Hindawi
Journal of Food Quality
Volume 2021, Article ID 9929049, 8 pages
https://doi.org/10.1155/2021/9929049

Research Article

Open Data Release and Privacy Concerns: Complexity in Mitigating Vulnerability with Controlled Perturbation

Shah Imran Alam,[1] **Ihtiram Raza Khan**,[1] **Syed Imtiyaz Hassan**,[2] **Farheen Siddiqui**,[1] **M. Afshar Alam**,[1] **and Anil Kumar Mahto**[1]

[1]*Department of Computer Science and Engineering, Jamia Hamdard, New Delhi, India*
[2]*Department of Computer Science and Information Technology, School of Technology, Maulana Azad National Urdu University, Hyderabad, India*

Correspondence should be addressed to Shah Imran Alam; shahimranalam@gmail.com

Received 4 April 2021; Accepted 18 May 2021; Published 23 June 2021

Academic Editor: Rijwan Khan

The benefits of open data were realised worldwide since the past decades, and the efforts to move more data under the license of open data intensified. There was a steep rise of open data in government repositories. In our study, we point out that privacy is one of the consistent and leading barriers among others. Strong privacy laws restrict data owners from opening the data freely. In this paper, we attempted to study the applied solutions and to the best of our knowledge, we found that anonymity-preserving algorithms did a substantial job to protect privacy in the release of the structured microdata. Such anonymity-preserving algorithms argue and compete in objectivethat not only could the released anonymized data preserve privacy but also the anonymized data preserve the required level of quality. K-anonymity algorithm was the foundation of many of its successor algorithms of all privacy-preserving algorithms. l-diversity claims to add another dimension of privacy protection. Both these algorithms used together are known to provide a good balance between privacy and quality control of the dataset as a whole entity. In this research, we have used the K-anonymity algorithm and compared the results with the addon of l-diversity. We discussed the gap and reported the benefits and loss with various combinations of K and l values, taken in combination with released data quality from an analyst's perspective. We first used dummy fictitious data to explain the general expectations and then concluded the contrast in the findings with the real data from the food technology domain. The work contradicts the general assumptions with a specific set of evaluation parameters for data quality assessment. Additionally, it is intended to argue in favour of pushing for research contributions in the field of anonymity preservation and intensify the effort for major trends of research, considering its importance and potential to benefit people.

1. Introduction

Open data have proved its importance in the field of research, open governance, development versus analysis, and business initiatives. The release of public open data has emerged as a critical need for the overall development of humanity as one nation. In the COVID-19 pandemic, the world as one entity has recently accepted the relevance of open data and its potential to help curb the spread and come up with a range of solutions. Researchers worldwide used open COVID-19 data to help governments and organizations like WHO enforce measures and suggest policies. However, the threat to individuals to whom the data refers is shown up intensely because of the fear of identity recognition or reidentification of it. In fact, this has always been a rising concern and is being criticized since long back throughout the world, not just in the COVID-19 data but in all such released data where the identity disclosure attack is possible. Researchers have been trying to find a balance between the quality of open data released and the possibility of identity revelation from attacks.

However, in the current pandemic crises, all major policy decisions have their critical base on the open data shared worldwide, which is an important lesson learned that should not

be allowed to be lost. At the same time, the inability to contain privacy threats has put nations and individuals in a weird defensive position. The concern of misuse is genuine in case the identity of individuals is released or regenerated from other existing sources of information which might be available. This has also resulted in regaining the focus towards anonymity concern [1–6].

Open data in a broader sense is defined with generosity in terms and conditions that do not limit people's right to see, use, investigate, and share data [7] although the work is relevant to all such releases which have limitations or controlled access to a small group of researchers.

Some of the important early initiatives which were mass encouragement towards public open data release are listed below for reference in yearwise ascending chronology [8–11].

The US [12], UK [13], and Canada [8, 14] government bodies pioneered and contributed to substantial commitments and efforts.

Open data for researchers with background from different business domains results in a complete critical analysis as well as exploring the hidden benefits from the data at each level. It leads to solutions for medical, technical business, and most important of all, the good and open governance [11, 15–19].

Food industry is the backbone of the health of the citizen in every country. Open data in the big picture is often talked about in health care. In the food technology domain, it generally marks its importance in terms of food quality assessment, regulatory measures, and genomics and to find the correlations between the food people consume with the overall health data of an area as well as consumer's preference analysis. There is enormous data that directly or indirectly comes from the ambit of food technology.

While the perception prevails that privacy concern is mostly about direct health care or defence data, the reality is that it is a barrier in the research and growth of almost every prominent industry. In the context of food technology also the open data struggles with privacy threats. The requirement to analyse consumer awareness and concerns is one such area that has privacy as a major concern and must be dealt with to protect the interest of consumer's personnel information.

Artificial intelligence plays a key role in recommendation systems, quality assessment, and deriving solutions that could improve the overall health of people in a particular locality and in government policy-making. Such AI powered systems need to include the best possible, privacy protection techniques and strengthen them transparently, to gain the trust of the people to whom the data actually refers.

2. Materials and Methods

2.1. Anonymity Algorithms. There has been a substantial effort by the researchers to mitigate the anonymity barrier of structured data release. Numerous anonymity-based algorithms have been proposed till date to preserve the anonymity concerns of the data. However, these algorithms work with suppression and generalization as the key techniques to prevent identity in the data [9, 20]. The sole purpose of these algorithms is to suggest the right balance between "data suppression and generalization" and "the anonymity preservation need" [21, 22]. These algorithms formulate and quantify the trade-off in their best possible ways. However, as they still depend on suppression and generalization as the key techniques, in fact, it means that the quality loss is bound to happen [21, 23].

2.2. Baseline Assumptions of K-Anonymity Algorithm. The K-anonymity algorithm establishes baseline assumptions which are followed by its successor algorithms in general. These assumptions are as follows and extracted from [24]:

(1) The value K emphasizes the minimal number of tuples of all combinations of quasi-identifiers to be exactly the same to make it impossible to be reidentified

(2) All explicit identifiers are assumed to be either suppressed or encrypted for the input data set. Hence, they are ignored completely

(3) Quasi-identifiers are attributes whose combination could be exploited for linking and reidentification in external data sets

Thus, keeping the distortion of the data at the bare minimum level with the least impact on the data distribution, similar parameters are assumed in enhancement to the K-anonymity. However, other statistical techniques like scrambling, swapping, or adding noise to the data values are not employed that can make the data unfit for investigation.

Baseline definition: the first assumption is derived from the definition of K-anonymity requirement; that is, in the released microdata, the combination of a tuple of a quasi-identifier must match at least k-1 tuples.

2.3. Vulnerabilities That Can Be Exploited in the Baseline Assumptions in the Context of Open Data. Although anonymity-based algorithm has been a success since the last decade, many platforms have been developed around them [25, 26]. However, we cannot deny the fact that the literature puts forth many reasoning that anonymity algorithms are not foolproof techniques and bypassing them is possible, and it also becomes much easier in today's data-age [27–29].

This section highlights the gaps that we found in the basic assumptions of the K-anonymity algorithm.

Consider the fictitious open data set of a hospital that is assumed to be released on a daily basis in Table 1.

Assume this release is done with $K = 2$. A generalization of level 1 is good enough on "Date of Birth" and "Zip" fields. Generalization of level 1 for "date Of Birth" field could be defined with the suppression of "dd" in "dd/mm/yyyy" and a generalization of level 1 could be defined as the suppression of the last two digits of the six-digit "Zip".

The table after applying K-anonymity algorithm for $K = 2$ will be as in Table 2.

Now, to illustrate vulnerability, we consider a voter's data as follows, say from an external database.

Table 1: Medical data released as anonymous data.

Date of birth (dd/mm/yyyy)	Zip	Sex	Medical symptoms
07/10/1982	233001	Male	Shortness of breath
18/01/1983	233001	Female	Obesity
23/10/1982	233022	Male	Shortness of breath
11/01/1983	233001	Female	Obesity
26/01/1983	233001	Female	Hypertension
09/09/1982	233052	Male	SLE
15/07/1982	233052	Male	SLE
03/09/1983	233005	Female	Hypertension
05/09/1983	233005	Female	Hypertension

Table 2: Medical data released as anonymous data with 2-anonymity applied.

Date of birth (dd/mm/yyyy)	Zip	Sex	Medical symptoms
XX/10/1982	2330 XX	Male	Shortness of breath
XX/01/1983	2330 XX	Female	Obesity
XX/10/1982	2330 XX	Male	Shortness of breath
XX/01/1983	2330 XX	Female	Obesity
XX/01/1983	2330 XX	Female	Hypertension
XX/09/1982	2330 5X	Male	SLE
XX/07/1982	2330 5X	Male	SLE
XX/09/1983	2330 XX	Female	Hypertension
XX/09/1983	2330 XX	Female	Hypertension

Name	Address	City	Zip	DOB	Sex	Party
Amit Kumar	E-31, Varanasi Road	Ghazipur	233001	07/10/1982	Male	Congress

The above record could be linked to two records in our medical data as depicted by the underlined rows in Table 3 provided someone attempting to figure out the medical symptom of "Amit Kumar" with external information that he visited the hospital that day for a consultation. The target for this reconstruction is to know the medical symptoms of the patient "Amit Kumar." Now as the data set releases satisfy K-anonymity for $K = 2$, it is guaranteed to be at least two such records which will be identical. However, the issue here is that the attacker might be interested in symptoms which are also identical in this case. Thus, the attacker could reidentify the data of the patient "Amit Kumar" in terms of medical symptoms, irrespective of the fact that the released data has been anonymized using K-anonymity.

Now, looking back at the mathematical foundation of the definition of K-anonymity requirement, see the following.

2.3.1. Definition (K-Anonymity). Let $\text{Tab}_X(X_1, X2, \ldots X_N)$ be a table and $\text{QI}_{\text{Tab}X}$ be the quasi-identifier associated with Tab_X. Tab_X is said to satisfy K-anonymity iff, for each quasi-identifier QI-$\text{QI}_{\text{Tab}X}$, each sequence of values in Tab_X [QI] should at least occur K times in T[QI].

Table 3 satisfies the above definition successfully and is compatible with $K = 2$; still, it could be deidentified for a particular critical field.

Thus, a possible way to handle this situation is the modification of the mathematical foundation requirement of the anonymity algorithm, which is an addon restriction imposed by the l-diversity algorithm. We have rephrased key aspects of it as follows.

2.3.2. (K-Anonymity) Leading to l-Diversity. A well-known improvement to tackle the gap in the K-anonymity algorithm's assumption is provided by the l-diversity algorithm [30]. Below is the discussion on the gap and implications of the solution provided by the l-diversity algorithm.

Let $\text{Tab}_X(X_1, X2, \ldots, X_N)$ be a table and $\text{QI}_{\text{Tab}X}$ be the quasi-identifier associated with Tab_X. Tab_X is said to satisfy K-anonymity iff, for each quasi-identifier QI-QITab_X, each sequence of values in Tab_X [QI] should at least occur K times in T[QI]. And the critical field values should be such that they have at least *'l'* enumeration or diversity in the group with k-anonymity.

There, the critical fields are those fields which could have high chances of attack.

Thus, to apply the modified definition of K-anonymity requirement, we need a generalization of level 2 to "Date of birth" field. Generalization of level 2 for "Date of birth" field could be defined with the suppression of "dd" and "mm" in "dd/mm/yyyy" So, Table 4 satisfies the modified definition.

Therefore, as we can see that, with this modified definition, there are four matching records in Table 4 with values of medical symptoms = {"shortness of breath", "SLE"}, which makes the record protected from reidentification. That implies the anonymized data that we get after applying 2-anonymity, 2-diversity on our data has left the granularity of the information to decrease substantially in exchange for improved privacy. A closer analysis reveals that the "Zip" column itself is left with just one piece of information which is "2330XX", which, however, may not be necessarily true for all such released data. However, in most cases where the records are collected in close geography, such occurrences may not be surprising as the "Zips" in the context of India are distributed accordingly in a range-based order. Similarly, but on all records, there is a visible loss in the "months" part of the "Date of birth" field. By assuming an equal weightage of "days," "month," and "year," we just lost $1/3^{\text{rd}}$ of the information fields.

Thus, the result of the application of k-anonymity and l-diversity, as it obliviously appears to be simply in black and white. That is, applying l-diversity over k-anonymized data with the identification of at least one sensitive field results in the released data more robust to the attack and course granule, hence clearly leading to quality loss. It further means that the analysis ability is more restricted and we are

Table 3: Medical data released as anonymous, with venerable records.

Date of birth (dd/mm/yyyy)	Zip	Sex	Medical symptoms
XX/10/1982	2330 XX	Male	Shortness of breath
XX/01/1983	2330 XX	Female	Obesity
XX/10/1982	2330 XX	Male	Shortness of breath
XX/01/1983	2330 XX	Female	Obesity
XX/01/1983	2330 XX	Female	Hypertension
XX/09/1982	2330 XX	Male	SLE
XX/07/1982	2330 XX	Male	SLE
XX/09/1983	2330 XX	Female	Hypertension
XX/09/1983	2330 XX	Female	Hypertension

Table 4: Medical data released as anonymous with 2-anonymity and 2-diversity applied.

Date of birth (dd/mm/yyyy)	Zip	Sex	Medical symptoms
XX/XX/1982	2330 XX	Male	Shortness of breath
XX/XX/1983	2330 XX	Female	Obesity
XX/XX/1982	2330 XX	Male	Shortness of breath
XX/XX/1983	2330 XX	Female	Obesity
XX/XX/1983	2330 XX	Female	Hypertension
XX/XX/1982	2330 XX	Male	SLE
XX/XX/1982	2330 XX	Male	SLE
XX/XX/1983	2330 XX	Female	Hypertension
XX/XX/1983	2330 XX	Female	Hypertension

losing on data quality. With real data, we will try to validate this apparent and obvious conclusion in the below section to understand the contradiction with the chosen standard set of evaluation parameters.

3. Application of Two Anonymity Algorithms to Nonfictitious Data from the Domain of Food Technology

We have used "ARX data Anonymization Tool" which is distributed under the "Apache License, Version 2.0". It is an open-source tool with a wide range of data anonymization techniques implemented for professional use.

We have used an attribute subset of the approved food establishments for November 2018 of the UK government [31]. The attribute selection is done with the purpose of analysing locationwise activities of the plant to provide skills training for job opportunities of local man power. However, no rows are reduced from the actual data. The fields are as shown in the tabular structure below. For better understanding, we have shown the top five rows in Table 5. The total number of records in this data set is 6455, which is sufficiently large data for practical analysis and evaluation purpose. We also want to address the privacy concern in recognition of the business owners, so the "App No" and the "Trading Name" fields are treated as identifiers and hence suppressed during the application of the anonymity algorithms.

For illustration purpose, Table 6 shows the 2-anonymized data sample of the records in the data set.

The quality evaluation is done with the following standard parameters:

(i) Gen. intensity
(ii) Granularity
(iii) N.-U. entropy
(iv) Discernibility
(v) Record-level squared error
(vi) Attribute-level squared error

The following strategy is used to anonymize data using the ARX tool with the following values of k and l as defined in k-anonymity and l-diversity algorithms and the observed results in the "Result and Discussion" which is the next section:

(i) 2-anonymity
(ii) 4-anonymity
(iii) 6-anonymity
(iv) 2-anonymity, 2-diversity
(v) 4-anonymity, 2-diversity
(vi) 4-anonymity, 3-diversity
(vii) 4-anonymity, 4-diversity
(viii) 6-anonymity, 2-diversity
(ix) 6-anonymity, 3-diversity
(x) 6-anonymity, 4-diversity
(xi) 6-anonymity, 5-diversity
(xii) 6-anonymity, 6-diversity

The value of l as in l-diversity is limited to be not greater than the value of k as in k-anonymity to let diversity not dominate the uniqueness of record in order to control data quality.

4. Results and Discussion

Table 7 presents the quality parameters evaluation in each column with ascending value of k as in the k-anonymization algorithm. We have not introduced the diversity factor in this part of the experiment to observe the behavior of the

Table 5: Attributes-subset of approved food establishments for November 2018 by the UK government, top five rows.

App. no.	Trading name	Town	Postcode	Country	All activities
AR 001	Monteum Ltd.	Shoreham	BN43 6RN	England	Dispatch centre (LBM), processing plant (fish)
AR 003	Southover Foods Ltd.	Southwick	BN42 4EN	England	Processing plant (meat)
AR 008	Higgidy Ltd.	Shoreham	BN43 6PD	England	Processing plant (meat)
AR 009	Little Tums	Shoreham-By-Sea	BN43 6NZ	England	Processing plant (meat)
AR 010	Malpass Markets	Shoreham	BN43 6RN	England	Mincemeat establishment, meat preparation establishment
...	...	...	...	...	...

Table 6: Record-samples of the food establishment data set after applying 2-anonymity.

App. number	Trading name	Town	Postcode	Country	All activities	Processing plant	Geographic local authority
*	*	Shetland	NA	Scotland	Auction Hall (fish)	NA	Shetland Islands
*	*	Shetland	NA	Scotland	Auction Hall (fish)	NA	Shetland Islands
*	*	Aberdeen	NA	Scotland	Cold store	NA	Aberdeen City
*	*	Aberdeen	NA	Scotland	Cold store	NA	Aberdeen City
...	...	...	...	...	...	...	...

Table 7: Results of quality parameters of the food establishment data set after applying 2-anonymity, 4-anonymity, and 6-anonymity.

Model	2-anonymity	4-anonymity	6-anonymity
Gen. intensity	15.42343	6.69984	6.69984
Granularity	16.12703	7.25019	7.25019
N.-U. entropy	13.06863	4.91985	4.91985
Discernibility	16.11756	7.2399	7.2399
Record-level squared error	14.99271	6.26457	6.26457
Attribute-level squared error	18.14774	8.74009	8.74009

k-anonymity algorithm independently as a sole factor. There are two interpretations that could be made with the following data:

(1) With the increase of the value of k, the data quality dips and so does the record-level and attribute-level squared error. However, both error metrics are reducing decisively due to strong generalization.

(2) There is a certain point where the data quality metrics stop descending and stabilize with the increase in the k-values.

Table 8 is an interesting abnormal behavior which is observed when both k and l values are too low and equal. The quality metrics results not only suggest a jump in data quality but also strangely high values of the record-level and attribute-level squared errors. In particular, the record-level squared error shoots close to 100%.

This is a false indicator as the quality improvement is driven by small but equal diversity leading to very high error quantification. It is presented just to conclude that it could be preferred to ignore such values while choosing the combination of k and l values.

Table 9 shows the data quality improved from 2-anonymity, 2-diversity which we chose to discard as an anonymization strategy in our case, with a higher value of $k = 4$. Thus, we observe steady and gradually. This gradual descending in quality is driven by l-value. The error metrics still remain high due to the high level of generalization even with a small value of k.

Table 8: Results of quality parameters of the food establishment data set after applying 2-anonymity, 2-diversity.

Model	2-anonymity, 2-diversity
Gen. intensity	21.73349
Granularity	22.83501
N.-U. entropy	19.25503
Discernibility	22.81606
Record-level squared error	99.96683
Attribute-level squared error	21.27768

Results in Table 10 are in line with the results in Table 9. The small and steady data quality fall continues with the increase in l-value, but the increase of k-value from 4 to 6 has a comparatively larger impact on data quality decline.

Figure 1 presents a single-frame observation to relate the above discussion. As it is observed from the stacked plot of all anonymization strategies, put together that the k-value is the more dominant factor in reducing the anonymized data quality compared to l-value. That is, in other words, generalization deteriorates the data quality more compared to diversity. Hence, diversity is good for privacy control and less evil for data quality. With this data sample, we reached stability in quality with a very less value of k both at $k = 4$ and $k = 6$ leading to unchanged values of quality metrics. The crux is that we reached an early and sharp data quality

Table 9: Results of quality parameters of the food establishment data-set after applying 4-anonymity, 2-diversity; 4-anonymity, 3-diversity and 4-anonymity, 4-diversity.

Model	4-anonymity, 2-diversity	4-anonymity, 3-diversity	4-anonymity, 4-diversity
Gen. intensity	14.21253	13.13973	11.91992
Granularity	15.52285	14.4694	13.27653
N.-U. entropy	11.54522	10.60076	9.35309
Discernibility	15.49844	14.44529	13.24972
Record-level squared error	99.90506	99.89778	99.87712
Attribute-level squared error	13.2648	12.21016	10.94421

Table 10: Results of quality parameters of the food establishment data-set after applying 6-anonymity, 2-diversity; 6-anonymity, 3-diversity; 6-anonymity, 4-diversity; 6-anonymity, 5-diversity; 6-anonymity, 6-diversity.

Model	6-anonymity, 2-diversity	6-anonymity, 3-diversity	6-anonymity, 4-diversity (%)	6-anonymity, 5-diversity	6-anonymity, 6-diversity
Gen. intensity	11.87162	11.34489	10.94	10.34937	8.76683
Granularity	13.27653	12.74981	12.36	11.71185	10.22463
N.-U. entropy	9.14062	8.70266	8.34	7.76506	6.37721
Discernibility	13.24748	12.72163	12.33	11.68218	10.19514
Record-level squared error	99.84593	99.84606	99.84	99.82315	99.78056
Attribute-level squared error	10.64448	10.16837	9.79	9.2042	7.41376

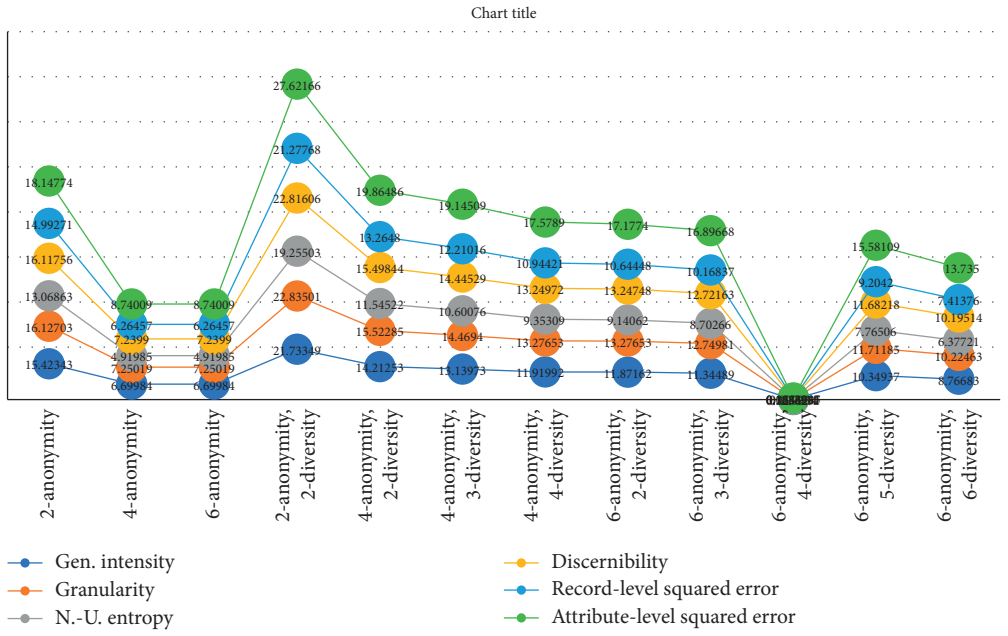

Figure 1: Single-frame result as stacked line plot of quality parameters.

decline before stability. That means a large amount of data loses its analytical opportunity with less anonymization. Quality loss appears overweighing to the privacy gains in our experiment. And if we try to control the quality loss with less generalization, that is, low value of k, then introducing diversity would result in abruptly high data quality loss.

5. Conclusions

This paper discusses the open data movement and its benefits in brief. It illustrates the need for opening public data that could give transparency to the citizen towards governance for the purpose of research, transparency, and fair opportunity to involve in governance to all its citizens.

It also identifies "anonymity" as one of the key barriers of open data. Although privacy for opening the data has been a concern for more than two decades, when there was very limited data compared to the data, the world holds today. There have been attempts to solve this barrier technically which had been a success at a major level. The K-anonymity algorithm and its successor algorithms borrow the basic assumption set established by the K-anonymity algorithm.

However, many critics focus on the aspect that the data which are claimed to be privacy-preserving after applying such algorithms, in fact, do not guarantee protection. Such critical reviews break open the loopholes to prove broken promises. On one side, this paper aims to provide a review with a simple example that could make readers understand that, on one side, algorithms compete to achieve better anonymity, on the other side they tend to compromise on data granularity. This makes it less usable for the analysts. It is just that the vulnerability has been fixed to break the usability of the released data. This makes more sense when imagining how important the granularity of the data was at the initial phase of the COVID-19 pandemic. As the leaders worldwide are worried about quick mitigation and fire-fighting in any such future unfateful circumstances, they must also be worried about doing more to open data to the maximum possible level and invest in finding solutions to the opposing force that is "privacy concern" while opening the data to the world. We studied a controlled strategy of k-value and l-value combinations and their impact on data quality. When the anonymity algorithms are needed to be applied on data with varying sizes, nature, and release purpose, the objective of choosing the right fit algorithm becomes an extremely difficult job with a range of algorithms and their individual variations. Moreover, the level of privacy control is a separate dimension, which applied with quality metrics parameters over a wide range of solutions would result in large 3-dimensional matrics or more that would be required to be translated to a one-dimensional result set to pick one suitable algorithm as a solution for that category of data set. Such a complex problem needs aggressive employment of machine learning and artificial intelligence, because of the complexity in finding the final most optimal solution for data anonymization.

Data Availability

Monthwise data for approved food establishments in the United Kingdom are available at https://data.food.gov.uk/catalog with many other food-related data sets. This data set is provided by "Food Standard Agency," UK. The link to the particular data we used is also provided in the reference section [31].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] https://techcrunch.com/2020/04/22/eu-privacy-body-urges-anonymization-of-location-data-for-covid19tracking/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAAPCYNlk1NPDYYq5GoZav8dmtRQr0D5IGArDD9suwh_qViS5Gpe0ANLTPkX68DfSTKr-nR_Nj98KZU62IUMR5FPDGJ7aTcAiF-fZksS9KuHyFs979CzhBvbdUKh-xlF6aHIaV8k6cLKjCB8qXGH_aFdaH1pOL.

[2] https://economictimes.indiatimes.com/magazines/panache/most-covid-19-apps-with-contact-tracing-feature-may-be-putting-your-personal-data-at-risk/articleshow/76279063.cms.

[3] https://www.who.int/about/ethics/integrity-hotline/en/.

[4] T.G Davies and A. Bawa, "The promises and perils of open government data (OGD," *The Journal of Community Informatics*, vol. 8, no. 2, 2012.

[5] https://privacyinternational.org/examples/tracking-global-response-covid-19.

[6] M. Ienca and E. Vayena, "On the responsible use of digital data to tackle the COVID-19 pandemic," *Nature Medicine*, vol. 26, no. 4, pp. 463-464, 2020.

[7] O. Data and S. Data, *Fact Sheet an Introduction to Open, Shared and Closed Data What is Open Data? What is Shared Data? What is Closed Data?*, 2012, http://www.alerc.org.uk/uploads/7/6/3/3/7633190/an_introduction_to_open_shared_and_closed_data.pdf.

[8] S. Khan and J. Foti, "Aligning supply and demand for better governance," *Open Data in the Open Government Partnership*, vol. 40, 2015.

[9] P. Conradie and S. Choenni, "On the barriers for local government releasing open data," *Government Information Quarterly*, vol. 31, pp. S10–S17, 2014.

[10] T. G. Davies and Z. A. Bawa, "The promises and perils of open government data," *Journal of Community Informatics,* vol. 8, pp. 7–13, 2012.

[11] Z. Irani and M. Kamal, "Transforming government: people, process, and policy," *Transforming Government: People, Process and Policy*, vol. 10, no. 2, pp. 190–195, 2016.

[12] R. Kitckin, *The Data Revolution*, SAGE Publication Ltd., Thousand Oaks, CF, USA, 2014.

[13] J. Bates, "The domestication of open government data advocacy in the United Kingdom: a neo-Gramscian analysis," *Policy & Internet*, vol. 5, no. 1, pp. 118–137, 2013.

[14] Het Informatiehuis Marien: Open Data. (2017).

[15] A. Larquemin and S. Buteau, "Open government data and evidence-based socio- economic policy research in India: an overview," *The Journal of Community Informatics*, vol. 12, pp. 120–147, 2016.

[16] F. Gonzalez-Zapata and R. Heeks, "The multiple meanings of open government data: understanding different stakeholders and their perspectives," *Government Information Quarterly*, vol. 32, no. 4, pp. 441–452, 2015.

[17] S. Baack, "Datafication and empowerment: how the open data movement re-articulates notions of democracy, participation, and journalism," *Big Data & Society*, vol. 2, no. 2, 2015.

[18] M. P. Canares, "Opening the local: full disclosure policy and its impact on local governments in the Philippines," in *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance 2014*, pp. 89–98, Guimaraes, Portugal, October 2014.

[19] http://opendatabarometer.org: 4th Global Ranking OpenData Barometer, (2017).

[20] P. Samarati, "Protecting respondents identities in microdata release," *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 6, pp. 1010–1027, 2001.

[21] S. Zhong, Z. Yang, and R. N. Wright, "Privacy-enhancing k -anonymization of customer data," in *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, Baltimore, ML, USA, June 2005.

[22] R. J. Bayardo and R. Agrawal, "Data privacy through optimal k-anonymization," in *Proceedings of the 21st International Conference on Data Engineering*, pp. 217–228, Tokyo, Japan, April 2005.

[23] M. S. Simi, K. S. Nayaki, and M. S. Elayidom, "An extensive study on data anonymization algorithms based on K-anonymity," *IOP Conference Series: Materials Science and Engineering*, vol. 225, 2017.

[24] P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppresion," in *Proceedings of IEEE Symposium on Security and Privacy*, pp. 384–393, Oakland, CA, USA, May 1998.

[25] L. Brankovic, Z. Islam, and H. Giggins, *Privacy-Preserving Data Mining*, Springer, Berlin, Germany, 2007.

[26] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: architecture and algorithms," *IEEE Transactions on Mobile Computing*, vol. 7, no. 1, pp. 1–18, 2008.

[27] J. Domingo-Ferrer and V. Torra, "A critique of k-anonymity and some of its enhancements," in *Proceedings of Third International Conference on Availability, Reliability and Security, ARES 2008*, pp. 990–993, Barcelona, Spain, March 2008.

[28] M. Bezzi, "An information theoretic approach for pri- vacy metrics," *Trans. Data Privacy*, vol. 3, pp. 199–215, 2010.

[29] P. Ohm, "Broken promises of privacy: responding to the surprising failure of anonymization," *UCLA Law Rev*, vol. 57, pp. 1701–1777, 2010.

[30] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "*L* -diversity: privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, 2007.

[31] https://data.food.gov.uk/catalog/datasets/1e61736a-2a1a-4c6a-b8b1-e45912ebc8e3.