analysis. In this data-rich environment, organisations can expect to have to deal with outlier data. Outliers can skew trends and have a serious impact on the accuracy data analysis. The presence of outliers can be a sign of concept drift, so ongoing outlier analysis is needed. Outliers can skew results, and anomalies in training data can impact overall model effectiveness. Outlier detection is a key tool in safeguarding data quality, as anomalous data and errors can be removed and analysed once identified. Outlier detection is an important part of each stage of the machine learning process. Accurate data is integral during the development and training of algorithms, and outlier detection is performed after deployment to maintain the effectiveness of models. This guide explores the basics of outlier detection techniques in machine learning, and how they can be applied to identify different types of outlier. In this paper we proposed various outlier detection methods and also there advantages and disadvantages***
Keywords: Attribute, Discovery, dependency, Integrity, constraints, normalization

I. INTRODUCTION

Outlier detection is a key consideration within the development and deployment of machine learning algorithms. Models are often developed and leveraged to perform outlier detection for different organisations that rely on large datasets to function. Economic modelling, financial forecasting, scientific research, and ecommerce campaigns are some of the varied areas that machine learning-driven outlier detection is used.Identifying and dealing with outliers is an integral part of working with data, and machine learning is no different. Algorithm development usually relies on huge arrays of training data to achieve a high level of accuracy. Once deployed, models will process huge amounts of data, providing insights into trends and patterns. In this data-rich environment, organisations can expect to have to deal with outlier data. Outliers can skew trends and have a serious impact on the accuracy of models. The presence of outliers can be a sign of concept drift, so ongoing outlier analysis in machine learning is needed. Machine learning models learn from data to understand the trends and relationship between data points. Outliers can skew results, and anomalies in training data can impact overall model effectiveness. Outlier detection is a key tool in safeguarding data quality, as anomalous data and errors can be removed and analysed once identified. Outlier detection is an important part of each stage of the machine learning process. Accurate data is integral during the development and training of algorithms, and outlier detection is performed after deployment to maintain the effectiveness of models. This guide explores the basics of outlier detection techniques in machine learning, and how they can be applied to identify different types of outlier.

An outlier is an individual point of data that is distant from other points in the dataset. It is an anomaly in the dataset that may be caused by a range of errors in capturing, processing or manipulating data. Outliers can skew overall data trends, so outlier detection methods are an important part of statistics. Outliers will be a consideration for any area that uses data to make decisions. If an organisation is gaining insight from data, outliers are a real risk.Outlier detection is particularly important within machine learning. Models are trained on huge arrays of training data. The model understands the relationship between data points to help predict future events or categorise live data. Outliers in the training data may skew the model, lowering its accuracy and overall effectiveness. Outlier analysis and resolution can lengthen the training time too. Outliers can be present in any data or machine learning use case, whether that's financial modelling or business performance analysis [9,10].

II. TWO TYPES OF OUTLIER DETECTION METHODS

There are various methods of outlier detection is as follows −

Supervised Methods − Supervised methods model data normality and abnormality. Domain professionals tests and label a sample of the basic data. Outlier detection can be modeled as a classification issue. The service is to understand a classifier that can identify outliers. The sample can be used for training and testing. In various applications, the professionals can label only the normal objects, and several objects not connecting the model of normal objects are documented as outliers. There are different methods model the outliers and consider objects not connecting the model of outliers as normal.

Unsupervised Methods − In various application methods, objects labeled as "normal" or "outlier" are not applicable. Therefore, an unsupervised learning approach has to be used. Unsupervised outlier detection methods create an implicit assumption such as the normal objects are considerably "clustered." An unsupervised outlier detection method predict that normal objects follow a pattern far more generally than outliers. Normal objects do not have to decline into one team sharing large similarity. Instead, they can form several groups, where each group has multiple features. This assumption cannot be true sometime. The normal objects do not send some strong patterns. Rather than, they are uniformly distributed. The collective outliers, share large similarity in a small area.
Unsupervised methods cannot identify such outliers efficiently. In some applications, normal objects are separately distributed, and several objects do not follow strong patterns. For example, in some intrusion detection and computer virus detection issues, normal activities are distinct and some do not decline into high-quality clusters.
Some clustering methods can be adapted to facilitate as unsupervised outlier detection methods. The main idea is to discover clusters first, and therefore the data objects not belonging to some cluster are identified as outliers. However, such methods deteriorate from two issues. First, a data object not belonging to some cluster can be noise rather than an outlier. Second, it is expensive to discover clusters first and then discover outliers.
Semi-Supervised Methods − In several applications, although obtaining some labeled instance is possible, the number of such labeled instances is small. It can encounter cases where only a small group of the normal and outlier objects are labeled, but some data are unlabeled. Semi-supervised outlier detection methods were produced to tackle such methods.Semi-supervised outlier detection methods can be concerned as applications of semisupervised learning approaches. For example, when some labeled normal objects are accessible, it can use them with unlabeled objects that are nearby, to train a model for normal objects. The model of normal objects is used to identify outliers those objects not suitable the model of normal objects are defined as outliers[7,8].

III. WHY DO WE NEED OUTLIER ANALYSIS

Data and analysis is increasingly becoming an integral part of everyday business decisions and management. Organisations rely on setting and measuring key performance indicators to evaluate business performance. Monitoring datasets is key to maintaining product quality, achieving high-value marketing campaigns, driving sales decisions, or ensuring user experience statistics are positive. With a growing emphasis on data-led decision making across different organisations, trust in the quality of data is vital. Outlier analysis plays an important role in maintaining this trust.

Outliers can skew trends and forecasts modelled from datasets, negatively impacting the quality and accuracy of decisions. Actively monitoring and performing outlier detection can flag errors in datasets and combat concept drift in machine learning models. If outliers are not identified and removed, models can become less accurate and effective.

IV. LITERATURE SUREY

In 2012 Karanjit Singh and Dr. Shuchita Upadhyaya proposed "**Outlier Detection: Applications And Techniques"** Outliers once upon a time regarded as noisy data in statistics, has turned out to be an important problem which is being researched in diverse fields of research and application domains. Many outlier detection techniques have been developed specific to certain application domains, while some techniques are more generic. Some application domains are being researched in strict confidentiality such as research on crime and terrorist activities. The techniques and results of such techniques are not readily forthcoming. A number of surveys, research and review articles and books cover outlier detection techniques in machine learning and statistical domains individually in great details. They attempt to bring together various outlier detection techniques, in a structured and generic description. With this exercise, we hope to attain a better understanding of the different directions of research on outlier analysis for ourselves as well as for beginners in this research field who could then pick up the links to different areas of applications in details.[1]

In 2013 Ji Zhang proposed "**Advancements of Outlier Detection: A Survey"** Outlier detection is an important research problem in data mining that aims to discover useful abnormal and irregular patterns hidden in large datasets. In this paper, we present a survey of outlier detection techniques to reflect the recent advancements in this field. The survey will not only cover the traditional outlier detection methods for static and low dimensional datasets but also review the more recent developments that deal with more complex outlier detection problems for dynamic/streaming and high-dimensional datasets. They proposed comprehensive survey is presented to review the existing methods for detecting point outliers from various kinds of vector-like datasets. The outlier detection techniques that are primarily suitable for relatively low-dimensional static data, which serve the technical foundation for many of the methods proposed later, are reviewed first. We have also reviewed some of recent advancements in outlier detection for dealing with more complex high-dimensional static data and data streams.
It is important to be aware of the limitation of this survey. As it has clearly stated in Section 2, we only focus on the point outlier detection methods from vector-like datasets due to the space limit. Also, outlier detection is a fast-developing field of research and more new methods will quickly emerge in the foreseeable near future. Driven by their emergence, it is believed that outlier detection techniques will play an increasingly important role in various practical applications where they can be applied to[2].

In 2014 Manish Gupta, Jing Gao proposed "Outlier Detection for Temporal Data: A Survey". They presented an organized overview of the various techniques proposed for outlier detection on temporal data. Modeling temporal data is a challenging task due to the dynamic nature and complex evolutionary patterns in the data. In the past, there are a wide variety of models developed to capture different facets in temporal data outlier detection. This survey organized the discussion along different data types, presented various outlier definitions, and briefly introduced the corresponding techniques. Finally, we also discussed various applications for which these techniques have been successfully used. This survey provides a number of insights and lessons as follows. • The methods for different data types are not easy to generalize to one another, though some of them may have similarity in the framework at the broader level. For example, change detection in continuous time series and discrete time series both require forecasting methods. However, the specific kind of forecasting method is extremely different in the two scenarios (regression models in one vs Markov Models in the other). • Most window-based models are currently offline, whereas online methods do exist for point based models. Therefore, there is significant opportunity for research in the former. • While the number of formulations of the temporal outlier detection problem are diverse, they are generally motivated by the most common applications which are encountered in the literature. Many recent applications, especially those corresponding to novel data types in the context [3].

In 2015 Christy.Aa , MeeraGandhi.Gb , S. Vaithyasubramanianc proposed "Cluster Based Outlier Detection Algorithm For Healthcare Data". They we utilize the concept of data preprocessing for outlier reduction. We propose two algorithms namely Distance-Based outlier detection and Cluster-Based outlier algorithm for detecting and removing outliers using a outlier score. By cleaning the dataset and clustering based on similarity, we can remove outliers on the key attribute subset rather than on the full dimensional attributes of dataset. Experiments were conducted using 3 built-in Health care dataset available in R package and the results show that the cluster-based outlier detection algorithm providing better accuracy than distance-based outlier detection algorithm. The goal of the algorithms presented is to improve the quality of data processing and capture the underlying patterns in the data by reducing the effect of outliers at the pre-processing stage. This outlier may be due to the unavailability or distortions in the data collection stage that consists of irrelevant or weakly relevant data objects. From the algorithms, it is shown that by choosing a valid outlier score, the overall performance of the algorithm can be improved[4].

In 2016 Kamaljeet Kaur Atul Garg proposed "**Comparative Study of Outlier Detection Algorithms".** As the dimension of the data is increasing day by day, outlier detection is emerging as one of the active area of research. Finding of the outliers from large data sets is the main problem. Outlier is considered as the pattern that is different from the rest of the patterns present in the data set. The detection of the outlier in the data set is an important process as it helps in acquiring the useful information that further helps in the data analysis. Various algorithms have been proposed till date for the detection of the outliers. This paper covers a study of various outlier detection algorithms like Statistical based outlier detection, Depth based outlier detection, Clustering based technique, Density based outlier detection etc. Comparison study of these outlier detection methods is done to find out which of the outlier detection algorithms are more applicable on high dimensional data. The speed of processing the data is to be increased that helps in the reduction of processing cost of data. There is no single universally applicable outlier detection approach of the current techniques. This paper presents the study of different existing outlier detection techniques and the way in which they are categorized. It is concluded that performance of clustering algorithms is comparatively better than other outlier detection algorithms on huge data sets. It is found that efficiency and computational complexity depends upon the data distribution and type of data. It is also observed that no individual algorithm is much suited for the high dimensional data. There is need of developing some new algorithms or improvement in the existing one is required[5].

In 2017 R. Muthukrishnan and G. Poonkuzhali proposed "A Comprehensive Survey on Outlier Detection Methods". In ancient days, outlier is viewed as noisy data in statistics, has turned out to be a vital problem which is being researched in different fields of application domains. Outlier detection is a primary step in many scientific research studies, because it has a negative impact on the results. Visual inspection alone cannot always identify an outlier and it can lead to mislabelling an observation as an outlier. Using a specific function of the observations leads to a superior outlier labelling rule. The estimation of parameters with classical measures such as mean is highly sensitive to outliers. Statistical methods were developed to accommodate outliers and to reduce their impact on the analysis. Numerous outlier detection methods have been developed specific to certain application domains, while few methods are more general. Outlier detection has been explored in a much broader area including discriminant analysis, experimental design, multivariate data, linear models etc. There are various approaches to outlier detection depending on the application and number of cases/variables in the data set. An attempt has been made to review the outlier detection methods which are entrenched and commonly used now-a-days. This paper provided a survey on the structure of existing outlier detection methodologies[6].

In 2017 Anwesha Barai (Deb), Lopamudra Dey proposed "Outlier Detection and Removal Algorithm in K-Means and Hierarchical Clusterin" An outlier in a pattern is dissimilar with rest of the pattern in a dataset. Outlier detection is an important issue in data mining. It has been used to detect and remove anomalous objects from data. Outliers occur due to mechanical faults, changes in system behaviour, fraudulent behaviour, and human errors. This paper describes the methodology or detecting and removing outlier in K-Means and Hierarchical clustering. First apply clustering algorithm K-Means and Hierarchical clustering on a data set then find outliers from the each resulting clustering. In K-Means clustering outliers are found by distance-based approach and cluster based approach. In case of hierarchical clustering, by using dendrogram outliers are found. The goal of the project is to detect the outlier and remove the outliers to make the clustering more reliable They presented outlier detection method in both K-Means and Hierarchical Clustering. To remove outlier is an important task. They proposed algorithms by which we can remove outliers. We work on benchmark dataset and after implementing our proposed algorithm it is proved that our proposed algorithm is more efficient than previous one. After removing the outliers' accuracy are increased. The approach needs to be implemented on more complex datasets. Future work requires approach applicable for varying datasets[7].

In 2018 Sampson Twumasi-Ankrah, Simon Kojo Appiah, Doris Arthur Proposed "Comparison Of Outlier Detection Techniques In **Non-Stationary Time Series Data".** They This study examined the performance of six outlier detection techniques using a non-stationary time series dataset. Two key issues were of interest. Scenario one was the method that could correctly detect the number of outliers introduced into the dataset whiles scenario two was to find the technique that would over detect the number of outliers introduced into the dataset, when a dataset contains only extreme maxima values, extreme minima values or both. Air passenger dataset was used with different outliers or extreme values ranging from 1 to 10 and 40. The six outlier detection techniques used in this study were Mahalanobis distance, depth-based, robust kernel-based outlier factor (RKOF), generalized dispersion, Kth nearest neighbors distance (KNND), and principal component (PC) methods. When detecting extreme maxima, the Mahalanobis and the principal component methods performed better in correctly detecting outliers in the dataset. Also, the Mahalanobis method could identify more outliers than the others, making it the "best" method for the extreme minima category. The kth nearest neighbor distance method was the "best" method for not over-detecting the number of outliers for extreme minima[8].

In 2019 Stefan Larson Anish Mahendran Andrew Lee Jonathan K. Kummerfeld Outlier Detection for Improved Data Quality and Diversity in Dialog Systems In a corpus of data, outliers are either errors: mistakes in the data that are counterproductive, or are unique: informative samples that improve model robustness. Identifying outliers can lead to better datasets by (1) removing noise in datasets and (2) guiding collection of additional data to fill gaps. However, the problem of detecting both outlier types has received relatively little attention in NLP, particularly for dialog systems. They introduce a simple and effective technique for detecting both erroneous and unique samples in a corpus of short texts using neural sentence embeddings combined with distance-based outlier detection. We also present a novel data collection pipeline built atop our detection technique to automatically and iteratively mine unique data samples while discarding erroneous samples. Outliers are often the most interesting parts of our data, but outlier detection has received relatively little attention in NLP beyond its application to finding annotation errors. They introduce the first neural outlier detection method for short text and demonstrates its effectiveness across multiple metrics in multiple experiments. We also propose a way to integrate outlier detection into data collection, developing and evaluating a novel crowdsourcing pipeline. This pipeline supports the creation of higher quality datasets to yield higher quality models by both reducing the number of errors and increasing the diversity of collected data. While the experiments discussed herein are concerned with components of dialog systems[9]

In 2019 Hongzhi Wang , Mohamed Jaward Bah , and Mohamed Hammad proposed **“Progress in Outlier Detection Techniques: A Survey**” They offer the fundamental concepts of outlier detection and then categorize them into different techniques from diverse outlier detection techniques, such as distance-, clustering-, density-, ensemble-, and learning-based methods. In each category, we introduce some state-of the- art outlier detection methods and further discuss them in detail in terms of their performance. Second, we delineate their pros, cons, and challenges to provide researchers with a concise overview of each technique and recommend solutions and possible research directions. They gives current progress of outlier detection techniques and provides a better understanding of the different outlier detection methods. The open research issues and challenges at the end will provide researchers with a clear path for the future of outlier detection methods. They have provided a comprehensive survey in a structured manner that reviews state-of-the-art methods of detecting outliers by grouping them into different categories. We have grouped the algorithms into density-based, statistical-based, distance-based, clustering-based, ensemble based, and learning-based approaches[10].

In 2020 Clement Franklin D and Ms. Sharon Femi proposed **“ Comparing the Performance of Anomaly Detection Algorithms**”. An Anomaly is a data point which differs in characteristics from other data points in the dataset. The detection of anomaly plays an important role in machine learning. But most of the algorithms provide anomaly detection only with limited generalization capacity. They, compare the efficiency of anomaly detection methods which has better robustness. The three outlier detection algorithms used are Local Outlier Factor, Isolation Forest and Autoencoders. Based on the accuracy, recall, precision, F1 score of the algorithms, the comparison graph is constructed for the three datasets and the efficient algorithm is determined. The anomaly has been detected for the identified datasets and the correlation matrix has been determined. The three algorithms are implemented i.e.. the LOF, Isolation Forest and the Autoencoder algorithms are implemented for the three different datasets namely Corona disease, Breast cancer, and Heart disease dataset the accuracies are compared and the efficient algorithm is found. In the future, the double level ensemble strategy will be used to combine algorithms and produce a better accuracy rate which in turn produces a better anomaly detection algorithm that detects outliers in datasets[11].

In 2021 Zoheir Sabeur, Gianluca Correndo and Galina Veres proposed **“Outlier and Anomaly Detection Methods with Applications to the 2021 Census”** The purpose here was to investigate the ability of machine learning techniques to detect the introduced ‘198x -> 189x’ error, along with any other anomalies present in the sample dataset. The following machine learning techniques were selected for implementation and investigation: K-means clustering, KAMILA, Rule-based approach, SPAD, One-class Support Vector Machine for novelty detection, Pattern Outlier Detector based on Support Vector Machines, and Pattern Outlier Detector based on Boosted Regression Trees. While most of these approaches were described in the Phase 1 Report, a short description for new approaches is given below. Given both the nature of the problem and the type of data considered (census data where all anomalies were already filtered out), there was no ground truth to use in training the models. All methods were therefore unsupervised the introduction of errors into the census microdata proved to be extremely challenging. The available datasets we acquired had already been edited and processed to remove the most obvious anomalies; however, this process also altered the underlying data distributions. The cleaned data thus did not carry the true statistical signatures of the various classes of embedded anomalies. These would ordinarily manifest themselves with their specific weights (energies) during census processing. We therefore recognize that it is difficult to synthetically introduce many of the types of errors and their respective natural statistical significance which have been identified by ONS for testing anomaly detection methods[12].

V. ADVANTAGES AND DISADVANTAGES

Comparison based on advantages and disadvantages

A. *Method Type*	B. *Advantages*	C. *Disadvantages*
D. ***Statistical Methods***	1. They are mathematically justified and if a probabilistic model is given, the methods are very efficient and it is possible to reveal the meaning of the outliers found .	They are typically not applied in a multi-dimensional scenario because most distribution models typically apply to the univariate feature space.
	2. The model constructed, often presented in a compact form, makes it possible to detect outliers without storing the original datasets that are usually of large sizes.	They are unsuitable even for moderate multi-dimensional data sets.
E. ***Distance-based Methods***	1. The major advantage of distance-based algorithms is that, unlike distribution-based methods, distance-based methods are non-parametric	Most of them are not effective in high-dimensional space due to the curse of dimensionality.
	2. They do not rely on any assumed distribution to fit the data. The distance-based definitions of outliers are straightforward and easy to understand and implement.	For many real databases where there are often millions of records. Thus, these approaches lack a good scalability for large data set.
F. ***Density-based Methods***	1. The density-based outlier detection methods are generally more effective than the distance-based methods.	To achieve the improved effectiveness, the density-based methods are more complicated and computationally expensive.
G. ***Clustering-based Methods***	1. Their objective is only to group the objects in dataset such that clustering functions can be optimized.	In contrast to the various definitions of outliers in outlier detection which are more objective and independent of how clusters in the input data set are identified.
	2. The aim to eliminate outliers in dataset using clustering is only to dampen their adverse effect on the final clustering result	In designing new outlier detection approaches is to directly model outliers and detect them without going though clustering the data first.

CONCUSION

Outlier detection is a key consideration within the development and deployment of any model and also data analysis. Identifying and dealing with outliers is an integral part of working with data, and data analysis. In this data-rich environment, organisations can expect to have to deal with outlier data. Outliers can skew trends and have a serious impact on the accuracy data analysis. The presence of outliers can be a sign of concept drift, so ongoing outlier analysis is needed. In this paper we presented a comparative study over some outlier detection method based on advantage and disadvantage. There are various methods are available to detect outlier form given dataset, each methods have some limitation and some advantages. We found that we need to select the methods based on the nature of the dataset. No Sigle method can we used for every kind of outlier detection.

REFERENCE

1. Karanjit Singh and Dr. Shuchita Upadhyaya Outlier Detection: Applications and Techniques IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 3, January 2012 ISSN (Online): 1694-0814 www.IJCSI.org
2. Ji Zhang Department of Mathematics and Computing University of Southern Queensland, Australia Advancements of Outlier Detection: A Survey ICST Transactions on Scalable Information Systems Research Article.
3. Manish Gupta, Jing Gao, Member, IEEE, Charu C. Aggarwal, Fellow, IEEE, and Jiawei Han, Fellow, IEEE Outlier Detection for Temporal Data: A Survey IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 1, JANUARY 2014.
4. Christy.Aa , MeeraGandhi.Gb , S. Vaithyasubramanianc Cluster Based Outlier Detection Algorithm For Healthcare Data © 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Peer-review under responsibility of scientific committee of 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15) doi: 10.1016/j.procs.2015.04.058.

5. Kamaljeet Kaur Atul GargComparative Study of Outlier Detection Algorithm International Journal of Computer Applications (0975 – 8887) Volume 147 – No. 9, August 2016.
6. R. Muthukrishnan and G. Poonkuzhal A Comprehensive Survey on Outlier Detection Method American-Eurasian Journal of Scientific Research 12 (3): 161-171, 2017 ISSN 1818-6785 © IDOSI Publications, 2017 DOI: 10.5829/idosi.aejsr.2017.161.171
7. Anwesha Barai (Deb), Lopamudra Dey Outlier Detection and Removal Algorithm in K-Means and Hierarchical Clustering World Journal of Computer Application and Technology 5(2): 24-29, 2017 http://www.hrpub.org DOI: 10.13189/wjcat.2017.050202
8. Sampson Twumasi-Ankrah, Simon Kojo Appiah Comparison of Outlier Detection Techniques In Non-Stationary Time Series Data Global Journal of Pure And Applied Sciences Vol. 27, 2021: 55-60 Copyright© Bachudo Science Co. Ltd Printed In Nigeria ISSN 1118-0579
9. Stefan Larson Anish Mahendran Ann Arbor, MI, USA Outlier Detection for Improved Data Quality and Diversity in Dialog System Proceedings of NAACL-HLT 2019, pages 517–527 Minneapolis, Minnesota, June 2 - June 7, 2019. c 2019 Association for Computational Linguistics .
10. Hongzhi Wang , Mohamed Jaward Bah , And Mohamed Hammad Progress in Outlier Detection Techniques: A Survey Received July 14, 2019, accepted July 29, 2019, date of publication August 2, 2019, date of current version August 19, 2019. Digital Object Identifier 10.1109/ACCESS.2019.2932769.
11. Clement Franklin D C Comparing the Performance of Anomaly Detection Algorithms International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV9IS070532 (This work is licensed under a Creative Commons Attribution 4.0 International License.) Published by : www.ijert.org Vol. 9 Issue 07, July-2020.
12. Zoheir Sabeur, Gianluca Correndo and Galina Veres Outlier and Anomaly Detection Methods with Applications to the 2021 Census IT Innovation Centre, Department of Electronics and Computer Science Paul A. Smith and James Dawber Southampton Statistical Sciences Research Institute University of Southampton University of Southampton.